Revisiting the standard for modeling functional brain network activity: application to consciousness

Grigis Antoine, Gomez Chloé, Frouin Vincent, Uhrig Lynn, Jarraya Béchir

▶ **To cite this version:**

Grigis Antoine, Gomez Chloé, Frouin Vincent, Uhrig Lynn, Jarraya Béchir. Revisiting the standard for modeling functional brain network activity: application to consciousness. 2024. hal-04511450

HAL Id: hal-04511450
https://hal.science/hal-04511450

Preprint submitted on 19 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Revisiting the standard for modeling functional brain network activity: application to consciousness

Grigis Antoine[1,*], Gomez Chloé[1,2], Frouin Vincent[1], Duchesnay Edouard[1], Uhrig Lynn[1,2], Jarraya Béchir[1,2,3]

1 Université Paris-Saclay, CEA, NeuroSpin, 91191, Gif-sur-Yvette, France
2 Cognitive Neuroimaging Unit, Institut National de la Santé et de la Recherche Médicale U992, Gif-sur-Yvette, France
3 Université Paris-Saclay (UVSQ), Neuroscience Pole, Foch Hospital, Suresnes, France

Abstract

Consciousness can be characterized by studying spontaneous fluctuations in brain activity, commonly measured with resting-state functional Magnetic Resonance Imaging (rs-fMRI). Previous rs-fMRI studies in monkeys and humans have shown that different levels of consciousness are defined by the relative prevalence of different dynamical functional connectivity patterns, also called brain patterns. These patterns closely match the underlying structural connectivity when consciousness is lost. The results suggest that changes in the state of consciousness lead to changes in connectivity patterns, not only at the level of co-activation strength between regions, but also at the level of entire networks. Here, we use a linear latent variable model that provides interpretable brain networks to reveal a new signature of consciousness and its chemically induced loss during anesthesia. To identify interpretable spatial signatures of consciousness, we apply a four-step framework by i) generating a list of atlases, ii) filtering and extracting the time series associated with the brain Regions of Interest (ROIs) of each atlas, iii) decomposing the signals into tailored brain networks with associated Brain Network Activities (BNAs), and iv) performing statistical inference and multivariate analysis of the BNAs. The novelty of the framework lies in the adoption of a constrained linear latent variable model that provides BNAs based on identifiable and disjoint ROIs, called brain networks, and the ability to offer a sound basis for atlas selection given the underlying clinical question. The model yields a set of tailored brain networks and associated BNAs that characterize states of consciousness. Our results suggest that a network composed of fronto-parietal and cingular cortices strongly influences the shift of consciousness state, especially between anesthesia and wakefulness. Interestingly, this observation is consistent with the global neural workspace theory of consciousness. We also decipher the level of anesthesia from rs-fMRI-derived BNAs. We identify neurobiologically relevant brain networks that provide novel interpretable signatures of consciousness and its loss during anesthesia. These findings pave the way for translational applications such as the diagnosis of consciousness disorders.

Introduction

Linking the state of consciousness to brain activity is a complex endeavor. In neuroscience, we study the relationship between consciousness and the brain by looking at how certain properties of the brain change between different states of consciousness. In particular, we compare normal wakefulness with altered states of consciousness experimentally induced by different anesthetic conditions. This comparison is conducted using computational models designed to accurately decode the level of consciousness from brain activity recordings such as electroencephalography (EEG) or functional magnetic resonance imaging (fMRI). Deciphering such activity remains a major challenge in neuroscience and medicine. It is also crucial for a deeper understanding of individual cognition, perception, and subjective awareness. Decoding levels of consciousness provides a

means to explore and categorize the different states of consciousness and mental processes that individuals can exhibit.

The anesthesia-induced Loss of Consciousness (LoC) was first studied using EEG [1–3]. The EEG signal depends on the type of anesthetic used and varies at different stages and depths of anesthesia [4]. Anesthesia can significantly modulate synaptic activity and neuronal responses and induce changes in measured cortical activity [4–8]. Each anesthetic also has its own specific effects on the brain [9–11]. The mechanisms involved may directly affect the neurovascular coupling [12] and therefore may have a direct effect on the vascular response measured in fMRI (called the blood oxygen level (BOLD) effect) [13, 14]. Working with the fMRI signal involves understanding how neural activity translates into observed changes in blood flow and oxygenation. A number of studies showed distinct profiles of altered Functional Connectivity (FC) from isoflurane and ketamine in Non-Human Primates (NHPs) [15, 16]. Although less sensitive to the specific anesthetic and molecular target, fMRI analysis at the level of cortical brain Regions Of Interest (ROIs) showed a clear conscious/unconscious separability [17–19]. Here, instead of a bottom-up investigation starting from molecules and cells that would take into account anesthetic specificities, an information-processing analysis of functional brain imaging is used to study different levels of consciousness.

Measuring consciousness with such ROI-based analysis is a challenging task due to its subjective nature. However, researchers have developed neuroimaging techniques that provide insight into the neural correlates of consciousness. In particular, resting-state fMRI (rs-fMRI), which is accessible to healthy subjects and patients, has been widely used to characterize spontaneous fluctuations in brain activity [20, 21]. The study of correlations between resting-state time-course signals in brain ROIs generate the so-called FC matrix. Some induced patterns have been shown to correlate with states of consciousness, such as sleep [22] or anesthesia [18, 23]. Regional connectivity patterns are classically studied over the entire fMRI recording (named FC to leave the notation uncluttered) or over time intervals to characterize the temporal dynamics of rs-fMRI activity (named dFC). FC captures the average or stable patterns of functional connections between brain regions, while dFC focuses on the time-varying nature of these connections, providing a more detailed understanding of dynamic brain processes. Both approaches contribute to our understanding of brain function and can provide valuable insights when studying Disorders of Consciousness (DoC). Here, we focus on characterizing the overall organization of functional networks in the brain (i.e., using FC) and understanding how different brain regions work together. Our goal is to develop an FC-driven computational model of the brain to identify patterns associated with conscious or unconscious experience. A key factor in the choice of this model will be the interpretability of the results. This promotes responsible use of machine learning models and enables users to understand and effectively use their results.

The proposed experiments will be performed on a retrospective rs-fMRI dataset composed of NHPs recorded across states of consciousness by modeling LoC with finely tuned EEG-controlled anesthesia [18, 19]. On this dataset, existing analyses fall either under unsupervised classification approaches [18, 19] or under a priori computational model identification [24]. We believe that reanalyzing the same dataset is valuable for several reasons. First, the neuroscience community has recently adopted robust data analysis frameworks that combine both supervised and unsupervised models. These frameworks aim to exploit as much experimental information as possible. Second, these new frameworks replace classical unsupervised approaches with factorial model fitting. These models produce latent variables that can be interpreted by design. With such a new analysis paradigm, we expect to automate the extraction of previous discoveries and to reveal new insights, alternative interpretations, or previously unnoticed patterns or relationships within the data. We also believe that it will allow the exploration of new research questions, such as the possibility of using the existing data in an application perspective, with the ultimate goal of providing tools to assist physicians in diagnosing DoC [25].

Using this dataset, our group has previously described two brain signatures of anesthesia-induced LoC [18, 19, 24]. The first signature is revealed by studying whole-brain dynamics from the rs-fMRI signal. Unsupervised clustering of the dFCs reveals distinct functional connectivity patterns, also called brain patterns. These states are ranked according to their similarity to the underlying structural connectivity. Analyzing the relationship between functional connectivity patterns and structural connectivity is crucial for gaining a comprehensive understanding of how the brain's structure influences its function. Indeed, both functional and structural connectivity are key components in unraveling the complexity of the brain's network

architecture. A comprehensive brain signature of anesthesia-induced LoC is provided by the relative prevalence of these ranked brain patterns in the different awake or anesthetized acquisition conditions. However, the interpretation of these results remains complex (see S1 Fig). Alternatively, a second signature is identified by examining the frequency of synchronization levels present in the rs-fMRI signal. Such an analysis of synchronization patterns provides a physiological proxy for the level of consciousness. Unfortunately, this proxy also shows poor detection ability (see S2 Fig). Overall, and in both cases, we believe that a single brain-wide metric limits the interpretability of the model and is probably inappropriate for modeling the underlying fine-grained spatial processes.

Finally, there are several theoretical frameworks to study consciousness. The two prominent theories are the Global Neuronal Workspace (GNW) [26,27] and the Integrated Information (II) [28,29] theories. Initially, a semi-automated approach was used to extend the GNW theory to monkeys using the considered retrospective dataset [19]. Thus, in the present work, we focus on the GNW theoretical model of conscious access. This model states that a piece of information becomes conscious when it is available to a widely distributed cortico-cortical network [26,27,30,31]. Although the GNW model of consciousness is not a localizationist approach, key neuroanatomical landmarks such as the prefrontal cortex, the parietal cortex, and the cingular cortex are responsible for broadcasting information to make it globally available [25]. Neuroimaging studies have shown disorganized GNW in patients with DoC [32–34] and in anesthetized monkeys [35,36]. Thus, anesthesia may induce LoC by reconfiguring cortical dynamics, primarily within the GNW nodes. In the proposed study, the target ROIs will consist of these cortical nodes.

In this work, we develop a computational framework for identifying interpretable spatial signatures of consciousness from rs-fMRI data. The proposed framework characterizes the different levels of consciousness. It consists of four steps (Fig 1): i) generating a list of atlases, ii) filtering and extracting the time series associated with each atlas ROI, iii) recovering disjoint brain networks and associated Brain Network Activities (BNAs) from ROI-based FC matrices that reflect the empirical covariance structures of the data [37], and iv) performing statistical inference and multivariate analysis on the BNAs. The brain networks are disjoint sets of brain regions, and the associated latent variables (the BNAs) form our spatial signatures. As input, standard FC is calculated by averaging time course signals in ROIs, assuming functional consistency within regions. Different atlases can describe these collections of ROIs. However, the functional relationships between them vary from atlas to atlas. At the heart of atlas selection is the question of whether different conditions lead to consistent choices, and whether genericity should be preferred over adaptive strategies. Thus, we extend the work of Monti and colleagues [37] to address the atlas selection problem. In Monti's seminal work, a strategy is developed to select the optimal number of components/brain networks when decomposing FC structures. In our framework, a machine learning paradigm also provides a sound basis for atlas selection given the underlying clinical question.

The remainder of this manuscript is organized as follows. First, we describe the retrospective dataset studied, including data collected during awake or anesthesia-induced LoC. Then, we provide a detailed explanation of the Modular Hierarchical Analysis (MHA) method and present the considered atlas collection. Finally, we demonstrate that the resulting tailored brain networks and associated BNAs are interpretable. They provide original insights into the study of brain dynamics and function. Overall, we establish a robust compendium of spatial biomarkers of consciousness. Among the identified brain networks, one is highly consistent with the GNW. This demonstrates that the MHA can be used in discovery-driven analysis to detect new interpretable signatures of consciousness and anesthesia.

Materials and methods

Dataset

This study is a retrospectively analysis of functional imaging data collected nearly 10 years ago in non-human primates in various states of consciousness [18,19]. The data were acquired in five rhesus macaques (macaca mulatta), one male (monkey J) and four females (monkeys A, K, L, and R), 5 to 8 kg, 8 to 12 years, either in the awake state or under general anesthesia using different molecular agents (ketamine, propofol, or sevoflurane) representing 6 anesthetic conditions. Three monkeys were scanned for each arousal state (awake:

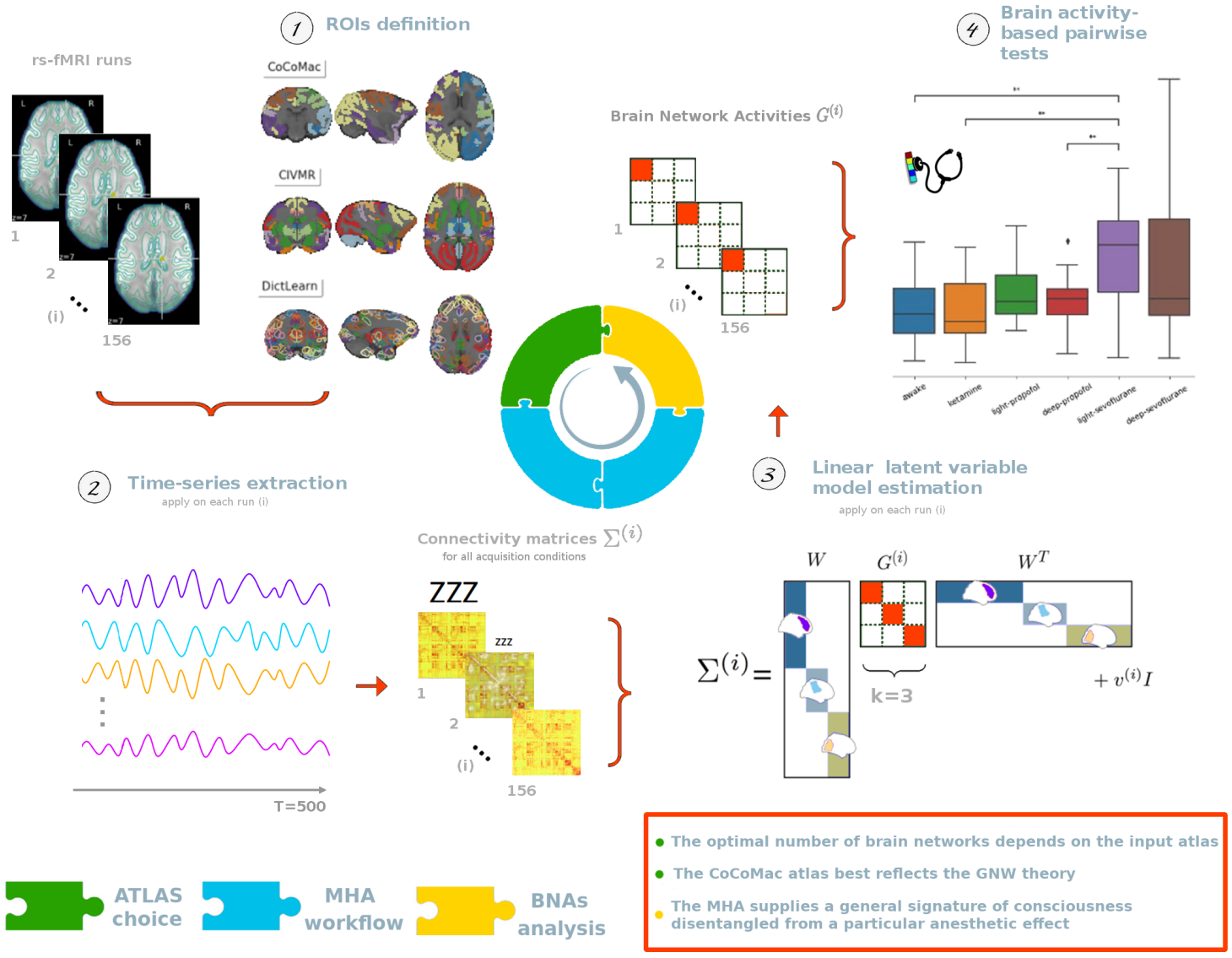

Fig 1. Illustrating the application of the proposed four-step framework to the characterization of different levels of consciousness. Note that the box plot shows the distribution of the functional activities $G^{(i)}$ that are associated with the brain network 1 for the CoCoMac atlas across all the acquisition conditions. The level of consciousness is indicated by the size of the "ZZZ" notation.

monkeys A, K, and J - propofol anesthesia: monkeys K, R, and J - ketamine anesthesia: monkeys K, R, and L - sevoflurane anesthesia: monkeys L, R, and J). For the anesthetics used in this retrospective study, the molecular targets are summarized in S1 Table. Levels of anesthesia were defined by a clinical arousal score (the monkey sedation scale) and continuous EEG monitoring (for details of the anesthesia protocol, see the anesthesia paper [19]). Two different levels of anesthesia are considered for propofol and sevoflurane, either moderate sedation or deep sedation equivalent to general anesthesia, and deep sedation for ketamine. 156 rs-fMRI runs (500 volumes per run, TR=2.4s) were acquired on a 3T Siemens with a customized single transmit-receiver surface coil (see S2 Table for a detailed description of the acquisition conditions distribution across monkeys). The spatial preprocessing is performed by the pipeline described in [19,38], which includes the following steps: slice timing correction, B0 inhomogeneities correction, motion correction, reorientation,

masking, realignment, and smoothing. A run-by-run visual quality control was performed by an expert neuroimager to ensure the quality of the data (i.e. absence of common artefacts, motion, ...). The same monkey is implicated in several sessions and runs. Within-session runs may exhibit less variability than between-session runs. Due to legitimate ethical constraints, NHP data are inevitably sparse. Therefore, we independently analyze all 156 rs-fMRI runs acquired under any of the six listed anesthetic conditions. To evaluate this hypothesis, we employ a leave-one-subject-out strategy during the model estimation and inference stages. The goal is to validate the generalization of the model to unseen data. In the following experiments, we leave out J to maximize the number of acquisition conditions in the test set. Alternatively, we could have assessed the consistent performance of the model across subjects, indicating its potential robustness and reliability in different settings. Unfortunately, this approach is not realistic in this retrospective dataset due to the heterogeneity in the distribution of acquisition conditions across monkeys.

The original study was conducted in accordance with the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes (Directive 2010/63/EU), the National Institutes of Health's Guide for the Care and Use of Laboratory Animals. Animal studies were approved by the institutional Ethical Committee (Commissariat à l'Énergie atomique et aux Énergies alternatives; Fontenay aux Roses, France; protocols 10-003 and 12-086).

Establishing a list of atlases

The atlas selection problem is investigated by considering either general reference atlases or a data-driven atlas. General anatomical atlases are defined on other structural or functional datasets, sometimes mixed with histological sections and microscopy. Conversely, data-driven atlases are learned directly from the rs-fMRI data. The former involve the analysis of data based on predefined anatomical or functional regions. The latter are designed to capture the inherent structure and variability of the data itself. While general reference atlases may not fully capture the complexity and variability of the data, data-driven atlases can aid in the discovery of novel biomarkers and reveal unexpected associations, relationships, or features not previously considered. Indeed, such an approach allows for a more individualized and nuanced understanding of functional connectivity (by increasing precision and improving sensitivity) and may be useful for identifying previously unrecognized functional networks.

Here we consider the CoCoMac atlas, a well-accepted general atlas of the rhesus macaque consisting of 82 cortical regions (41 cortical regions within each hemisphere) [39]. We also consider the CIVMR atlas, which contains a selection of 222 cortical and subcortical regions [40]. To build an atlas based on the dataset, we use the Online Dictionary Learning (DictLearn) from the Nilearn development kit [41]. Using this learning process, we define ROIs directly from rs-fMRI data. A comprehensive analysis of several state-of-the-art strategies recently ranked DictLearn among the best with high robustness and accuracy [42]. DictLearn first masks the data with the cortical/subcortical spatial support of the CIVMR atlas. Next, it extracts connex brain activation regions from the dictionary maps using connected components analysis. We obtain a set of 246 adapted ROIs mapping cortical/subcortical contiguous areas, called the DictLearn atlas. The cross-validation scheme does not evaluate the potential variability of the DictLearn atlas estimation. Thus, the data-driven DictLearn atlas, like the CoCoMac and CIVMR atlases, is considered as prior knowledge of the proposed framework.

Time series filtering and extraction

Let $I^{(i)} \in \mathbb{R}^{x \times y \times z \times t}$, $i \in [1, N]$ be a rs-fMRI time series volume collected across our cohort of N=156 experimental runs. i refers to complete runs, i.e. multiple observations (statistical samples) of the same (time x space) stochastic process related to the different states of anesthesia. x, y, and z represent the spatial dimension of each volume, and t is the number of time points. Time series denoising operations are applied [18,19]. Specifically, voxel time series are detrended, filtered with low-pass (0.05-Hz cutoff), high-pass (0.0025-Hz cutoff), and zero-phase fast Fourier notch (0.03 Hz, to remove an artifactual pure frequency present in all the data) filters, regressed out from motion confounds, and z-score standardized. Let p be the fixed number of ROIs defined by the atlas. The feature matrix $X^{(i)} \in \mathbb{R}^{p \times t}$ contains the averaged time series computed across all voxels within each ROI. The high number of ROIs in some atlases encourages the

computation of the mapping between an atlas and the rs-fMRI data in the high-resolution template space. This way, topology issues induced by the atlas down-sampling are avoided (i.e., a region vanishing due to high contraction in the deformation field). From each feature matrix $X^{(i)}$, a FC matrix reflecting the data empirical covariance structure is derived and will be analyzed in the sequel.

Linear latent variable model estimation

The Modular Hierarchical Analysis (MHA)

A latent variable model is a statistical model that relates high-dimensional observed variables to low-dimensional unseen latent variables. Such a model can capture complex brain properties that are challenging to quantify or measure directly. From the empirical ROI-based covariance structures or FC matrices, the model estimates a set of distinct brain networks common to all experimental runs, and associated BNAs. This supports the identification of common patterns of brain activity that are consistently observed during different states of anesthesia. The identification of a common brain pattern repertoire across the six considered anesthetic states strengthens the proposed hypothesis [19]. In such a model, the latent variables are the BNAs that are specific to each run i. In a small number of networks regime, the MHA linear latent variable model has proven to yield more reproducible and explainable results than others such as Principal Component Analysis (PCA) or Independent Component Analysis (ICA) [37]. The MHA approach follows the probabilistic PCA formulation [43]. Briefly, the rs-fMRI observations $X^{(i)}$ are generated as a linear projection from low-dimensional latent variables $Z^{(i)} \in \mathbb{R}^k$. Both observations and latent variables are taken to follow a multivariate Gaussian distribution. We obtain the following generative model for observed data [37]:

$$\begin{aligned} Z^{(i)} &\sim \mathcal{N}(0, G^{(i)}) \\ X^{(i)}|Z^{(i)} = z^{(i)} &\sim \mathcal{N}(Wz^{(i)}, v^{(i)}\mathbb{I}) \end{aligned} \tag{1}$$

where $G^{(i)} \in \mathbb{R}^{k\times k}$ is the covariance of latent variables, k denotes the number of disjoint brain networks, and $v^{(i)} \in \mathbb{R}_+$ is the measurement noise. By capturing the low-rank covariance structure via the shared across-runs loading matrix W, the MHA model can reconstruct the covariance matrix in $\Sigma^{(i)}$:

$$\Sigma^{(i)} = WG^{(i)}W^T + v^{(i)}\mathbb{I} \tag{2}$$

$W \in \mathbb{R}^{p\times k}$ describes brain networks that are reproducible across the entire population. Each column j of W encodes the j^{th} brain network. For each run i, the matrix $G^{(i)}$ contains the latent variables of run i. More specifically, the j^{th} diagonal element of $G^{(i)}$ estimates the so-called BNA associated with the j^{th} brain network for run i. To compute the model parameters, the optimization maximizes the model log-likelihood $\mathcal{L}$ between $\Sigma^{(i)}$ and the empirical covariance structure $K^{(i)} = X^{(i)}X^{(i)'} \in \mathbb{R}^{p\times p}$ across all runs as follows:

$$\begin{aligned} \mathcal{L}(W, G^{(i)}) &= \sum_{i=1}^{N} p\ log(2\pi) + log\ det\ \Sigma^{(i)} + tr({\Sigma^{(i)}}^{-1}K^{(i)}) \\ \hat{W} &= \underset{W:W^TW=\mathbb{I};\ W\geq 0}{\operatorname{argmax}} \mathcal{L}(W, G^{(i)}) \end{aligned} \tag{3}$$

Over PCA, the MHA model adds a non-negativity constraint to the orthonormal constraint. It gives MHA the ability to uncover disjoint brain networks in W and associated run-wise BNAs in $G^{(i)}$. W has a block structure and is uniquely defined and identifiable. W can be thought of as a shared basis of k non-overlapping brain networks across all runs.

As in the work of Monti and colleagues [37], the choice of the optimal number of disjoint networks k in the model is treated as hyperparameter tuning. A leave-one-subject-out split is performed to generate a training set and a test set. The MHA model is fitted to the training set, and the log-likelihood $\mathcal{L}$ is maximized over the unseen test set. Using unseen data is crucial to evaluate the performance and generalization ability of the model and to avoid overfitting. However, maximizing the log likelihood over unseen data for hyperparameter

selection should be done with caution. We know that overfitting can occur critically in our dataset because multiple anesthetic states were administered to the same monkey. To mitigate this risk, we include all of a monkey's data in the test set.

Brain networks matching

For each atlas l, $l \in \{1 = CoCoMac,\ 2 = CIVMR,\ 3 = DictLearn\}$ we get k_l Brain Networks (BNs). In the following, we will compare these networks with a geometric criterion. Let BN_i^l, $i \in [1, k_l]$ be a brain network of a given atlas l. BN_i^l is composed of $N_{i,l}$ ROIs, defined as the number of non-zero values in the i^{th} columns of W^l. The goal here is to match brain networks across atlases. We start by simplifying a ROI (composed of V connected voxels) to its centroid (a 3D coordinate x,y,z). We define the mapping function $\phi : BN_i^l \in \mathbb{R}^{N_{i,l} \times V \times 3} \rightarrow \bar{BN}_i^l \in \mathbb{R}^{N_{i,l} \times 3}$, where ϕ is the mean operator. To compare brain networks between two atlases, we then introduce a proximity measure between a brain network i of atlas l_1 ($BN_i^{l_1}$) and a brain network j of atlas l_2 ($BN_j^{l_2}$). The proximity metric is defined as the Mean of the Closest (MoC) distance between the set of centroids defined in $\bar{BN}_i^{l_1}$ and $\bar{BN}_j^{l_2}$:

$$\begin{aligned} d_{MoC}(\bar{BN}_i^{l_1},\ \bar{BN}_j^{l_2}) &= \frac{d(\bar{BN}_i^{l_1},\ \bar{BN}_j^{l_2}) + d(\bar{BN}_j^{l_2},\ \bar{BN}_i^{l_1})}{2},\ where \\ d(\bar{BN}_i^{l_1},\ \bar{BN}_j^{l_2}) &= mean_{g_i \in \bar{BN}_i^{l_1}} min_{g_j \in \bar{BN}_j^{l_2}} ||g_i - g_j|| \end{aligned} \tag{4}$$

and $||.||$ is the euclidean norm. Note that the d_{MoC} proximity metric is symmetric to avoid inconsistencies between sets of different sizes.

Decoding brain network activities

Only the k-diagonal elements in $G^{(i)}$ are considered in the proposed analysis. Let $S^{(i)} = (s^1_{G^{(i)}}, ..., s^k_{G^{(i)}})$ denote the associated individual activities over the k discovered BNs. For all runs $i \in [1, N]$, let $G \in \mathbb{R}^{N \times k}$ be the concatenation of the individual activities $S^{(i)}$. Below we are interested in comparing the different BNAs contained in each column of G: $BNA^j = (s^j_{G^{(1)}}, ..., s^j_{G^{(N)}})$, $j \in [1, k]$. As such, we can interpret the BNAs as a measure of the activity within the corresponding brain networks. In other words, the BNAs represent the amount of variability carried by each brain network. We hypothesize that the off-diagonal entries of the latent variable covariances are not the most discriminative features for the clinical question under investigation. The non-zero off-diagonal values indicate the presence of redundancy in the data or some degree of correlation between variables.

BNA-based statistical inference

Group analysis is performed on the BNAs on an atlas basis to highlight the main discrepancy between anesthetic conditions. Applying the Shapiro-Wilk test reveals that the BNAs do not satisfy the normal assumptions. Therefore, pairwise nonparametric Wilcoxon signed-rank tests are used between paired grouped BNAs. The null hypothesis (H0) states that no significant difference exists between two awake/anesthetized conditions. p-values are adjusted for multiple comparisons using the Benjamini / Yekutieli False Discovery Rate (FDR) correction.

BNA-based multivariate analysis

Let the aforementioned $G \in \mathbb{R}^{N \times k}$ be the decomposed BNAs, and $y \in \mathbb{Z}_+^N$ the labels encoding the anesthetic conditions (awake state or moderate/deep ketamine, propofol, or sevoflurane anesthesia). Supervised machine learning can predict the outcomes y from the input features G. The proposed classification relies on Support Vector Machines (SVM) with a Radial Basis Function (RBF) kernel (as implemented in scikit-learn [44]). The gamma hyperparameter is automatically determined, while the C hyperparameter is set to 1. To limit overfitting during training in our small-size dataset, bagging is implemented. It aggregates multiple models

trained from the base SVM-RBF estimator by randomly taking training subsets. It enables the definition of an overall stronger predictor. As previously described, the model is trained using a leave-one-subject-out splitting to generate a training set and a test set. Model fitting is performed using five-fold cross-validation on the train set. Note that the above classifier treats each class as being non-ordinal and might miss the inherent relationship among the categories to learn. Thus, we also assess the benefit of using another base classification estimator that implements an Ordinal Logistic model with l2 regularization [45] (as implemented in mord [46]). The regularization parameter is set to 1. In both cases, the described hyperparameters are not evaluated in an internal cross-validation.

Brain network importance

Finally, it is reasonable to ask which brain networks are involved in the different predictions. Using a model-agnostic permutation importance technique (as implemented in scikit-learn [44]), we measure the importance of features by randomly permuting the values of a feature and evaluating the impact on the model's performance. By comparing the performance of the model with the permuted features to its original performance, one can determine which features have the most influence on the predictions. Features with a significant drop in performance after permutation are considered more important. Such a technique helps to identify which features or BNAs and their associated brain networks have the greatest impact on the model's performance, and can provide insight into the relationships between BNAs and anesthetic conditions.

Results

Consciousness-related connectivity can be decomposed into few consistent brain networks

The choice of atlas controls the number of input regions p fed into the model. Maximizing $\mathcal{L}$ yields salient $k = 4$, $k = 6$, and $k = 7$ optimal brain networks for the CoCoMac, DictLearn, and CIVMR atlases, respectively (Fig 2). For the CoCoMac atlas, the likelihoods for $k = 3$ and $k = 4$ are very close. We choose $k = 4$ to maximize the coverage of the ROIs (see S3 Table and S4 Table for a listing of this coverage for $k = 3$ and $k = 4$). Recall that the MHA constraints drive this coverage by conditioning the loading matrix W to have at most one non-zero entry per row and imposing sparsity with non-negativity. For the comparison of the brain networks from the three atlases, the similarity metric d_{MoC} is applied in a bottom-up fashion. Paired brain networks show remarkably symmetric and spatially consistent patterns of connectivity (Fig 3). Atlases that provide additional brain networks are globally symmetric (see S3 Fig). It should be noted that by considering all data, we maximize the possibility of obtaining a result that is valid across subjects. Furthermore, given the distribution of acquisition conditions across the monkeys, we felt that this decision to consider all data was justified. Nevertheless, to validate the robustness of the discovered BNs, we re-ran the MHA decomposition using a leave-one-subject-out strategy with the CoCoMac atlas. Interestingly, the generated BNs and BNAs, with a particular focus on BN1 and BN4, exhibit remarkable similarity (see S5 Fig). Furthermore, when listing the differences in terms of the ROIs included in each BN, the differences remain small (see S5 Table).

Changes in brain networks across states of consciousness

In terms of BNAs, the most significant differences between conditions (i.e., states of consciousness) are highlighted using pairwise statistics by preserving the network order (see Fig 4 for the CoCoMac atlas and S4 Fig for the CIVMR and DictLearn atlases). In contrast to the sliding window synchronization patterns [24], our statistics allow for highlighting a larger number of significant differences, with a notable one appearing between the awake state and all anesthetic conditions. The latter result emphasizes the interest of the Brain Network 1 (BN1), which is indicated by a star in Fig 4 and S4 Fig. Interestingly, when focusing on this network for the CoCoMac atlas (BN1 in Fig 4), the ROIs underlying this difference are perfectly symmetric and closely match the macaque GNW nodes [35]: the posterior cingulate cortex (CCp), anterior cingulate cortex (CCa), intraparietal cortex (PCip), frontal eye field (FEF), dorsolateral prefrontal cortex (PFCdl),

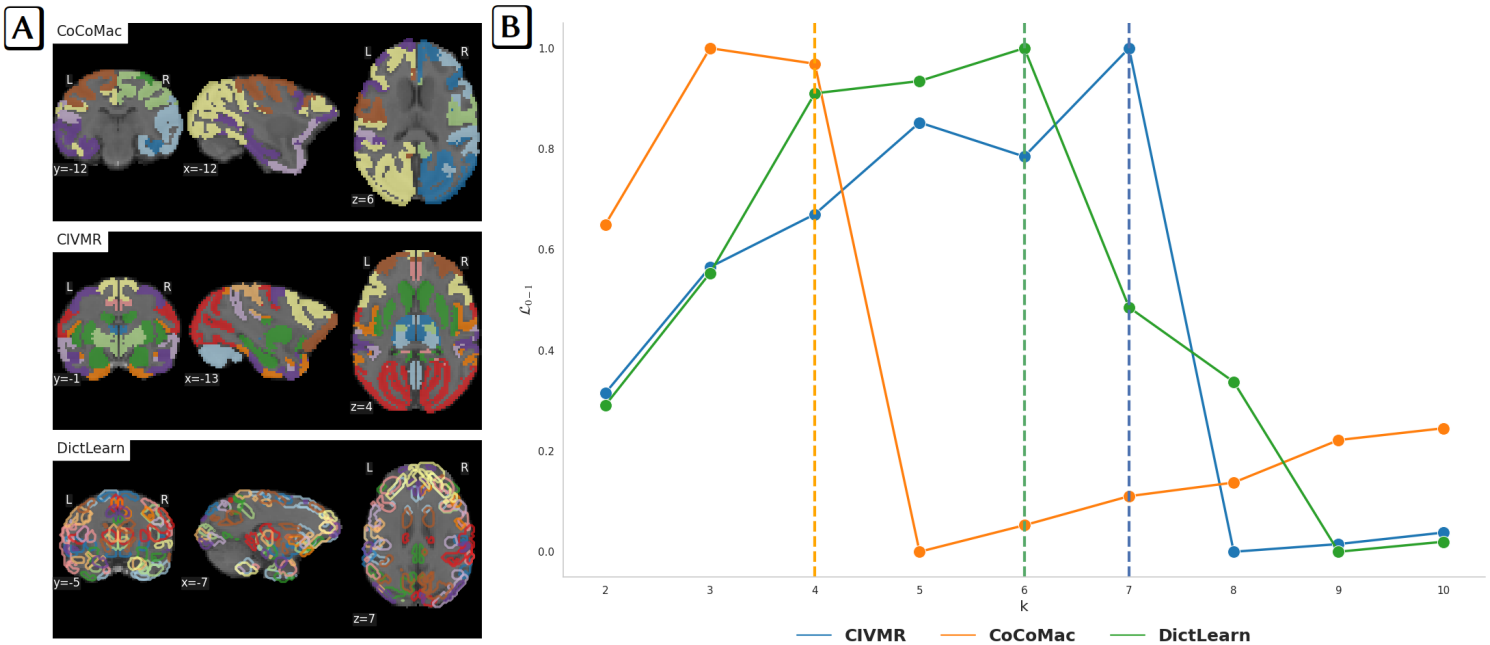

Fig 2. Illustrating how to select the best number of brain networks. A) the CoCoMac, CIVMR, and DictLearn atlases, and B) associated 0-1 normalized log-likelihood $\mathcal{L}_{0-1}$ on the unseen validation set for $k \in [2, 10]$. The optimal number of networks is represented by a vertical dashed line: k=4 brain networks for the CoCoMac atlas, k=6 for the DictLearn atlas and k=7 for the CIVMR atlas.

prefrontal polar cortex (PFCpol) and dorsolateral premotor cortex (PMCdl) areas, which includes the sensory regions of the primary motor cortex (M1), primary somatosensory cortex (S1), primary visual cortex (V1), and primary auditory cortex (A1) (Table 1-A). Although obtained in an unsupervised constrained manner, BN1 closely supports the GNW theory (7/11 nodes). The identification of the GNW can only be conclusive on the CoCoMac atlas, as it is the only atlas available to date with the macaque GNW description. To repeat the analysis with other atlases, we should first follow the methodology developed by Uhrig and colleagues [35] to establish atlas-specific GNW nodes.

Table 1. Listing of Brain Networks (BNs) inferred from the CoCoMac atlas.

	name	hemi	location
CCp	posterior cingulate cortex	left, right	cingulate cortex
CCa	anterior cingulate cortex	left, right	cingulate cortex
S1	primary somatosensory cortex	left, right	parietal cortex
PCi	inferior parietal cortex	left, right	parietal cortex
PCm	medial parietal cortex	left, right	parietal cortex
PCip	intraparietal cortex	left, right	parietal cortex
PCs	superior parietal cortex	left, right	parietal cortex
M1	primary motor cortex	left, right	frontal cortex
FEF	frontal eye field	left, right	frontal cortex
PMCm	medial premotor cortex	left, right	frontal cortex
PMCdl	dorsolateral premotor cortex	left, right	frontal cortex

(A)

	name	hemi	location
Amyg	amygdala	left, right	temporal cortex
TCc	central temporal cortex	left, right	temporal cortex
TCi	inferior temporal	left, right	temporal cortex
PHC	parahippocampal cortex	left, right	temporal cortex
HC	hippocampus	left, right	temporal cortex
TCv	ventral temporal cortex	left, right	temporal cortex
VACv	anterior visual area (ventral)	left, right	occipital cortex
V2	visual area 2	left, right	occipital cortex
VACd	anterior visual area (dorsal)	left, right	occipital cortex
V1	visual area 1	left, right	occipital cortex
CCr	retrosplenial cingulate cortex	left, right	cingulate cortex

(B)

A) the BN highlighting the difference between the awake state and anesthesia (the BN1 indicated by a star in Fig 4-A), and B) the inferred BN4 that is driven by the visual pathway. The detected GNW areas are shown in blue, and the corresponding sensory areas in green.

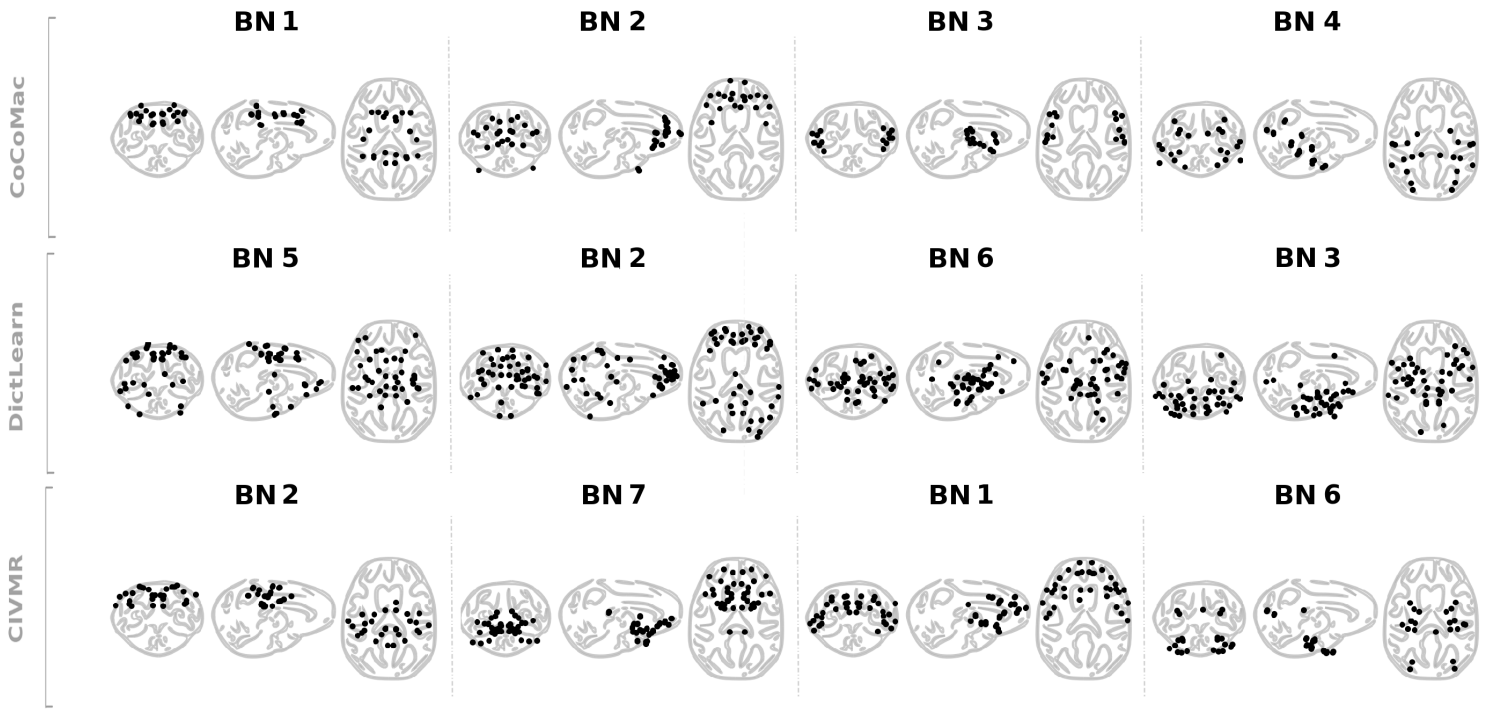

Fig 3. The derived brain networks (BNs). BNs consist of sets of unique ROIs represented by their centroids for the CoCoMac (k=4), DictLearn (k=6) and CIVMR (k=7) atlases. The BNs are sorted according to the geometric criterion d_{MoC}. Only the BNs that match in each atlas are displayed. The remaining BNs are listed in S3 Fig.

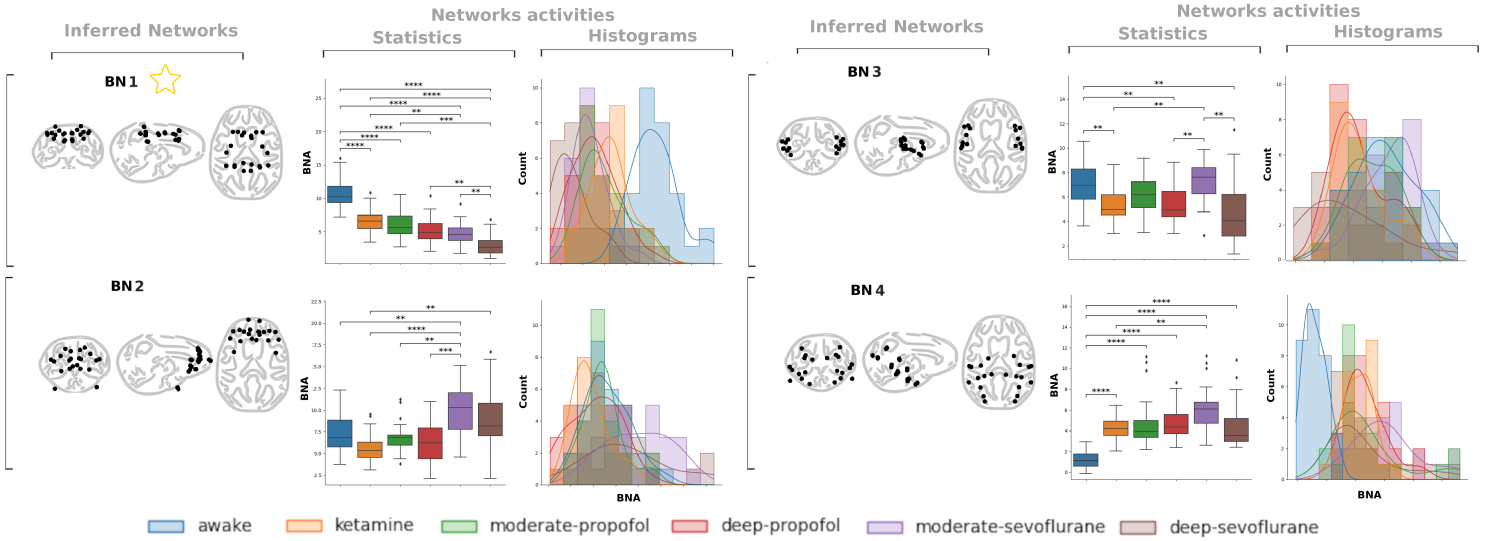

Fig 4. Inferred Brain Networks (BNs) and associated BN activities (BNAs). Four BNs are inferred from the MHA model using the CoCoMac atlas (BN1, BN2, BN3, BN4). Pairwise statistical analysis of the associated BNAs and visualization of these BNAs against the different acquisition conditions are proposed in addition to the display of the inferred networks. These plots illustrate how clearly awake can be distinguished from all anesthetic states and, to a lesser extent, how all anesthetic states can be distinguished from each other. The legend for the p-value annotation is as follows:
$** : 1.0e-3 < p \leq 1.0e-2, *** : 1.0e-4 < p \leq 1.0e-3, **** : p \leq 1.0e-4.$

Table 2. Brain activity based prediction of acquisition conditions using the CoCoMac, CIVMR and DictLearn atlases.

	CoCoMac			CIVMR			DictLearn		
SVM RBF	**All**	**DeepModerate**	**Anesthesia**	**All**	**DeepModerate**	**Anesthesia**	**All**	**DeepModerate**	**Anesthesia**
train	0.76 ± 0.06	0.84 ± 0.02	1.0 ± 0.0	0.8 ± 0.02	0.89 ± 0.03	0.97 ± 0.02	0.8 ± 0.04	0.88 ± 0.01	1.0 ± 0.0
validation	0.58 ± 0.09	0.78 ± 0.14	0.99 ± 0.012	0.61 ± 0.11	0.77 ± 0.08	0.96 ± 0.06	0.71 ± 0.1	0.79 ± 0.02	0.99 ± 0.01
test	0.24 ± 0.05	0.84 ± 0.05	0.97 ± 0.03	0.39 ± 0.08	0.54 ± 0.04	0.76 ± 0.03	0.44 ± 0.11	0.65 ± 0.08	0.75 ± 0.0
Ordinal Logistic	**All**	**DeepModerate**	**Anesthesia**	**All**	**DeepModerate**	**Anesthesia**	**All**	**DeepModerate**	**Anesthesia**
train		0.7 ± 0.02	1.0 ± 0.0		0.63 ± 0.05	0.9 ± 0.03		0.79 ± 0.02	1.0 ± 0.0
validation		0.73 ± 0.08	0.99 ± 0.02		0.56 ± 0.07	0.85 ± 0.21		0.75 ± 0.08	1.0 ± 0.01
test		0.79 ± 0.09	0.98 ± 0.01		0.63 ± 0.03	0.68 ± 0.04		0.67 ± 0.05	0.83 ± 0.02

The Balanced Accuracy (BAcc) metric is used to evaluate model performances. Three settings are considered: the awake state and all anesthetics are considered separately (All), the anesthetics are grouped by dosage (DeepModerate), or all anesthetics are encoded in the same group (Anesthesia). Two models are evaluated: the SVM RBF and Ordinal Logistic models. The CoCoMac atlas performs best on the DeepModerate and Anesthesia settings (blue). It remains difficult to predict all conditions, but the CIVMR and DictLearn atlases give the best performance (orange). In this case, the difference with the CoCoMac remains weak.

Which atlas best predicts depth of anesthesia from brain network activity?

With the BNA distributions, we can clearly distinguish the state of wakefulness from the state of anesthesia, regardless of the anesthetics administered to suppress consciousness (Fig 4). Further differences between anesthetics remain. Therefore, we perform a multivariate analysis of the BNAs to establish a decision rule for the atlas selection problem, taking into account the downstream anesthetic state classification task. To further explore BNAs, SVM-RBF and Ordinal Logistic models are considered for solving three classification tasks, corresponding to three sets of target labels: 1) the awake state and each anesthetic are considered separately (label set name: All), 2) the anesthetics are labeled by sedation level (label set name: DeepModerate), or 3) all anesthetics are encoded in the same group (label set name: Anesthesia). For the CoCoMac, CIVMR, and DictLearn atlases, BNA-driven predictions are listed in Table 2. The CoCoMac atlas performs best on the DeepModerate and Anesthesia tasks, demonstrating its relevance in characterizing the depth of anesthesia. It remains competitive to predict the conditions separately. In the following experiments, the SVM-RBF is retained as it gives better performance. Our goal now is to provide more insight into the brain networks discovered with the selected atlas. A brain network importance analysis promotes two brain networks (Fig 5-A). Second, BN1, mostly parieto-cingular, is described in the previous paragraph and contains most of the GNW nodes. First, BN4 (Table 1-B), mainly temporo-occipital, contains the visual pathway and may correspond to the fact that awake monkeys have open eyes with potential visual stimulation.

Sensitivity analysis of brain activity-driven predictions

Deciphering the level of consciousness from neural activity holds important potential for clinical application, i.e. the development of novel tools to objectively monitor the depth of anesthesia. To investigate the versatility of the method, we evaluate its learning curve as a function of the acquisition duration. This experiment is performed in the case of the CoCoMac atlas, using the three different labeling settings described in the previous paragraph: All, DeepModerate, and Anesthesia (Fig 5-B). The duration of the truncated time series ranges from 10 to 500 TR by steps of 10 TR increment. Despite the small dataset size, these plots suggest that a 200 TR run length gives accurate performances (~ 0.8 balanced accuracy for the Anesthesia setting). Thus, the proposed solution needs 200 TR to make reliable predictions. This learning curve analysis for a given time series duration is the only practical way to determine a steady state for the model. Although more data would be necessary to confirm this result, it provides an order of magnitude of the minimum buffer size for such an approach.

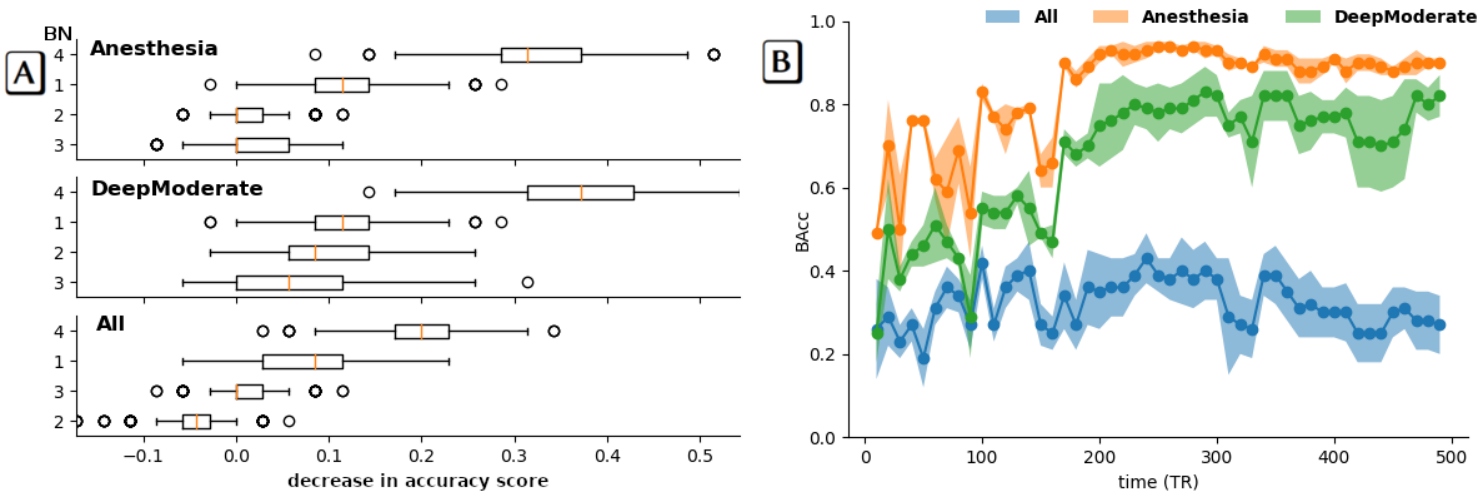

Fig 5. In-depth understanding of how the model works. A) Brain networks (BNs) with the greatest impact on the prediction based on feature of importance analysis for the CoCoMac atlas. Three settings are considered: the awake state and all anesthetics are considered separately (All), the anesthetics are grouped by dosage (DeepModerate), or all anesthetics are encoded in the same group (Anesthesia). Values on the x-axis encode how much model performance decreases with a random shuffling (using balanced accuracy as the performance metric). The amount of randomness is assessed by repeating the process several times. Strong positive values indicate features of interest, while negative values indicate predictions that are more accurate than the real data. The latter can happen when the feature is not important, but randomness makes the predictions more accurate. This is a known behavior with small datasets, which are more prone to random error. B) Learning curves with respect to the acquisition time for the CoCoMac atlas. The Balanced Accuracy (BAcc) metric is used to evaluate the performance of the SVM RBF model. For reasonable performance we need at least 200 TR.

Discussion

In the present study, we developed a novel approach for the analysis of rs-fMRI data obtained during different states of consciousness in macaque monkeys. We could extract anatomically relevant cortical brain networks that underlie different states of consciousness. The novelty of the framework lies in the adoption of a constrained linear latent variable model that provides BNAs over identifiable and disjoint ROIs, called brain networks. In line with the GNW theory of consciousness, the brain network involving the frontal, parietal, and cingular cortices plays a prominent role in identifying the level of consciousness. This new approach allows the construction of an interpretable brain decoding model that provides a unique signature of consciousness and anesthesia-induced LoC. Within the brain activity recordings, we highlight specific properties (the BNAs) that predictably differ between conscious and unconscious states and that could be used to accurately and objectively classify individuals according to their state of consciousness. A BNA-driven prediction paradigm also ensures that the selected atlas is well suited to investigate the underlying clinical question and provides a sound empirical basis for solving the optimal atlas selection problem.

A successful benchmark limited by overfitting

fMRI neuroimaging and neurophysiological maps provide unique data on brain activity, providing an excellent opportunity to build a whole-brain computational model of LoC. We present such a data-driven machine learning model, the MHA. This model scales to any dataset, and the associated hyperparameters are tuned numerically from maximum likelihood estimation. Specifically, a four-step framework generates a coherent, interpretable, and robust model of consciousness. The discovered brain networks are tailored, spatially consistent, and symmetric. The optimal number of brain networks depends on the input atlas, but a clear decision can be made by monitoring the log-likelihood. Stability analyses of resting-state networks in humans often suggest that 7 to 17 networks are appropriate [47]. In the macaque, ICA allows the detection of 11

prominent resting-state networks involving multiple levels of neural processing that show a remarkable degree of similarity to the human organization [48]. Here, following the seminal work of Monti and colleagues in humans [37], the number of optimal brain networks obtained for the three considered atlases varies from 4 to 7. Capturing the functional organization with few BNs clearly increases interpretability. The question of whether this organization is adequately captured is challenging and will be the subject of dedicated work.

As input, we used a retrospective rs-fMRI dataset acquired in macaque monkeys in different states of consciousness, where key results regarding the network organization are expected [18,19]. The potential of the proposed framework is demonstrated by modeling brain activity on this dataset. Because fMRI acquisition in monkeys is typically performed on a small number of subjects, there is a risk of overfitting. Overfitting occurs when a model becomes too complex and starts to fit the noise or random fluctuations in the training data, rather than the underlying patterns or relationships that are of interest. When working with datasets with a small number of individuals, overfitting can be particularly problematic because there may not be enough examples to capture the true distribution of the data. The small sample size can thus limit the statistical power of the study, making it more difficult to detect true effects and increasing the risk of false negatives. Here, the use of a linear latent variable model is less prone to overfitting than more complex models. A leave-one-subject out strategy is also used in the inference stage. However, it's important to note that the effectiveness of a linear model in limiting overfitting depends on the characteristics of the data and the complexity of the underlying relationships.

Increasing the size of the dataset is a common and effective strategy to reduce overfitting in machine learning. To increase the size of the dataset without additional acquisition, we are exploring the benefits of including simulated data [49,50]. The MHA model, coupled with a rich simulated repertoire of functional brain configurations, would provide a unique framework for exploring artificial LoC in silico and other potential target areas described in different theoretical frameworks (such as the II theory). Note that it is not easy to externally validate the model and subsequent findings. In fact, such retrospective data are very rare.

In conclusion, the detection of residual consciousness is still an open and challenging problem. BNAs derived from the MHA model provide reliable biomarkers of consciousness. Recent work has focused on capturing the dynamics of spatiotemporally overlapping functional networks encoded in the dFCs. Training the MHA model on the dFCs will be the subject of dedicated work. The results will be compared with those obtained by conventional dynamic FC analysis or co-activation pattern analysis such as the iCAPs [51]. Such a comparison would reinforce the unique insights provided by the MHA approach.

The parcellation strategy

It is common practice in neuroimaging to use a specific template that defines anatomical regions of the brain. The atlas serves as a common spatial landmark for analysis and interpretation of the data. In our case, it allows us to map and compare brain activity across individuals or groups. Developing a framework for selecting the optimal atlas is essential for defining the relevant ROIs in the study and improving the reproducibility and comparability of research. Indeed, to obtain accurate measurements and meaningful interpretations of the data, it is essential to choose an atlas, whether structural and/or functional, that precisely defines the ROIs suitable for the subsequent analysis. In addition, the development of a framework for selecting the optimal atlas will lay the groundwork for standardizing approaches and guidelines for atlas selection. This will also promote improved reproducibility and facilitate meta-analyses. We propose a machine learning paradigm that provides a sound basis for atlas selection given the underlying clinical question. In our analysis, the most accurate predictions of depth of anesthesia using the proposed BNAs are achieved with the CoCoMac atlas, specifically designed to study cortical regions. This is consistent with GNW theory, which suggests that the neural correlates of consciousness are predominantly cortical in nature. In a functional connectivity study, the choice of brain atlas is usually a trade-off between the characterization of brain structure and signal averaging for data reduction and noise reduction. Fine-grained brain areas can capture brain activity descriptions with more functionally-specific regions at the expense of signal-to-noise ratio loss. Fine-grained brain areas also generate high-dimensional input features that are challenging to learn in generic predictive models, a problem known as the curse of dimensionality. Increasing the dataset size will alleviate this problem, and can be achieved through the simulation paradigm proposed in the previous paragraph, or through emerging initiatives such as the PRIMatE Data Exchange [52], which provides free access to an

increasing number of fMRI recordings.

Potential applications of the proposed framework

The brain is a highly interconnected system consisting of multiple regions that communicate and interact with each other. The proposed framework rethinks the state-of-the-art data-driven strategies. It decomposes the brain signal into brain networks. This provides valuable information about the functional organization of the brain and allows the study of individual differences in brain activity. In fact, each individual has a unique pattern of brain connectivity. Characterizing these individual differences can provide insights into variations in cognitive abilities, behavior, and susceptibility to brain disorders. In addition, understanding individual differences in network connectivity may facilitate the development of personalized treatment approaches, where interventions can be tailored to target specific network dysfunctions in a given individual. The study of network-level properties offers the potential to identify specific biomarkers that can be used for diagnostic purposes, disease monitoring, or prediction of treatment outcomes. Indeed, many neurological and psychiatric disorders are characterized by alterations in brain connectivity. By decomposing the brain signal into networks, we can study how these changes manifest at the network level. By comparing network properties between healthy individuals and patients, it is possible to identify aberrant connectivity patterns associated with specific disorders. This approach can lead to a better understanding of the underlying mechanisms of the disorders and potentially help to develop diagnostic or therapeutic strategies.

Decoding consciousness

In this study, our focus is on detecting differences directly related to interactions between cortical regions by applying statistical analyses to BNAs categorized by level of consciousness. Unlike alternative methods such as PCA or ICA, BNAs derived from BNs emphasize a significant advantage inherent to the MHA model: a notable improvement in interpretability. Indeed, we indentify two main brain networks that account for the ability to decode the state of consciousness. One of them, previously described in the literature, supports the above hypothesized GNW theory. The associated BNAs effectively discriminate between awake and all anesthetic states, which is a strong finding. Additionally, to a lesser extent, these BNAs also distinguish among anesthetic states, underscoring that the use of different anesthetics may alter FC. The other is a new, unnoticed pattern. Specifically, this network is driven by the visual pathway and may be an artifact of our experimental setting, where awake monkeys had open eyes with potential visual stimulation. In the method of [19], eye position was tracked during the scans. To move forward, we plan to remove windows of saccades or increased visual activity from the awake data. Saccades may selectively generate small motion artifacts (which may have been partially preprocessed) or artificially induce neural coherence in the awake data. This could have confounded the modeling results. Overall, the MHA approach yields few tailored brain networks and associated BNAs, which promotes interpretability. The model does not rely on any biological assumption about anesthetics, and provides results that are insensitive to different anesthetics. Thus, one might assume that we are getting a general signature of consciousness, disentangled from potential markers related to a particular anesthetic effect.

Supporting information

S1 Fig. Brain state patterns. Illustrating the prevalence of different dynamic functional connectivity patterns. Unsupervised clustering of dFC matrices yields brain states: A) the within-state probability of occurrence allows the discrimination of specific brain patterns, B) the brain states (numbered 1 through 7) can be ranked according to their spatial correlation with the underlying anatomical connectivity structure, and C) the averaging of within-state saliency maps highlights different brain connectivity patterns, which are the basis of consciousness markers.

S2 Fig. Sliding windows global synchronization patterns analysis. Replication of the sliding windows synchronization patterns analysis under our experimental settings using either A) the negative

or B) the positive peaks, and considering the 3 atlases (CoCoMac, CIVMR, and DictLearn). From left to right: i) the positive/negative peaks of each time-series are detected and regrouped in a raster plot, ii) the positive/negative peaks occurrence for each sliding window are computed, and iii) the global synchronization metric as the Fano Factor (FF) distributions across acquisition conditions (dashed lines), and a gamma fit (solid lines) are estimated.

S3 Fig. Brain networks derived from each atlas. The derived brain networks (BNs) consist of sets of unique ROIs represented by their centroids for the CoCoMac (k=4), DictLearn (k=6) and CIVMR (k=7) atlases. The BNs are sorted according to the geometric criterion d_{MoC}: A) BNs that are matched in each atlas, and B) BNs from atlases that provide additional brain networks.

S4 Fig. Inferred Brain Networks (BNs) and associated BN Activities (BNAs). BNs are inferred from the MHA model using the A) CIVMR (k=7) and B) DictLearn (k=6) atlases. Pairwise statistical analysis of the associated BNAs and visualization of these BNAs against the different acquisition conditions are proposed in addition to the display of the inferred networks. These plots illustrate how clearly awake can be distinguished from all anesthetic states and, to a lesser extent, how all anesthetic states can be distinguished from each other. The legend for the p-value annotation is as follows: $** : 1.0e-3 < p \leq 1.0e-2, *** : 1.0e-4 < p \leq 1.0e-3, **** : p \leq 1.0e-4$.

S5 Fig. Inferred Brain Networks (BNs) and associated BN Activities (BNAs) when using a leave-one-subject out strategy. BNs are inferred from the MHA model using the CoCoMac (k=4) atlas when all data are available (first row) or when one subject is removed during training (second row). For comparison, a pairwise statistical analysis of the associated BNAs and a visualization of these BNAs against the different acquisition conditions are proposed.

S1 Table. Summary of the molecular targets of the anesthetics used. Information taken from [9, 11, 53]. Up" means potentiation, and "down" means inhibition. Under the heading "Potassium channels" effects are given for two-pore/inwardly rectifying/voltage-gated channels.

S2 Table. Description of the acquisition conditions across monkeys. The data were collected between July 2011 and August 2016 in five rhesus macaques (macaca mulatta), one male (monkey J) and four females (monkeys A, K, L, and R), 5 to 8 kg, 8 to 12 years, either in the awake state or under anesthesia (ketamine, propofol, or sevoflurane). Three monkeys were scanned for each arousal state (awake: monkeys A, K, and J - propofol anesthesia: monkeys K, R, and J - ketamine anesthesia: monkeys K, R, and L - sevoflurane anesthesia: monkeys L, R, and J) with the following repartition.

S3 Table. Listing of the networks inferred from the CoCoMac atlas (k=3). Listing of A) the network 1, B) the network 2, and C) the network 3. The detected GNW areas are depicted in blue, and the associated sensory areas in green.

S4 Table. The remaining networks inferred from the CoCoMac (k=4). Listing of A) the network 2, and B) the network 3. The detected GNW areas are depicted in blue, and the associated sensory areas in green.

S5 Table. The networks inferred from the CoCoMac (k=4) when using a leave-one-subject out strategy. Listing of A) the network matched to BN1, and B) the network matched with BN4. The detected GNW areas are depicted in blue, and the associated sensory areas in green. The +/- signs indicate the addition or deletion of a region in the considered BN.

Acknowledgments

We thank Alexis Amadon, Jérémy Bernrard and the NeuroSpin MRI and informatics teams for help with imaging tools, and Jean-Robert Deverre for support. We thank Marion Gay, Sébastien Mériaux, Thomas Lillin and all the members of the NeuroSpin preclinical platform for their support to the experimental work.

Author contributions

Conceptualization: B. Jarraya.
Data curation: L. Uhrig.
Formal analysis: A. Grigis, C. Gomez.
Funding acquisition: B. Jarraya.
Investigation: B. Jarraya, L. Uhrig.
Methodology: A. Grigis.
Project administration: B. Jarraya.
Resources: B. Jarraya.
Software: A. Girigs, C. Gomez.
Supervision: V. Frouin, E. Duchesnay.
Visualization: A. Grigis.
Writing – original draft preparation: A. Grigis.
Writing – Review & Editing: A. Grigis, C. Gomez, V. Frouin, E. Duchesnay, L. Uhrig, B. Jarraya.

Data and code availability

All data and codes underlying the results of this study are available at https://doi.org/10.5281/zenodo.10572216.

References

1. Gibbs FA, Davis H, Lennox WG. The electro-encephalogram in epilepsy and in conditions of impaired consciousness. Journal of Nervous and Mental Disease. 1935;34:1133–1148.

2. Gibbs FA, Gibbs EL, Lennox WG. Effect on the electro-encephalogram of certain drugs which influence nervous activity. JAMA Internal Medicine. 1937;60:154–166.

3. Loomis AL, Harvey EN, Hobart G. Cerebral states during sleep, as studied by human brain potentials. Journal of Experimental Psychology. 1937;21:127–144.

4. Purdon PL, Sampson AL, Pavone KJ, Brown EN. Clinical Electroencephalography for Anesthesiologists: Part I Background and Basic Signatures. Anesthesiology. 2015;123:937–960.

5. Mhuircheartaigh RN, Warnaby CE, Rogers R, Jbabdi S, Tracey I. Slow-Wave Activity Saturation and Thalamocortical Isolation During Propofol Anesthesia in Humans. Science Translational Medicine. 2013;5:208ra148–208ra148.

6. Akeju O, Westover MB, Pavone KJ, Sampson AL, Hartnack KE, Brown EN, et al. Effects of Sevoflurane and Propofol on Frontal Electroencephalogram Power and Coherence. Anesthesiology. 2014;121:990–998.

7. Wang K, Steyn-Ross ML, Steyn-Ross DA, Wilson MT, Sleigh JW. EEG slow-wave coherence changes in propofol-induced general anesthesia: experiment and theory. Frontiers in Systems Neuroscience. 2014;8.

8. Akeju O, Song AH, Hamilos AE, Pavone KJ, Flores FJ, Brown EN, et al. Electroencephalogram signatures of ketamine anesthesia-induced unconsciousness. Clinical Neurophysiology. 2016;127:2414–2422.

9. Rudolph U, Antkowiak B. Molecular and neuronal substrates for general anaesthetics. Nat Rev Neurosci. 2004;5(9):709–720.

10. Hentschke H, Schwarz C, Antkowiak B. Neocortex is the major target of sedative concentrations of volatile anaesthetics: strong depression of firing rates and increase of GABAA receptor-mediated inhibition. Eur J Neurosci. 2005;21(1):93–102.

11. Franks NP. General anaesthesia: from molecular targets to neuronal pathways of sleep and arousal. Nat Rev Neurosci. 2008;9(5):370–386.

12. Masamoto K, Kanno I. Anesthesia and the quantitative evaluation of neurovascular coupling. J Cereb Blood Flow Metab. 2012;32(7):1233–1247.

13. Aladj LJ, Croughwell ND, Smith LR, Reves JG. Cerebral Blood Flow Autoregulation Is Preserved During Cardiopulmonary Bypass in Isoflurane-Anesthetized Patients. Anesthesia & Analgesia. 1991;72:48–52.

14. Matta BF, Heath KJ, Tipping K, Summors AC. Direct cerebral vasodilatory effects of sevoflurane and isoflurane. Anesthesiology. 1999;91(3):677–680.

15. Hori Y, Schaeffer DJ, Gilbert KM, Hayrynen LK, Cléry JC, Gati JS, et al. Altered resting-state functional connectivity between awake and isoflurane anesthetized marmosets. Cereb Cortex. 2020;30(11):5943–5959.

16. Lv Q, Yang L, Li G, Wang Z, Shen Z, Yu W, et al. Large-Scale Persistent Network Reconfiguration Induced by Ketamine in Anesthetized Monkeys: Relevance to Mood Disorders. Biological Psychiatry. 2016;79(9):765–775.

17. Allen EA, Damaraju E, Plis SM, Erhardt EB, Eichele T, Calhoun VD. Tracking Whole-Brain Connectivity Dynamics in the Resting State. Cerebral Cortex. 2014;24(3):663–676.

18. Barttfeld P, Uhrig L, Sitt JD, Sigman M, Jarraya B, Dehaene S. Signature of consciousness in the dynamics of resting-state brain activity. Proceedings of the National Academy of Sciences. 2015;112(3):887–892.

19. Uhrig L, Sitt JD, Jacob A, Tasserie J, Barttfeld P, Dupont M, et al. Resting-state Dynamics as a Cortical Signature of Anesthesia in Monkeys. Anesthesiology. 2018;129(5):942–958.

20. Drysdale A, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nature Medicine. 2016;23:28–38.

21. Taylor JJ, Kurt HG, Anand A. Resting State Functional Connectivity Biomarkers of Treatment Response in Mood Disorders: A Review. Frontiers in Psychiatry. 2021;12.

22. Horovitz SG, Braun AR, Carr WS, Picchioni D, Balkin TJ, Fukunaga M, et al. Decoupling of the brain's default mode network during deep sleep. Proc Natl Acad Sci U S A. 2009;106(27):11376–11381.

23. Boveroux P, Vanhaudenhuyse A, Bruno MA, Noirhomme Q, Lauwick S, Luxen A, et al. Breakdown of within- and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. Anesthesiology. 2010;113(5):1038–1053.

24. Hahn G, Zamora-López G, Uhrig L, Tagliazucchi E, Laufs H, Mantini D, et al. Signature of consciousness in brain-wide synchronization patterns of monkey and human fMRI signals. NeuroImage. 2021;226:117470.

25. Mashour GA, Roelfsema P, Changeux JP, Dehaene S. Conscious processing and the global neuronal workspace hypothesis. Neuron. 2020;105(5):776–798.

26. Dehaene S, Kerszberg M, Changeux JP. A neuronal model of a global workspace in effortful cognitive tasks. Proceedings of the National Academy of Sciences. 1998;95(24):14529–14534.

27. Dehaene S, Changeux JP. Experimental and Theoretical Approaches to Conscious Processing. Neuron. 2011;70(2):200–227.

28. Tononi G. An information integration theory of consciousness. BMC Neurosci. 2004;5:42.

29. Oizumi M, Albantakis L, Tononi G. From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. PLoS Comput Biol. 2014;10(5):e1003588.

30. Bernard J B, Steven L. One, not two, neural correlates of consciousness. Trends in Cognitive Sciences. 2005;9(6):269.

31. Dehaene S, Charles L, King JR, Marti S. Toward a computational theory of conscious processing. Current Opinion in Neurobiology. 2014;25:76–84.

32. Faugeras F, Rohaut B, Weiss N, Bekinschtein TA, Galanaud D, Puybasset L, et al. Probing consciousness with event-related potentials in the vegetative state. Neurology. 2011;77(3):264–268.

33. King JR, Sitt JD, Faugeras F, Rohaut B, Karoui IE, Cohen L, et al. Information Sharing in the Brain Indexes Consciousness in Noncommunicative Patients. Current Biology. 2013;23(19):1914–1919.

34. Sitt JD, King JR, El Karoui I, Rohaut B, Faugeras F, Gramfort A, et al. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. Brain : a journal of neurology. 2014;137(Pt 8):2258–2270.

35. Uhrig L, Dehaene S, Jarraya B. A Hierarchy of Responses to Auditory Regularities in the Macaque Brain. Journal of Neuroscience. 2014;34(4):1127–1132.

36. Uhrig L, Janssen D, Dehaene S, Jarraya B. Cerebral responses to local and global auditory novelty under general anesthesia. Neuroimage. 2016;141:326–340.

37. Monti RP, Gibberd A, Roy S, Nunes M, Lorenz R, Leech R, et al. Interpretable brain age prediction using linear latent variable models of functional connectivity. PLOS ONE. 2020;15(6):1–25.

38. Tasserie J, Grigis A, Uhrig L, Dupont M, Amadon A, Jarraya B. Pypreclin: An automatic pipeline for macaque functional MRI preprocessing. NeuroImage. 2020;207:116353.

39. Bakker R, Wachtler T, Diesmann M. CoCoMac 2.0 and the future of tract-tracing databases. Frontiers in Neuroinformatics. 2012;6:30.

40. Calabrese E, Badea A, Coe CL, Lubach GR, Shi Y, Styner MA, et al. A diffusion tensor MRI atlas of the postmortem rhesus macaque brain. NeuroImage. 2015;117:408–416.

41. Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, et al. Machine learning for neuroimaging with scikit-learn. Front Neuroinform. 2014;8:14.

42. Dadi K, Abraham A, Rahim M, Thirion B, Varoquaux G. Comparing functional connectivity based predictive models across datasets. In: 2016 International Workshop on Pattern Recognition in Neuroimaging (PRNI); 2016. p. 1–4.

43. Tipping ME, Bishop CM. Probabilistic Principal Component Analysis. Journal of the Royal Statistical Society Series B. 1999;61(3):611–622.

44. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. Journal of machine learning research. 2011;12(Oct):2825–2830.

45. Rennie JD, Srebro N. Loss functions for preference levels: Regression with discrete ordered labels. In: Proceedings of the IJCAI multidisciplinary workshop on advances in preference handling. vol. 1. AAAI Press, Menlo Park, CA; 2005.

46. Pedregosa-Izquierdo F. Feature extraction and supervised learning on fMRI : from practice to theory [Theses]. Université Pierre et Marie Curie - Paris VI; 2015. Available from: https://theses.hal.science/tel-01100921.

47. Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol. 2011;106(3):1125–1165.

48. Hutchison RM, Leung LS, Mirsattari SM, Gati JS, Menon RS, Everling S. Resting-state networks in the macaque at 7T. NeuroImage. 2011;56(3):1546–1555.

49. Kringelbach ML, McIntosh AR, Ritter P, Jirsa VK, Deco G. The rediscovery of slowness: Exploring the timing of cognition. Trends Cogn Sci. 2015;19(10):616–628.

50. Deco G, Kringelbach ML, Jirsa VK, Ritter P. The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core. Sci Rep. 2017;7(1):3095.

51. Scott E, Lynn U, Aymeric G, Guylaine H, Younes F, Jordy T, et al. Signature of Consciousness in the Transient Dynamics of fMRI brain activitySignature of Consciousness in the Transient Dynamics of fMRI brain activity. In: OHBM; 2023.

52. PRIMatE Data Exchange Consortium. Accelerating the evolution of nonhuman primate neuroimaging. Neuron. 2020;105(4):600–603.

53. Alkire MT, Hudetz AG, Tononi G. Consciousness and anesthesia. Science. 2008;322(5903):876–880.