DISS. ETH NO. 24191

# NETWORK ANALYSIS AND MULTIVARIATE PATTERN RECOGNITION METHODS TO STUDY DRUG EFFECTS ON FUNCTIONAL CONNECTIVITY IN MICE BASED ON RESTING STATE fMRI

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

SYED QASIM BUKHARI

Master Erasmus Mundus CIMET
"Color in Informatics and Media Technology"

born on 03.10.1985

citizen of Pakistan

accepted on the recommendation of

Supervisor: Prof. Dr. Markus Rudin

Co-examiner 1: Prof. Dr. Klaas Enno Stephan

Co-examiner 2: Prof. Dr. Irene Tracey

2017

بسم الله الرحمن الرحيم

رب زدني علماً

اللهم انى أسالك علماً نافعا

| | |
|---|---|
| Read in the name of your Lord who created. | اقْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَق |
| Created man from a clinging substance. | خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ |
| Read, and your Lord is the most Generous. | اقْرَأْ وَرَبُّكَ الْأَكْرَم |
| Who taught by the pen. | الَّذِي عَلَّمَ بِالْقَلَم |
| Taught man that which he knew not. | عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَم |

## Acknowledgements

I would like to thank my main supervisor Prof. Markus Rudin for his unconditional support and guidance during my PhD. The freedom he gave me allowed me to do many collaborations and learn many new things at my own pace. I would also like to thank my thesis committee members, Prof. Klaas Enno Stephan and Prof. Irene Tracey for their guidance and suggestions on improving the project. I would also like to thank my colleagues at AIC, specially Dr. Aileen Schröter for her help in data acquisition and useful comments with regards to the biological part of the project, Dr. Alien Seuwen and Dr. Joanes Grandjean also for their help in data acquisition, David Bühlmann for his help in translating many of my documents in to German; Wu Wei, Andreas, Markus Vaas, Giovanna and Marco Dominietto for an unforgettable time at the ETH Hönggerberg office and the rest of the AIC colleagues for making my time at AIC very special.

I would also like to thank my other friends at ETH and UZH, Ahmed Moursya, Bledar Fazlija, Arber Fazlija, Babu Düfner, Adnan Siddique, Waiz Karim, Omar Fawzi and Adel Benlagra for making my life outside ETH an amazing experience and making me feel at home in Zurich.

I would also like to thank my family, first and foremost my mother, Shireen Gul, who was my very first teacher and made me who I am today. I would also pay my gratitude to my late father, Ahmed Bukhari, whom I lost during my PhD. He had always been a source of inspiration to me and he will always live in my heart. I would also like to thank my brother Tahir Bukhari who had always been a great friend and brother. I would also like to thank my wife, Rabia Mazhar, who cared for me when I needed and gave me her love and compassion to pass through the difficult times of PhD.

**Summary**

Resting state fMRI aims at establishing the functional relationship between two spatially distinct brain regions based on temporal correlation of the respective signals. The method has been applied to study functional connectivity (FC) across species and was found of special value for studies in mice fMRI studies due to its experimental simplicity, which allowed eliminating confounds typically observed in stimulus-evoked fMRI studies. Yet, analysis of fMRI data in animals face a number of problems. First, fMRI studies in mice typically require the use of anaesthetics, which are known to alter responses to stimuli or functional networks at rest. Proper interpretation of fMRI data collected in animals under anaesthesia animals requires investigation of the effects of these drugs on brain processing per se. Second, since fMRI analysis tools have been typically developed for processing of human fMRI data, translation to animals including mice may be challenging due to anatomical and physiological differences across species. It therefore appears appropriate to evaluate several state of the art analysis tools for human fMRI for their potential use in mice fMRI.

In the first study, we have used Dual Regression analysis Network Modelling to test its feasibility in mouse resting-state fMRI analysis to investigate effects of two commonly used anaesthetics, isoflurane and medetomidine, on rs-fMRI derived functional networks, and in particular to study to what extent anaesthesia affected the interaction within and between brain networks. The analysis revealed both similarities and specific differences in network patterns of the two groups. Under isoflurane anaesthesia, intra- and interhemispheric cortical interactions have been predominantly observed, with only minor interactions involving subcortical structures. In particular, cortico—thalamic connectivity appeared significantly attenuated in line with previous observations. In contrast, medetomidine anesthetized mice displayed significant subcortical functional connectivity including interactions between cortical and thalamic ICA components. Combining the two anaesthetic drugs at low dose resulted in network interactions that by large constituted the superposition of the interactions observed for each of the two agents alone. In conclusion, the study revealed that with suitable adaptations the DR based

network modelling can be used for analyzing mouse fMRI data and the results are comparable to those obtained with classical seed based analysis.

In a second study, we investigated whether the method was sensitive enough to detect changes in FC in mouse brain in response to varying the dose dependent effects of isoflurane using resting state fMRI. Stationary FC analysis was complemented by analysis of dynamic functional connectivity (dFC), i.e. looking for short-term changes in the interaction of brain functional networks. Stationary network analysis using FSL Nets revealed that increasing isoflurane dose led to a reduction of functional connectivity between the bilateral homotopic cortical regions as well as between cortical and thalamic areas. In addition, dFC analysis revealed a dominance of functional states (dFS) exhibiting pronounced modular structure in mice anaesthetized with a low dose of isoflurane, while at high isoflurane levels dFS showing widespread unstructured correlation displayed highest weights. This indicates that spatial segregation across brain functional networks is lost upon increasing dose of the anaesthetic drug. In conclusion, by combining the results of stationary and dynamic FC analysis of mouse resting-state fMRI data we found that increasing isoflurane levels led to loss of modular network organization, which includes the loss of strong bilateral interactions between homotopic brain areas.

In a third study, we evaluated to what extent machine learning methods could be applied for unsupervised classification of subjects according to their resting-state fMRI derived FC pattern. Features extracted from stationary as well as dynamic functional connectivity analysis derived from mice exposed to the anaesthetic isoflurane at different doses were subjected to machine learning algorithms for both support vector machines (SVM) and deep belief networks (DBN). The results show that we were able to successfully classify, i.e. assignments to group above chance level, between anaesthetic doses using features extracted from stationary and dynamic functional connectivity analysis. Not surprisingly, the classification accuracy increased when comparing extreme groups, e.g. lowest and highest dose of isoflurane. The features extracted from dFC analysis were found to be more discriminative with regard to the different anaesthetic doses than those derived from stationary FC. This illustrates the potential of using dFC features. A major limitation regarding the use of machine learning in the context of our study was small sample size (N=12 per group),

which led to an accuracy of less than 70% for most comparisons. In conclusion, classification based on machine learning tools yielded results that were clearly above chance levels though classification accuracy was likely compromised by the small size of the training data sets. Future studies are needed to assess the value of machine learning in mouse fMRI.

Finally, we applied the Random Forest (RF) classification to detect differences in the pharmacological MRI (phMRI) response of rats to treatment with an analgesic drug (buprenorphine) at two doses as compared to control (saline). RF analysis was able to identify drug effects based on differential phMRI responses in the hippocampus, amygdala, nucleus accumbens, superior colliculus, and the lateral and posterior thalamus for drug versus saline. These structures have high levels of mu opioid receptors associated to the drug response. In addition, these regions are involved in aversive signalling, which is inhibited by mu opioids. In conclusion, the results demonstrate that buprenorphine mediated phMRI responses comprise characteristic features that allow a supervised differentiation from placebo treated rats as well as the proper allocation to the respective drug dose group using the RF method.

In this thesis, network analysis methods originally developed for analysis of human fMRI data have been applied for processing of mouse fMRI data. Networks identified were found biologically meaningful as were the within and between network interactions. In particular, dFC analysis appears an attractive approach yielding insight into network interactions that appear indicative of specific condition. Classification attempts revealed that machine learning approaches work in principle, though accuracy is largely compromised by the typical small size of data sets available in animal fMRI. In this context, the establishment of open-access databases, where researches can deposit their original data, may become attractive.

**Zusammenfassung**

Funktionelles MRI im Ruhezustand zielt darauf ab, die funktionelle Beziehung zwischen zwei räumlich getrennten Hirnregionen auf Grundlage der zeitlichen Korrelation der jeweiligen Signale zu etablieren. Das Verfahren wurde angewendet, um die funktionelle Konnektivität (FK) in verschiedenen Spezies zu untersuchen und wurde aufgrund der guten experimentellen Durchführbarkeit für Studien in Mäuse als besonders wertvoll befunden, da typische experimentelle Limitationen welche in stimulus-evozierten fMRI-Studien auftreten überwunden werden können. Allerdings steht die Analyse der fMRI-Daten bei Tieren vor einer Reihe von Problemen. Zum einen erfordern fMRI-Studien an Mäusen typischerweise die Verwendung von Anästhetika, von denen bekannt ist, dass sie Reaktionen auf Stimuli oder funktionelle Netzwerke in Ruhe beeinflussen. Daher erfordert die korrekte Interpretation von fMRI-Daten, die bei Tieren unter Anästhesie gesammelt wurden, die Untersuchung der Wechselwirkungen dieser Medikamente auf die Gehirnverarbeitung per se. Zum anderen sind fMRI-Analysetools typischerweise für die Verarbeitung von menschlichen fMRI-Daten entwickelt worden, daher kann die Translation auf Tiere, einschließlich Mäusen, aufgrund anatomischer und physiologischer Unterschiede zwischen den Spezies eine Herausforderung darstellen. Es erscheint daher angebracht, Analysewerkzeuge die für menschliche fMRI-Daten entwickelt wurden, auf ihre Eignung zur Analyse von Maus fMRI-Daten zu evaluieren.

In einer ersten Studie haben wir „Dual Regression Analysis Network Modeling" verwendet, um ihre Eignung zur Analyse von Maus fMRI-Daten zu untersuchen, um die Effekte von zwei häufig verwendeten Anästhetika, Isofluran und Medetomidin, auf funktionelle Netzwerke zu untersuchen. Insbesondere wurde untersucht, inwiefern die Anästhesie die Interaktion innerhalb und zwischen verschiedenen Hirnnetzwerken beeinträchtigt. Die Analyse ergab sowohl Ähnlichkeiten als auch spezifische Unterschiede in den Netzwerkmustern der beiden Gruppen. Unter Isoflurananästhesie wurden überwiegend intra- und interhemisphärische kortikale Wechselwirkungen beobachtet, während Interaktionen mit subkortikalen Strukturen nur schwach ausgeprägt waren. Insbesondere wurde die kortiko-thalamische Konnektivität deutlich gedämpft, was im Einklang mit vorhergehenden Studien steht. Im Gegensatz dazu zeigten Mäuse unter Medetomidinanästhesie signifikante funktionelle Interaktionen mit subkortikalen Strukturen, einschließlich der Wechselwirkung zwischen kortikalen

und thalamischen funktionell unabhängigen Komponenten. Die Kombination beider Anästhesien in geringer Dosis führte zu Netzwerkinteraktionen, welche die Überlagerung der isoliert beobachteten Wechselwirkungen darstellt. Zusammenfassend ergab die Studie, dass mit geeigneten Anpassungen die DR-basierte Netzwerkmodellierung zur Analyse von Maus fMRI-Daten verwendet werden kann und die Ergebnisse vergleichbar sind mit denen der klassischen „Seed-based"-Analyse.

In einer zweiten Studie untersuchten wir, ob die Methode sensitiv genug ist, um Veränderungen der FK im Maushirn als Antwort auf Veränderungen der dosisabhängigen Wirkung von Isofluran nachzuweisen. Die statische FK-Analyse wurde durch Analyse der dynamischen funktionellen Konnektivität (dFK) ergänzt, d.h. um dynamische Veränderungen in der Interaktion von funktionellen Netzwerken über die Zeit zu finden. Die statische Netzwerkanalyse unter Verwendung von FSL Nets ergab, dass eine erhöhte Isofluran-Dosis zu einer Verminderung der funktionellen Konnektivität zwischen homotopen, kortikalen Regionen sowie zwischen kortikalen und thalamischen Regionen führte. Zusätzlich zeigte die dFC-Analyse eine Dominanz dynamisch funktionelle Zustände (dFZ), die eine ausgeprägte modulare Struktur in Mäusen mit einer geringen Isofluran-Dosis zeigten während bei hohen Isofluran-Dosen die am stärksten gewichteten dFZ eine weitgehend unstrukturierte Korrelation zeigten. Dies zeigt, dass die räumliche Segregation in funktionellen Netzwerken des Gehirns bei steigender Dosis des Anästhetikums verloren geht. Durch die Kombination der Ergebnisse einer statischen und dynamischen FK-Analyse von Maus fMRI-Daten im Ruhezustand lässt sich zusammenfassend feststellen, dass erhöhte Isofluran-Dosen zum Verlust der modularen Netzwerkorganisation führen, was den Verlust starker bilateraler Wechselwirkungen zwischen homotopischen Hirnarealen mit sich bringt.

In einer dritten Studie wurde untersucht, inwiefern maschinelle Lernmethoden für die automatische Klassifikation von Probanden anhand von Ruhezustand fMRI abgeleiteter FK-Muster angewendet werden können. Merkmale welche aus statisch sowie dynamisch funktioneller Konnektivitätsanalyse in Mäusen abgeleitet wurden, die dem Anästhetikum Isofluran in verschiedenen Dosen ausgesetzt waren,
wurden mittels maschinellen Lernalgorithmen wie Support-Vektor-Maschinen (SVM) und „Deep Belief Networks" (DBN) in die verschiedenen Gruppen klassifiziert. Die

Ergebnisse zeigten dass wir in der Lage sind anhand dieser Daten Subjekte den unterschiedlichen Anästhesie-Gruppen zuzuweisen mit einer Erfolgsquote über dem Zufallsniveau. Es überraschte nicht, daß die Klassifikationsgenauigkeit stieg beim Vergleich von Extremgruppen, z.B. Niedrigste und höchste Dosis von Isofluran. Dabei erwiesen sich Merkmale der dFK-Analyse als robuster für die Klassifizierung in die verschiedenen Gruppen als jene, die von der statischen FK-Analyse stammen. Dies illustriert das Potenzial der Verwendung von dFK-Merkmalen. Eine wesentliche Einschränkung hinsichtlich des Einsatzes von maschinellem Lernen im Rahmen unserer Studie war eine geringe Stichprobengröße (N = 12 pro Gruppe), wodurch die Genauigkeit der Klassifikationen unter 70% lag für die meisten Gruppenvergleiche. Zusammenfassend kann gesagt werden dass Klassifizierung mittels maschinellen Lernalgorithmen Resultate klar über dem Zufallsniveau lieferte, wobei die Genauigkeit der Klassifizierung sicherlich durch die kleine Gruppengrösse des Trainings-Datensatzes kompromittiert wurde. Demzufolge sind weitere Studien nötig um den Wert maschineller Lernalgorithmen zur Analyse von Maus fMRI-Daten zu evaluieren.

# Table of Contents

**Aims of the thesis**

Resting state connectivity has gained immense popularity for studying brain networks using fMRI. However fMRI analysis tools are typically developed for humans. Due to inter-species differences, increased susceptibility to magnetic inhomogeneity at large magnetic field strengths leading to images distortion and intra-voxel dephasing (signal voids), as well as differences in basic physiology (respiration frequency, heart rate, blood pressure), the application of these analysis tools to analyse small animal fMRI data is not straightforward and requires adaptations. Similarly, machine-learning tools have been largely restricted to application in human fMRI data. Furthermore, a major difference between human and animal (rodent) fMRI is the use of anaesthesia, which is common in small animal fMRI studies despite their known intrinsic effects on brain activity, but rarely used in human studies, unless the purpose of the study is to investigate the effects of an anaesthetic drug on brain function/networks. In animal studies, anaesthesia is essential for immobilizing the animal during data acquisition. It is therefore important to assess and understand the alterations in brain activity patterns/network induced by the anaesthetic drug. We aimed to relate functional changes as derived from MRI to clinical, behavioural measures of anaesthesia depth such as loss of sensation, analgesia, muscle relaxation, and loss of consciousness, as a function of anaesthesia depth (dose of the anaesthetic drug). A prerequisite for such studies are reliable tools for analysing fMRI, and in particular resting-state fMRI data, that yield semi-quantitative information on brain networks and changes in interactions strength when types or dose of the anaesthetic agent is altered.

The aim of this thesis was to evaluate the suitability of several state of the art fMRI analysis tools originally developed for the analysis of human fMRI data for processing of resting-state fMRI data of mice exposed to different anaesthesia regimens. Specifically, we applied

1) dual regression and network analysis for identifying major brain network modules and studying within and between network interactions

2) dynamic functional connectivity analysis for probing the hidden information on network interactions not apparent from conventional stationary resting-state fMRI analysis, and

3) machine learning tools in classifying resting-sate fMRI data obtained for different anaesthesia regimes.

## 1. Introduction

In the past few decades, neuroscience community has made rapid scientific progress in understanding the brain function. Before the advent of non-invasive neuroimaging techniques, neuropsychology played a major role in increasing our understanding of the brain function by examining patients with known brain injuries. Several patients have been reported in the literature that suffered specific kinds of brain damage due to lesions or injury. The disabilities in such patients gave indications of the specific cognitive functions of the damaged brain regions. For example patient HM [1; 2] whose hippocampi, parahippocampal cortices, entorhinal cortices, piriform cortices, and amygdalae were surgically resected in an attempt to cure his epilepsy, helped neuroscientists understanding how these particular areas in the brain may be linked in the memory[3] formation process. Similarly patient Tan had a lesion affecting his frontal part of the brain, specifically posterior inferior frontal gyrus. The clinical phenotype associated to this specific lesion was loss of speaking ability also called as the cognitive deficit of aphasia, which allowed linking this specific cognitive ability of speaking to the brain regions affected [4; 5]. This also proved that the speech function in the brain in localised. However despite such interesting and important results, this approach is rather limited since specific brain deficits are rarely found without other deficits. This limitation motivated the advent the non-invasive methods to determine brain functions of specific brain areas.

Communication between human brain nerve cells were first recorded at the scalp in 1924 through a technique called Electroencephalography (EEG). Though the EEG signals measured at the scalp are the result of the activity from a large number of neurons, EEG allows distinguishing between different frequency patterns with precision. Delta, theta, alpha and beta rhythms in EEG have characteristic amplitude, size and frequency. EEG also allows high temporal resolution in the order of 1 ms, however the EEG has a poor spatial resolution due to the signal collected at the scalp (around 5 to 9 cm) [6; 7; 8]. This made EEG an ideal technique to investigate the fast occurring neuronal activity in specific brain areas. However for understanding the interaction between different brain regions, EEG is not an optimal tool due to its lack of spatial resolution [8].

It is beyond the scope of the thesis to review all methods suggested or applied to derive the functional topology of the human brain. Instead we will focus on

approaches that are related to functional neuroimaging[9]. Blood flow in the brain regions is shown to relate with the brain function by observing the local pulsations of the subjects under a task [9]. The first quantitative method for measuring whole brain blood flow and metabolism in humans was developed in 1946 when Kety and Schmidt measured the arterial-venous concentration of nitrous oxide, to quantitatively measure cerebral blood flow (CBF) [10; 11]. The advent of X ray Computed Tomography (CT) by Hounsfield marked a major breakthrough in neuroimaging. CT quickly became popular for the identification of problematic brain tissues and replaced dangerous clinical practices for example pneumoencephalography. However CT was limited to image the brain anatomy only and was not useful for understanding the brain functional principles. Nevertheless, the basic principle of CT image reconstruction has been transferred to other modalities that are more sensitive to changes in metabolism and blood flow (providing energy substrates and removing waste products) intrinsically associated with neural activity. Among those positron emission tomography (PET), first suggested in 1951 [9; 12] is attractive as it uses radionuclides such as 18F or 11C that could be readily introduced into metabolic precursors and might thus provide relevant information on local metabolic activity. Among the precursors, [18F]-2-fluoro-2-deoxy-glucose (FDG), proposed as a measure of cerebral metabolic rate of glucose (CMRglu), takes an important role, and was for a long time the only PET tracer approved for clinical use [13]. FDG is based on the method proposed by Sokoloff et al [9; 14], who has used 14C labelled 2-deoxyglucose (DG) to measure local CMRglu using autoradiographic techniques. They found increased DG activity in brain areas that displayed increased activity. This approach was translated in vivo, and FDG PET has become a standard method to assess the glucose utilization in various tissues including the brain. Validation studies in animals have revealed that the approach is sensitive to changes in neural activity [15]. A shortcoming of FDG-PET is slow kinetics, i.e. changes in activity have to be analysed over a period of 40 to 45 min. Therefore measurements of activity induced changes in cerebral blood flow using 15O-H2O, was suggested as an alternative. Here the limitations are, that an on-site cyclotron is required for such experiments as the half-life of 15O is only of the order of 2 min. It is not surprising the PET based blood flow measurements were to be replaced by alternative methods not dependent on an exogenous tracer, in particular magnetic resonance imaging based measurements.

In 1973, Paul Lauterbur recorded the first nuclear magnetic resonance image ((N)MRI) exploiting the intrinsic magnetic properties of the hydrogen nuclei in the water molecule, which constitute 70 to 80% of the tissue composition. Four years later, in 1977, the first human MRI scan was acquired. However, MRI techniques during those years were rather limited and mechanisms determining image contrast mechanisms hardly investigated. As a result, only anatomical information on the brain could be recorded though it was readily recognized that the multiparametric dependence of image contrast rendered the method rather versatile.

In 1990 Ogawa et al. [16] observed that the signal intensity in MR images depended on the degree of blood oxygenation (blood oxygenation level dependent (BOLD) contrast). While oxygenized haemoglobin is diamagnetic, deoxygenized haemoglobin is paramagnetic and hence an intrinsic contrast agent. Its efficiency depends on the ratio of oxy- versus deoxyhaemoglobin, i.e. on the oxygen saturation of blood. This rapidly led to the development of functional MRI (fMRI), which is based on the assumption that the increased metabolic activity of the activated brain areas should lead to alterations in BOLD contrast. In 1992, three independent research groups demonstrated proof of principle: Kwong et al [17] and Ogawa et al [16] revealed specific signal changes in the visual cortex upon visual stimulation and Bandettini and colleagues in motor cortex during a finger tapping task[18]. It took several years until useful quantitative models explaining the changes in MRI contrast were first suggested [19; 20], though many issue still remain and need further clarification.

With the introduction of fMRI neuroscience community received a powerful tool for studying brain function under a wide range of conditions. An important advantage of BOLD fMRI is the fact that the method is based on an intrinsic contrast, and does not depend on the administration of contrast agent as in MRI bolus tracking, or radioactive compounds, as used with PET.

In the following we will focus our discussion on fMRI techniques. Before doing so below we give a brief summary of the principle of MRI and its progress in the past years.

## 1.1 Basic Principle of MRI

The principle of MRI is based on the magnetic moment associated with the motion of charged particles. MRI uses the principle that each nucleus with an odd number of protons and/or neutrons has a particular nuclear spin, to which a magnetic moment is associated that can be polarised by applying the magnetic field. Our body is mainly comprised of water, and therefore the most abundant nuclei in our bodies are hydrogen nuclei, with a nuclear spin of I = ½. According to quantum mechanics theory, such a system exists in a state, which constitutes a linear combination of the two spin eigenstates $m_I$ = ½ or $m_I$ = -½. Under the effect of an externally applied magnetic field $B_0$, the spins of the hydrogen nuclei will align either parallel or antiparallel to the direction of the magnetic field with a slight predominance of the $m_I$ = ½ state (parallel to the field), which is energetically slightly more favourable. This leads to a net polarization of the sample corresponding to the vector sum of all spin vectors. The energy difference between the parallel or anti parallel state is given by

$$\Delta E = \gamma_H \cdot \hbar \cdot B_0$$

where $\gamma_H$ is the gyromagnetic ratio, $\hbar$ is the Dirac constant ($\hbar$ = h/ 2π) and $B_0$ is the magnetic field strength. Incase of hydrogen, $\gamma_H$= 42.6 MHz/Tesla

The frequency at which resonance is achieved is called Larmor frequency and is given by

$$\omega_0 = \gamma_H \cdot B_0$$

which for typical magnetic field strength of several Teslas is in the radiofrequency domain.

A time dependent radiofrequency magnetic field ($B_1$) with components perpendicular to the main magnetic field is applied to deflect the net magnetization from its equilibrium position and generates magnetization components transverse to $B_0$, referred to as transverse magnetization. . The transverse magnetization vector precesses around the main field $B_0$ at the Larmor frequency.

A receiver coil measures the voltage induced by the precessing transverse component of the net magnetization the induced voltage constituting the MRI signal. Once the electromagnetic field $B_1$ is turned off by turning off the radiofrequency radiation, the nuclei return to their original state. The signal behaviour is captured by two relaxation

processes: Longitudinal magnetisation $R_1=1/T_1$ ($T_1$ being the longitudinal relaxation time) is the rate at which the average magnetization returns to its original state aligned along the magnetic field, and transverse magnetisation $R_2=1/T_2$ ($T_2$ being the longitudinal relaxation time) is the rate at which transverse components vanishes due to loss of phase coherence. Relaxation parameters are the most important source of contrast among tissues as they depend on both tissue composition and microstructure.

We know that precession speed depends on the gyromagnetic ratio and the strength of the magnetic field. However even within a tissue, each proton experiences a different magnetic field strength and therefore precesses at slightly different rate, which results in getting them out of synchronisation. This is called as dephasing and the rate of dephasing is termed as T2*, process is called as free induction decay or FID. Similar to T1 recovery, T2* decay is also an exponential process. It is also important to realise the difference between T2 and T2*. While T2 is determined by the interactions of neighbouring spins, which is of stochastic nature due to Brownian motion, T2* is in addition governed by variations in local magnetic susceptibility, which is an intrinsic characteristic of tissue and deterministic in nature. As a result T2*≤T2.

### 1.1.1 Spatial encoding in MRI

Spatial encoding in MRI is performed through frequency and phase encoding. As the resonance frequency is directly dependent on the magnetic field $B_0$, changing the value of $B_0$ can alter it. This is achieved by applying a magnetic field gradient, say in the x direction, which renders the resonance frequency dependent on the location x, i.e.

$$\omega(x) = \gamma_H \cdot (B_0 + G_x \cdot x)$$

Hence the location $x$ is encoded by the frequency value $\omega(x)$. Encoding in the second dimension (and eventually third dimension for full 3D imaging) has to be separated in time for proper allocation of frequencies in two (three) dimensions. The most frequently applied approach, spin-warp imaging, uses a constant encoding time $t_1$ in order to keep effects due to relaxation constant. As a result of phase encoding the signals at the beginning of the readout period carry a signal phase that depends on its location in the phase encoding direction (say y-direction), i.e.

$$\varphi(y; i) = \gamma_H \cdot (B_0 + G_y(i) \cdot y) \cdot t_1$$

As a single phase value is not sufficient to derive the frequency information required for spatial encoding in the second dimension, the acquisition has to be repeated using a different gradient value $G_y(i + 1)$, yielding a phase values $\varphi(y; i + 1)$. For full data domain coverage in the phase-encoding direction, the amplitude of the applied magnetic field gradient has to be incremented from $-G_{y,max}$ to $+G_{y,max}$ in steps $\delta G$, the values of which have to be chosen to be in line with the sampling theorem and the desired nominal resolution.

Finally, a cross sectional slice is selected by combining frequency encoding with frequency selective excitation. By simultaneous application of a frequency selective radiofrequency pulse $\left(\omega_c - \frac{\Delta\omega}{2} < \omega < \omega_c + \frac{\Delta\omega}{2}\right)$ and a magnetic field gradient perpendicular to the slice selected (slice selection gradient), only the magnetic nuclei with resonance frequencies in this frequency domain, will contribute to the signal. Here, $\omega_c$ defines the central frequency of the slice and $\Delta\omega$ the slice width.

Below we briefly discuss the common experimental paradigms used for image acquisition in MR systems.

### 1.1.2 Spin Echoes

The FID is dependent on T2* and thus susceptible to the magnetic field inhomogeneity. Hahn [21] introduced the so-called spin-echo sequence, which consists of an excitation pulse (typically 90°) followed by a refocusing pulse (180°) applied after a time delay TE/2. This then leads to an spin echo signal that occurs exactly at the time TE, the echo time.

This can be rationalized considering three stages. The 90° pulse flips the magnetization from the z-axis (direction of $B_0$) into the transverse plane. Following the pulse, the individual magnetization vectors precess around the static magnetic field at slightly different frequencies due to difference in local magnetic susceptibility. Hence they will get out of phase and the overall signal decays (FID). After a time TE/2 an 180° refocusing pulse is applied, which mirrors the magnetization vectors in the transverse plane (for example from $M_{xy}(x)$ to $M_{xy}(-x)$ if the pulse is applied around the y-axis. Following this pulse, the vectors precess again around the static field, and the magnetization vectors that have experienced the largest phase shift during the first

precession period will catch up with the slower precessing vectors for a spin echo that occurs exactly at TE. While this experiment refocuses static variations in magnetic susceptibility it cannot account for any stochastic processes due to molecular motion. As a result the echo amplitude is reduced by a factor $\exp\left(-\frac{TE}{T2}\right)$, i.e the signal is determined by the transverse relaxation time T2. Combining spin-echo acquisition with the previously described encoding principles will generate so-called T2-weighted images.

### 1.1.3 Gradient Echoes

An echo can also be generated by gradient reversal. Basically, the transverse magnetization is dephased during a period $t_1$ by applying a magnetic field gradient $G_1$ in the readout direction, leading to a phase shift $\varphi_1 = \gamma_H \cdot (\Delta B_0(x) + G_1 \cdot x) \cdot t_1$, where $\Delta B_0(x)$ accounts for the deviation of the static magnetic field from its nominal value due to local differences in magnetic susceptibility. During the readout phase the gradient $G_r$ is applied and the magnetization vectors experience a phase shift $\varphi_2 = \gamma_H \cdot (\Delta B_0(x) + G_r \cdot x) \cdot t_2$, hence the total phase shift experienced is $\varphi = \varphi_1 + \varphi_2$. The values becomes minimal when the condition $G_1 \cdot t_1 = G_2 \cdot t_2$ is fulfilled (gradient-recalled echo). Note that even under these conditions, there is a remaining phase shift of $\varphi_1 = \gamma_H \cdot \Delta B_0(x) \cdot (t_1 + t_2) = \gamma_H \cdot \Delta B_0(x) \cdot TE$ , hence the echo amplitude is reduced by a factor determined by the local magnetic susceptibility and therefore retains T2* dephasing. Gradient echoes constitute building blocks of fast MRI sequences, in fast low angle shot MRI (FLASH), the repetition delay between subsequent excitation pulses is chosen such that $TR \ll T1$. To account for incomplete relaxation, excitation pulse angles are reduced to values $\alpha < 90^o$. The maximal signal is obtained at the so-called Ernst angle. Very fast acquisition protocols are based on multiple gradient reversals to generate an echo train, with each echo individually phase encoded. This echo planar imaging experiment, introduced as early as 1977 by Mansfield [22], allows the acquisition of complete image with a single excitation pulse, and constitutes the backbone of most fMRI protocols currently in use.

## 1.2. Functional magnetic resonance imaging

### 1.2.1 The BOLD effects and principle of fMRI

Functional MRI aims to identify the brain activity on different parts of the brain repeatedly simultaneously, thus producing a method to observe brain activity over time. In 1990, Ogawa and colleagues [16] performed an experiment to demonstrate that in vivo changes in blood oxygenation can be detected using MRI using the fact that oxygenated hemoglobin is diamagnetic while deoxygenated hemoglobin is paramagnetic and acts as an intrinsic contrast agent. Hence, the overall effect on the MRI signal depends on the ratio oxy-/deoxy-hemoglobin. Active neurons need energy, which is supplied in the form of nutrients and oxygen from blood. This leads to a local increase of cerebral blood flow, which is larger than the increase in metabolic way due to the fact that extraction efficiency decreases when flow rates are higher. As a result, the venous blood in the active brain area is better oxygenated which leads to an increase in the relaxation time $T_2^*$ (and $T_2$), which translates into a net signal increase when using gradient-echo (spin-echo) pulse sequences. The fMRI signal is hence a hemodynamic signal that depends on the integrity of neurovascular coupling.

Though the temporal resolution of fMRI is limited by the hemodynamic response function, it is especially useful for analysing whole brain connectivity analysis, since it provides whole brain coverage. In classical fMRI experiments neural activity is triggered by an event, for example a cognitive task or a sensory input. The experimental paradigm in which BOLD response is determined during a particular task is called as task-based-fMRI. Under these conditions, activated brain areas are identified by comparing BOLD signal intensity during the activated to that at baseline conditions.

### 1.2.2 Resting state fMRI

Resting state functional connectivity analysis is a method to understand the function of brain at rest in the absence of any explicit task. Biswal and colleagues found [23] that this spontaneous brain activity present is reflected in the transient fluctuation of the BOLD baseline signal, the so-called resting state fMRI (rs-fMRI) signal. Analyses of the temporal coherence of these fluctuations allows identifying brain regions that display a high degree of synchronicity, which therefore are believed to be functionally connected. The concept of rs fMRI was first introduced by Biswal et al. [23; 24],

however did not gain popularity until Raichle and coworkers identified the so-called default mode network (DMN) [25; 26], consisting of brain areas that are consistently active and highly correlated with each other during rest. Regions included in the DMN have been known to be involved in different functions including self-awareness and theory of mind, etc. This includes posterior cingulate cortex (PCC), precuneus, medial prefrontal cortex and angular gyrus. Typically, during a task DMN regions display anti-correlation with task related regions, in which activity, and hence the BOLD signal, increases. In the meantime, rs-fMRI has become an established tool in clinical and experimental neuroscience and has been consistently applied across species for network identification. The approach is of particular interest for examining functional connectivity and brain networks in order to find alterations in brain connectivity in neurological or psychiatric disorders.

Figure 1 shows the difference between resting state fMRI (a) and task-based fMRI (b). Figure 1(a) shows the time series extracted from a single voxel, a measure for the spontaneous activity in this voxel, which is then correlated with the time series extracted from other regions. The value of the correlation coefficient is a measure for the degree of connection between two voxels (regions-of-interest; ROIs). Figure 1(b) shows the example of a task based fMRI data, where the task is represented by "ON", and baseline or no-task is represented by 'OFF' state. The activations in the brain regions associated with the task are calculated by fitting a general linear model (GLM) on the ON – OFF (task-baseline) states, that makes inferences about the effects of interest by decomposing the data in to effects and error and performs statistics using estimates of those effects and error.
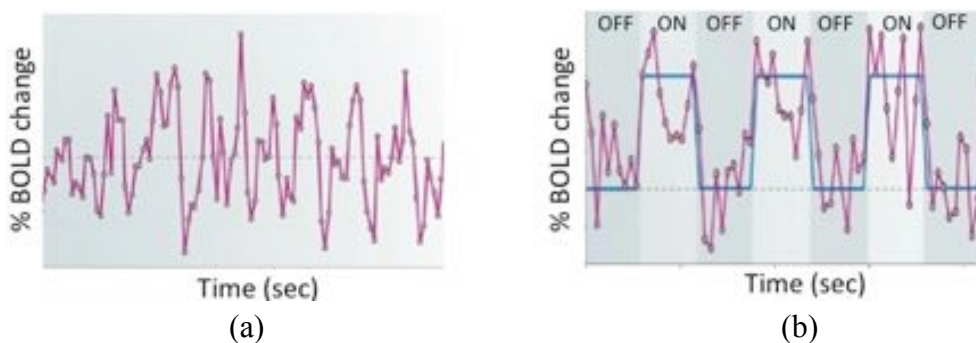


(a)  (b)

Figure 1: Resting-state fMRI and task based fMRI. (a) BOLD signal intensity as a function of time from a single voxel in absence of a task. Such temporal profiles are compared between pairs of voxels (ROIs) across the brain. The value of the correlation

coefficient is a measure for the degree of functional connection between these voxels (ROIs). (b) Task-based fMRI using an experimental paradigm comprising two states (on/off). The BOLD intensity extracted from a specific ROI is plotted as a function of time. Using the theoretical paradigm (eventually convolved with the hemodynamic response function; see text) as a correlator, activated areas may be identified on the basis of the correlation coefficients. Adapted from Fox et al. [26].

As already mentioned, fMRI measures changes in blood oxygenation levels and hence is an indirect measure of neuronal activity. As a consequence, the BOLD signal is influenced by physiological parameters like heart rate, blood pressure, breathing, baseline blood oxygenation, fluctuations of which may affect the baseline fluctuations in the BOLD signal. Because of these potential confounds (noise sources), the concept of analyzing the synchronicity (and amplitudes) of BOLD fluctuations at 'rest' has been criticized [27; 28; 29]. Nevertheless, functional, or rather metabolic connectivity has also been demonstrated on the basis of measurements of glucose metabolism using PET, a readout that is independent of the integrity and stability of neurovascular coupling. Electroencephalographic (EEG) and magnetoencephalographic (MEG) recordings have confirmed networks inferred from rs-fMRI data, indicating a neuronal basis of these signals. Resting state fMRI has also been criticized for using the term "rest", because the mammalian brain is continuously functioning and is never truly at rest. The term paradigm-free would be probably more appropriate.

There are many potential clinical applications of functional connectivity analysis at rest, for example for studying brain plasticity in response to physiological or pathological alterations. Profound changes in functional networks have been reported for patients suffering from Alzheimer's disease and risk populations [30; 31; 32; 33], psychiatric disorders [34; 35], chronic pain [36], etc. Resting state fMRI is also attractive for animal studies for both mechanistic purposes and as models of human disease. Since animals are typically anesthetized while imaged, rs-fMRI provides a natural way to explore the brain networks of these animals without disturbing the intrinsic functional connectivity. This method is also of great value to explore differences in brain pathologies and other cognitive disorders.

## 1.3. Analysis methods for rs-fMRI

### 1.3.1 Seed based correlation analysis

Seed based correlation analysis or ROI analysis is the simplest of the fMRI analysis methods. It is based on the principle of cross-correlation and is implemented by estimating the Pearson correlation coefficient between two time series vectors. These time series vectors may be either generated from any voxel in the brain or by averaging the time series from an anatomically constrained ROI. Figure 2 shows the seed in the brain region and its functionally connected regions as found through correlation analysis.
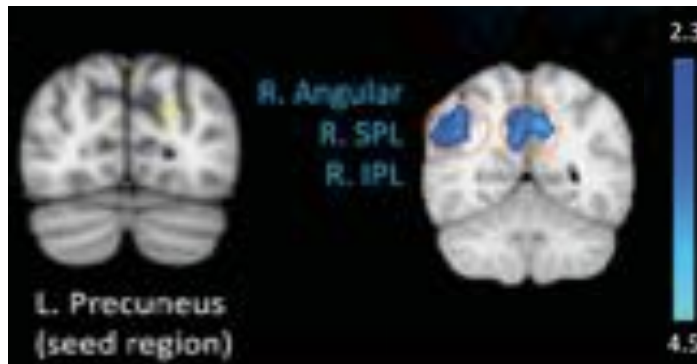


Figure 2: Figure showing the z correlation in Angular, SPL and IPL when seed (in yellow) was placed in the left precuneus cortex. The activations (in blue) indicate the connectivity in the regions with respect to the seed as indicated in yellow. Adapted from Martucci et al [37]

In order to compare across the population, the Pearson correlation values are usually normalized using Fisher's z score transformations. Z-scores transformation converts the correlation values $r(v)$ into a normal distribution

$$z(v) = 0.5 \ln(\frac{1+r(v)}{1-r(v)})$$

To assess the significance of the correlations between each pair of regions in each seed, statistics are performed on the z-transformed correlation coefficients.

Figure 3 illustrates the individual processing steps in the ROI based correlation analysis. Instead of defining individual ROIs, whole brain parcellation can be performed to obtain whole brain connectivity matrix.
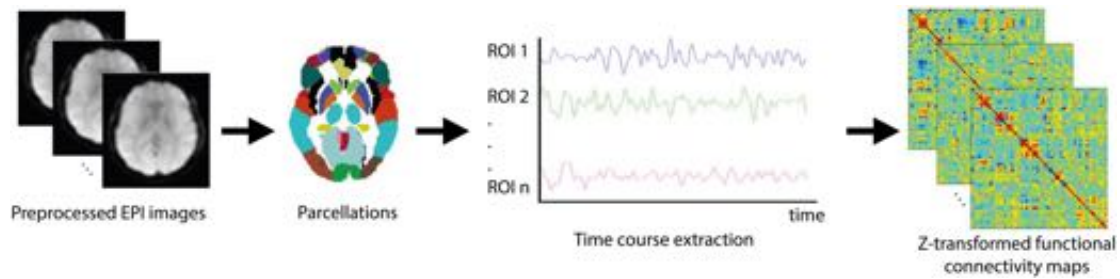
Figure 3: Processing steps for ROI based correlation based on whole brain parcellation. From left to right: raw data set consisting of $N$ sequentially acquired image volumes; definition of $n$ ROIs; extraction of time series for each of the $n$-ROIs; functional $n \times n$ connectivity maps. Adapted from Wen and Hsieh [38]

Initially, seed based correlation analysis has been widely used for analysing rs-fMRI data. Strong functional connectivity between the homotopic regions across the two brain hemispheres has been observed consistently in the normal population. In patients, impaired functional connectivity between distinct brain ROIs has been reported [39; 40; 41; 42]. Analysis of the ROIs affected may shine light on the underlying mechanisms leading to the symptoms of the brain disease.

There has also been considerable criticism in the literature on use of pre-defined ROIs. If the ROIs have been taken from an atlas, then they disregard individual variability among the population. For example, the precise size and location of subdivisions of thalamus of an atlas might be different from the study group due to pathology or age. If ROIs are defined separately for each study then there is a risk of including operator's bias in ROI definitions. These critics encourage the use of data driven approaches to identify spatial maps by analysing the specific data independent of operator bias.

### 1.3.2 ICA analysis

ICA analysis has also been widely applied in the fMRI data analysis[43; 44; 45; 46]. The theory of the ICA analysis is that it recovers independent source signals from mixtures with unknown mixing coefficients [47]. In case of fMRI, ICA can be applied either in the spatial or temporal domain to the fMRI data, which are represented as a $n_v \times n_t$ matrix consisting of $n_v$ rows ($n_v$ number of voxels) and $n_t$ columns ($n_t$ number of sample time points). In spatial ICA it is assumed that the $n_v$ columns of the

fMRI data matrix are statistically independent processes, whereas in temporal ICA the $n_v$ rows of *fMRI data matrix* are considered independent. Spatial versus temporal independence has been thoroughly discussed by Friston and Calhoun [48; 49; 50].

ICA analysis can be used to remove noise from the signal, as well as to construct group IC maps that may be used as an alternative to defining ROIs. [43; 51; 52] described a method to reduce noise using ICA analysis. There are several methods in the literature that aim at determining spatial maps from fMRI data. [43; 53] described a method using probabilistic ICA that estimates spatial IC maps from a group of fMRI data that has been including in FSL's MELODIC package. [43; 44; 53] discuss this approach in detail. fMRI data from different subjects and all groups are input in a concatenated fashion in to the ICA algorithm, which then estimates the spatial IC maps that can be used for further analysis. The advantage of using ICA analysis is that it determines components comprising of activation clusters or attributed to artifacts without any explicit time series model being specified [43].

### 1.3.3 Dual regression

Dual regression is a state-of-the-art method for identifying brain activation patterns from resting state fMRI (rsfMRI) data. As rs-fMRI does not involve an explicit experimental paradigm, it is not possible to relate brain activity to a design-based model function using e.g. a GLM approach.

Dual regression generates subject-specific spatial maps from the set of spatial maps in the group-average analysis [54; 55]. The goal of dual regression analysis is to derive subject-specific networks corresponding to ICA group components. The algorithm comprise of three steps. First, the concatenated multiple fMRI data sets are decomposed into IC maps by applying ICA analysis. The model order of the ICA analysis is estimated using the Laplace approximation to the Bayesian evidence for a probabilistic principal component model. In mathematical terms we can write [56],

$$X = A \cdot S + E \, , \tag{1}$$

with the matrix $X = (x_{ij})$ representing the group averaged fMRI volumes recorded at time points $1 \leq i \leq T$ and $j$ indicating the individual voxels ($1 \leq j \leq N_v$), the rows of

14

matrix $A = (a_{ip})$ indicate the contributions of the $1 \leq p \leq P$ independent components to the fMRI BOLD volume at time point $i$, the rows of matrix $S = (s_{pj})$ each spatial IC, and the matrix $E = (\epsilon_{ij})$ the noise contribution.

The second step is to perform the spatial regression to obtain subject specific time series. The full set of group-ICA spatial maps ($S$) is used in a linear model fit (spatial regression) against the individual fMRI data sets ($X_k$), resulting in matrices describing temporal dynamics for each component and subject ($A_k$). In other words, group ICA maps are used as spatial regressor in order to find the time series associated with the voxels in that map. This can be formulated as

$$\widetilde{X}_k = \widetilde{S} \cdot \widetilde{A}_k + \widetilde{E}_k \tag{2}$$

where $k$ is the scan number ($1 \leq k \leq N_k$) and tilde indicates transposed.

The third step is to perform temporal regression to obtain subject specific spatial maps $S_k$, which means that the individual time-course matrices $A_k$ are used in a linear model fit (temporal regression) against the associated fMRI data set $X_k$ to estimate subject-specific spatial maps. The time series found by spatial regression is used as a temporal regressor to find the full set of voxels associated with that time series,

$$X_k = A_k \cdot S_k + E'. \tag{3}$$

Hence, the result of the dual regression algorithm are a subject-specific spatial maps $S_k$ based on the group averaged spatial map $S$. The procedure has been discussed in detail [54; 56].

These individual component maps are concatenated across subjects into single 4D files. Statistical significance testing is performed by applying nonparametric permutation testing (typically 5,000 permutations) [57]. This results in spatial maps characterizing the between-subject/group differences.

### 1.3.4 Dynamic functional connectivity analysis

Typically when extracting information on FC on the basis of rs-fMRI data, the whole time series is used to compute correlation values, assuming that the functional networks are stationary in time. That essentially means that information on events/interactions that occur at a shorter time scale is averaged out. Yet, neural processing occurs at a much shorter time scale and EEG studies led to the identification of microstates that involve network rearrangements in the millisecond time domain [58; 59]. Dynamic functional connectivity (dFC) analysis [60] is a concept that aims at estimating the FC changes over relatively short time intervals constituting a small fraction of the full time series recorded. Limiting factors when using fMRI data are the low-pass temporal filtering imposed by the hemodynamic response function and the sequential nature of MRI data acquisition, which limits the time required to capture an image volume to one or a few seconds. Despite these limitations, dFC has been suggested as a more accurate representation of functional brain networks since brain networks are dynamic. There have been many algorithms proposed in the literature to estimate dFC from resting state fMRI data. Sliding window based correlation analysis is a simple method to define a window of duration $\tau$, for which a conventional FC analysis is carried out. The window must comprise a sufficient number of sample points to allow for a meaningful correlation analysis, yet should not be too long in order to still capture dynamic aspects. The window is then shifted by an increment $\Delta t$ (typically one sampling point), while maintaining its length $\tau$ and the FC analysis is carried out again. Repeating this procedure several times allows monitoring the changes in FC networks over time. These resulting correlation matrices are then z-transformed and are called as dFC matrices $\boldsymbol{Z}_s(t_i; \tau)$. This procedure is illustrated in Figure 4.
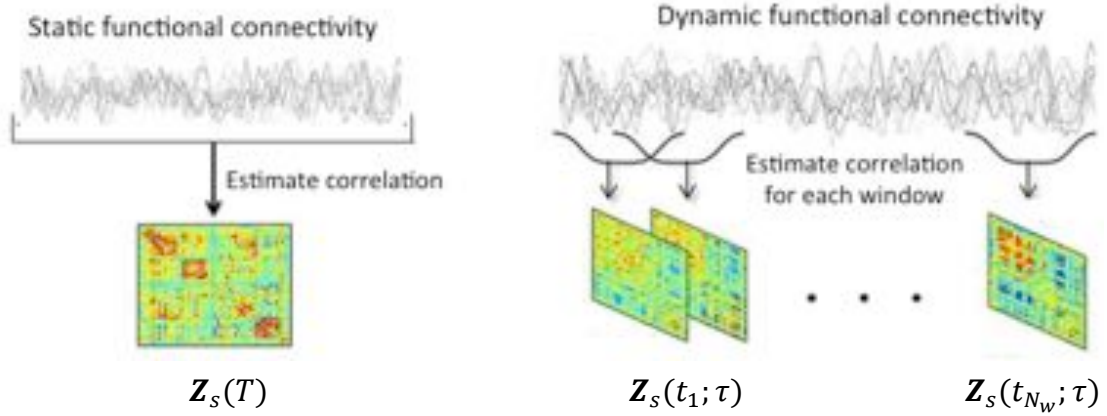
Figure 4: Static FC matrix $\mathbf{Z}_s(T)$ is obtained for each subject $s$ by computing the coefficients of correlation between the whole time series of duration $T$ for all pairs of regions (voxels) $i$ and $j$. dFC is estimated by carrying out the analogous analysis for time windows of duration $\tau$ separately, the starting point of which are incremented by $\Delta t$, yielding correlation matrices $\mathbf{Z}_s(t_i; \tau)$ with $1 \leq i \leq N_w$, $N_w = (T - \tau)/\Delta t$ being the number of windows. Adapted from [61].

Typically dFC had been used for analysing resting state fMRI data, however in 2010 Sakoglu et. al. applied dFC technique to the task related fMRI data [62]. They developed an approach that uses spatial ICA to estimate correlations between windowed time-courses of different brain networks (components). After estimating the spatial ICAs, they used the method developed by Calhoun and colleagues (http://mialab.mrn.org/software) based on the maximal lagged correlation approach [62] for estimating both stationary and dFC. The dFC was estimated by applying the window size of 64 time points.

**Algorithm**

We used the algorithm developed by Leonardi et al for estimating the dFC states. Feasibility of this algorithm has been demonstrated in few recent studies [63; 64; 65].

Sliding-window correlation between the time series x and y using the formula

$$r_{xy}(t) = corr(x[t, t + \tau], y[t, t + \tau]), \tag{4}$$

where $\tau$ was the window length in TRs and the window was shifted by $\Delta t$ for each

estimation. The coefficients of correlation were computed according to Pearson correlation as given in eq. (5)

$$r_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \ \sum(y-\bar{y})^2}} \ , \tag{5}$$

where $r$ is the coefficient of correlation between $x$ and $y$.

The correlation coefficient computed were then z-transformed according to

$$z_{xy}(t) = \text{atanh}\,(r_{xy}(t)) \tag{6}$$

The dFC analysis is carried out for each subject individually. . The resulting correlation matrices for each window were vectorized yielding a $K \times N_w$ matrix $\boldsymbol{C}_s$ for each subject and condition $s(1 \le s \le N_s)$, with $K$ being the number of pair-wise correlations and $N_w = (T - \tau)/\Delta t$ the number of windows. The matrix $\boldsymbol{C}_s$ was row-wise de-meaned to solely address the fluctuations of connectivity over time regardless of their mean value (Leonardi et al., 2013). This has been discussed further in Grandjean et al. (2017) with a detailed comparison of with and without this de-meaning step. The matrix $\boldsymbol{C}_s - \overline{\boldsymbol{C}}_s$ after demeaning then represents the increase/decrease in correlation strengths with respect to the mean (stationary FC).

The $N_s$ subject and condition-specific matrices $\boldsymbol{C}_s - \overline{\boldsymbol{C}}_s$ were then concatenated into a data matrix $\boldsymbol{X'}$ of dimension $K$ times the product $N_w \cdot N_s$

$$\boldsymbol{X'}^{(K \times (N_w \cdot N_s))} = \left(\boldsymbol{C}_1 - \overline{\boldsymbol{C}}_1 \big| ... \big| \boldsymbol{C}_{N_s} - \overline{\boldsymbol{C}}_{N_s}\right) = \left(x'_1 \big| ... \big| x'_{N_w \cdot N_s}\right) \tag{7}$$

with $x_i$ being vectors of length $K$.

## 1. Estimation of eigenconnectivities

The matrix $\boldsymbol{X'}$ describing the dynamic functional connectivity for a group of $N_s$ subject (or subjects times conditions) can be reduced, e.g. by principle component analysis (PCA), with a few PCs accounting for most of the variation across in the data matrix $\boldsymbol{X'}$. Leonardo et al. (2013) derived such components by estimating the eigenvectors and eigenvalues of $\boldsymbol{X'}$ according to

$$X' \cdot \widetilde{X}' = U \cdot \Lambda \cdot \widetilde{U} \qquad (8)$$

with $U$ being a unitary matrix containing the orthonormal eigenvectors as columns and $\Lambda$ being a diagonal matrix $\Lambda_{ik} = \delta_{ik} \cdot \lambda_i$ comprising the eigenvalues $\lambda_i$ (Fig. 5). The eigenvectors, i.e. the columns of $U$, were termed eigenconnectivities.
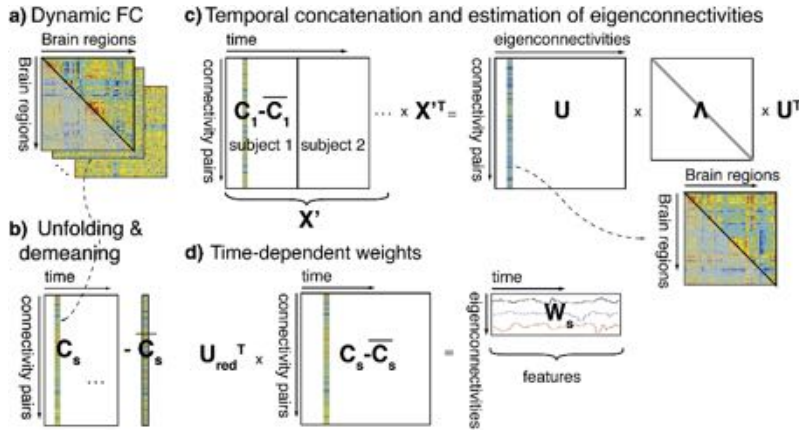


Figure 5: Schematics illustrating the dFC method. (from Leonardi et al 2013, reproduced with permission). (a) Dynamic FC between $n$ brain regions was computed as sliding window correlations between the activities of all regions for each subject. (b) The upper triangular part of each correlation matrix $Z_s(t_i; \tau)$ of dimension $n \times n$ was unfolded and concatenated across time to form a dynamic FC matrix $C_s$ of dimension $K \times N_w$, $K$ being the number of pairwise correlations and $N_w$ the number of time windows The mean correlation value across time was subtracted from all connectivity pairs (i.e., row-wise centering) yielding the matrix $C_s - \overline{C}_s$. (c) The dynamic FC matrices were concatenated across subjects to form a matrix $'$, for which eigenvalues $\lambda_i$ and eigenvectors (columns of unitary matrix $U$) were computed, so-called eigenconnectivities. Eigenconnectivities can be visualized after rearranging the eigenvectors of dimension $K \times 1$ them into a matrix of dimension $n \times n$ and symmetrizing, representing a indicating the interaction matrix corresponding to a specific eigenvalue in the basis of the brain regions . (d) The weight matrix $W_s$ containing the time dependent weights of each eigenconnectivity and subjectwas calculated by projecting the demeaned dynamic subject-specific FC matrix $C_s - \overline{C}_s$ onto a few eigenconnectivities, accounting for the major part of variation in the data.

Other algorithms may be used to estimate eigenconnectivities. For example, Grandjean et al. (2017) estimated eigenconnectivities based on dictionary learning algorithm. Dictionary learning algorithm aims to find the sparse representation of the input data and has been shown to produce good results in the fields of image classification and processing [66; 67]. In this work, we used dictionary-learning algorithm to estimate eigenconnectivities. Figure 6 shows that functional correlation matrices can be transformed in to dFC states by applying dictionary learning algorithm over it
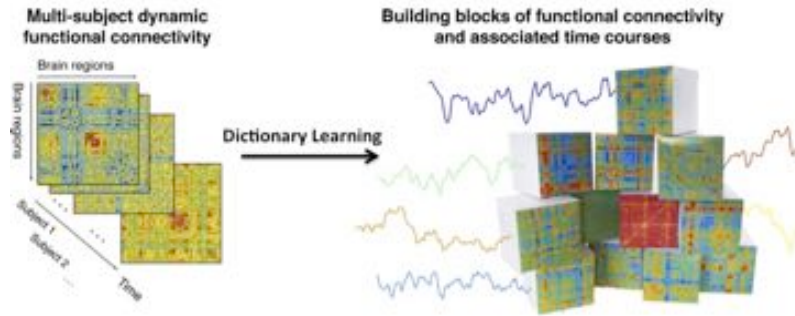


Figure 6: Functional correlation matrices estimates for each window and different subjects are concatenated in order to apply dictionary-learning algorithm for the estimation of dynamic functional states, or atoms; also referred to as the basic building blocks of dFC. Adopted from Leonardi et al 2013.

Dictionary learning algorithm involves the generation of a dictionary with

$$\boldsymbol{D}^{(K \times M)} = (d_1 | \dots d_M) \tag{9}$$

with $M$ $K$-dimensional column vectors as simple building blocks, the atoms, capturing whole-brain connectivity to a large extent, and a coefficient matrix of the concatenated subjects.

$$\boldsymbol{W}^{M \times (N_w \cdot N_s)} = \left( w_1 | \dots w_{N_w \cdot N_s} \right). \tag{10}$$

The columns of the concatenated coefficient matrix describe the approximation of the set of signals in $\boldsymbol{C}_s - \overline{\boldsymbol{C}}_s$ to combine these atoms. Like the number of IC components in ICA analysis, $M$ can be chosen arbitrarily, but following other papers we generated M=20 atoms.

Cost function minimization was implemented using a sparsity-enforcing algorithm giving $\boldsymbol{W}$ and $\boldsymbol{D}$ as outputs.

$$f(\boldsymbol{W}, \boldsymbol{D}) = \frac{1}{N_w \cdot N_s} \sum_{i=1}^{N_w \cdot N_s} \|\boldsymbol{w}_i\|_1 \, s.t. \, \|\boldsymbol{c}_i - \boldsymbol{D} \cdot \boldsymbol{w}_i\|_2^2 \leq r \tag{11}$$

In the above equation cost function tries to minimize the squared distance between $\boldsymbol{c}_i$ and the product of $\boldsymbol{D}$ and $\boldsymbol{w}_i$ and upper bounding it with a regularization parameter $r$

In order to retrieve easily interpretable atoms, we required them to be energy-bounded and positive by imposing the set of constraints

$$\boldsymbol{C} \triangleq \left\{ \boldsymbol{D} \in \mathbb{R}^{K \times M} s.t. \forall j = 1, \dots, M, \forall i = 1, \dots, K, \ \tilde{\boldsymbol{d}}_j \cdot \boldsymbol{d}_i \leq 1, D_{ij} \geq 0 \right\} \tag{12}$$

The first part of the equation shows $\boldsymbol{D} \in \mathbb{R}^{K \times M}$ that the dictionary atoms belongs to the real numbers with the matrix dimension of K (number of pairwise correlation) times M (number of atoms estimated). Furthermore, eq. (12) states that the atoms are energy-bounded, $\boldsymbol{d}_i \leq 1$, and positive, $D_{ij} \geq 0$.

Dictionary learning was performed 100 times (or folds; 400 iterations for the first fold, 200 for the subsequent ones). The obtained atoms were matched to the first fold using the Hungarian algorithm (Kuhn, 2010), with spatial correlation as the similarity metric. In order to get robust results, the fold instances exhibiting similarity above the median value across all folds were averaged for each atom

Animal-specific time-dependent contributions of atoms were obtained by back-projection of the dictionary onto the original dFC dataset using least-square fitting.

### 1.3.5 Network Modelling

Network modelling allows estimating interacting brain networks from fMRI time series data. Typically, functional connectivity (FC) between the brain regions is estimated using Pearson correlation given by

$$r_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \; \sum(y-\bar{y})^2}} \; , \tag{13}$$

where $r$ is the coefficient of correlation between $x$ and $y$. This covariance constitutes the simplest measure of pairwise similarity between two time-series is covariance and is commonly referred to refer to as full correlation. Smith and colleagues argued [68] that partial correlation analysis is more suited for analysing (FC) in the brain, since partial correlation refers to the normalized correlation between the time series measured for two brain regions, after each has been adjusted by regressing out all other time-series in the data, i.e. the interaction with all other regions (network nodes). Partial correlation attempts to distinguish direct from indirect connections, which is relevant as direct functional connection are considered representing effective connectivity between nodes (Friston, 1994)

Mathematically, partial correlation estimates regress the effects of all the other nodes of the networks from the correlation coefficients. Partial correlation between two regions is given as below

$$\rho_{xy} = corr \, [x, y \, | \, Z \, ] \tag{14}$$

$\rho_{xy}$ is the partial correlation between the time series of regions $x$ and $y$. Contributions of all other nodes $Z \neq x, y$ were regressed out beforehand from the time series $x, y$.

As we know that the pearson correlation can be calculated as follows

$$r_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \; \sum(y-\bar{y})^2}} \tag{15}$$

where $\bar{x}$ and $\bar{y}$ are sample means of random variables $r_x$ and $r_y$

To remove indirect effects, we can calculate the second order correlation between $r_x$ and $r_y$ by removing the effects of a node set $Z \subseteq (V - \{r_x, r_y\})$. This second order

correlation is called the partial correlation $\rho_{x,y.Z}$ controlled the node set Z. Therefore the partial correlation estimation can be written as:

$$\rho_{x,y} = \frac{\sum_{t=1}^{T}(\varepsilon_{x.Z}^t - \tilde{\varepsilon}_{x.Z})(\varepsilon_{y.Z}^t - \tilde{\varepsilon}_{y.Z})}{\sqrt{\sum_{t=1}^{T}(\varepsilon_{x.Z}^t - \tilde{\varepsilon}_{x.Z})^2 \ \sum_{t=1}^{T}(\varepsilon_{y.Z}^t - \tilde{\varepsilon}_{y.Z})^2}} \tag{16}$$

where $\varepsilon_{x.Z}^t$ is a set of the residuals of the linear regression with $r_x$ as the response variable and Z as the predictor variable set. So is $\varepsilon_{y.Z}^t$. This means that the partial correlation is a sampled full correlation when the controlling factors are regressed out

Figure 7 illustrates the difference between full and partial correlation for a simple network comprising 3 nodes and two edges.
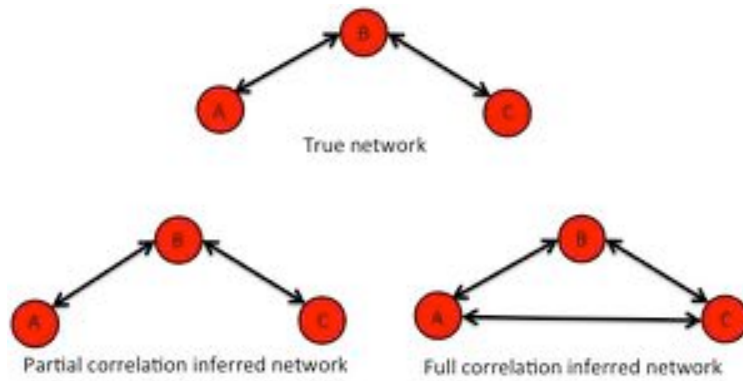


Figure 7: The true network is assumed to comprise three nodes A, B, C and two edges (A-B, B-C). This is correctly inferred by partial correlation analysis, while full covariance analysis (Pearson correlation) would also predict A_C to be a network edge. Hence partial correlation allows eliminating unwanted spurious edges for the network analysis [68].

There are many different toolboxes available to estimate the brain networks from fMRI data [68; 69; 70; 71; 72; 73] including the one developed by Smith et al. (FSL Nets) that performs network modeling from fMRI time series data based on taking spatial ICs, i.e. the output from stage 1 of dual regression analysis, as input. These group IC spatial maps are then regressed from each subject to generate subject-specific time courses which are preprocessed further, by removing outliers and bad components, followed by partial correlation estimation between all pairs of time series.

This generates a correlation matrix (network) of weights/connections, which is subdivided in to modules using hierarchical clustering. Afterwards statistical tests are performed to find group differences through a design matrix and randomized permutations method [74].

The networks obtained can help identifying differences among groups. Analysis, elements (edges) from the network matrix displaying the highest p-values for group differences may be used as fingerprint for differentiation and may hint a neurophysiological cause underlying the relevant response. As shown in Figure 8, the boxplots summarizes the distribution of the correlation values (connections strengths) in the groups. This allows interpreting the cause of the significant group difference. For example, it can be found that whether one of the connections was higher or lower than the other and whether any of the connections was not present in either of the group, allowing a simplistic interpretation of the connectivity changes between the groups.
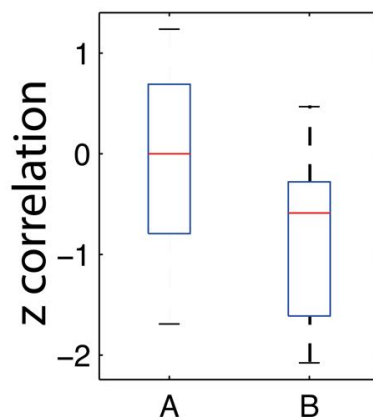


Figure 8: Boxplots showing the change in correlation value between the groups. The correlation estimated between the regions was close to zero in group A, while the regions are anti correlated in group B.

### 1.3.6 Effective connectivity analysis

Integration of segregated brain areas can be characterized in terms of functional connectivity using correlation analysis of the time series extracted from the regions. Functional connectivity is typically inferred on the basis of correlation between the brain regions. However, this approach just tells what is connected with what, but is not suitable for studying distributed neuronal processing in the brain. For such studies,

analysis of the effective connectivity has been suggested as an approach for studying the influence of one neural system over the other.

Functional connectivity can be quantified with statistical measures such as correlations, coherence etc. however effective connectivity determines the model parameters that best explains the observed dependencies (functional connectivity). Effective connectivity explicitly requires definition of models to select the best model, that mean effective connectivity estimation can essentially be reduced to model comparison.

In 2003 Friston et al. [75] proposed a method to reduce the "neuronal" evolution function to the most simple and generic form possible, i.e., a bilinear interaction between neuronal states $x$ and inputs $u$. They proposed to estimate effective connectivity [71; 76; 77] by fitting the neuronal model convolved with so-called "Balloon model", [19], an extension of hemodynamic response function, to the observed fMRI data. Then the deconvolution of the BOLD signal is performed to estimate neuronal response. This procedure is called dynamic causal modelling (DCM) as it attempts to infer causality between neuronal networks. Figure 7 shows the basic DCM model.

In order to estimate effective connectivity, several models are compared against each other to find the best fit to the experimental data (observed fMRI signal). Model comparison in effective connectivity plays an important role by selecting one of the several models, that best explains the cause of the observed fMRI data. Each alternative causal model represents distributed brain responses and therefore represents as many alternative hypotheses.
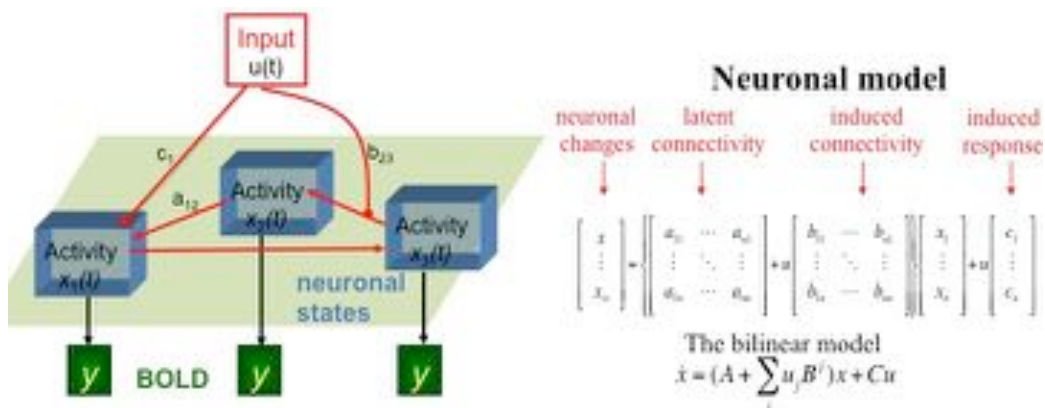
Figure 7: Figure shows a basic DCM model and its corresponding neuronal state equation.[71; 75; 77]

There have been numerous studies that use DCM to estimate the effective connectivity in humans as well as in other species [71; 78; 79; 80; 81; 82; 83; 84]. Due to the higher temporal resolution of local field potential (LFP) to infer on sequel of neuronal events, results of DCM have been validated by simultaneously recording the local LFP measurements that verified the ability of DCM to infer on synaptic processes. Razi et al [85] showed that functional connectivity can be derived from the effective connectivity, whereas the opposite is not true.

### 1.3.7 Machine learning

Typically fMRI analysis use linear regression models with several assumptions in a so-called mass univariate approach. This approach assumes that each voxel in the brain is independent of the other voxels, and therefore neglects the basic principles of brain functions that neurons interact with each other through synaptic connections and the activity in brain voxels is dependent on many other voxels through a complex underlying function. Machine learning allows inference over distributed patterns of activity in fMRI data. Machine learning as a multivariate approach addresses the shortcomings of mass univariate approaches by disregarding the independence assumption over voxels, and takes in to account the inter dependence of voxels over each other. While typical functional connectivity analysis methods can be considered as encoding models to predict brain activity from experimental context, machine-learning models are often termed as decoding models to predict experimental context from brain activity.

In typical machine learning classification analysis, the fMRI data is the observed brain activity. The patterns are learned from the brain activity to differentiate between two or several classes. Then an entirely new dataset consisting of the same or similar classes are used to validate the model. The high accuracy on the unseen test data reflects the higher ability to distinguish between different samples based on their neuroimaging features.

There have been several potential applications of applying machine learning algorithms to the fMRI data. One important application is in the diagnosis of different neurological disorders. Recent studies have also shown potential of machine learning algorithms for treatment response prediction. Other than that, FSL FIX algorithm used machine learning to learn the noise patterns in IC estimations, in order to remove noise at the single subject level.

Machine learning algorithms typically require low dimensional features with large number of samples to learn the features properly. However fMRI data is high dimensional in nature (>10000 voxels) and the data samples are typically small in number (<200 samples/subjects). Therefore it is challenging to apply machine-learning algorithms to the fMRI data because of the intrinsic need of machine learning algorithms of large number of samples and small number of features, without overfitting the data. Several dimensionality reduction algorithms have been proposed in the literature [86; 87] that aims to minimize the dimensionality of the data to avoid overfitting. Appropriate feature selection methods for neuroscience data have been discussed in [88]. Many of the machine learning algorithms also face the "black box" issue, where it is not exactly clear what have the algorithms learned. This makes them prone to finding spurious associations.

Feature selection for machine learning consists of various methods. The common goal of these methods is to find patterns that enable the accurate discrimination of cognitive states. For dimensionality reduction, several techniques are used among which Principle Component Analysis (PCA) and Independent Component Analysis (ICA), identifying features that explain the largest fraction of variability in the data, are the most common ones. Another strategy is the use of searchlight maps [89]. This method takes the average of the neighboring voxels, to encode similar information. It uses searchlight spheres comprised of spatially adjacent voxels to extract brain states. The spheres in the searchlight are used as features after averaging the voxels contained within each of them. The size of the spheres is iteratively changed in order to achieve better accuracy by the classifier. A more biologically inspired approach is the use anatomical regions instead of the spheres for generating feature vectors. This approach is useful as it also combines our knowledge of the brain anatomy. The brain is parcellated in to several anatomical regions and then each region is represented via its mean time series or first eigenvector after applying PCA. These eigenvectors are then

27

used to estimate the correlation matrix, which is fed in to the classification algorithm as feature maps. Several studies have utilized this approach [90; 91; 92]. However, the approach is dependent on the accuracy of the anatomical maps. Brain regions vary across subjects and therefore anatomical maps drawn on one subject will not perfectly match with the anatomical regions of other subjects. Furthermore operators' bias is introduced in the definition of anatomical regions. Some of these issues can be easily addressed by automatization of the procedure by using ICA components extracted from the dataset. Estimating ICA maps after concatenating the complete training dataset would generate dataset-specific spatial maps, and therefore a more accurate representation of the functional regions in the dataset.

Recently Broderson et. al. [93; 94] reported a method called generative embedding that uses effective connectivity estimated from DCM as features for the classifier. This procedure captures the network interactions at the neuronal levels and therefore transforms a high dimensional matrix to a low dimensional neuronal network model. They used SVM to successfully classify schizophrenic subjects from the control group. Similar attempts on using the network interactions as feature vectors in classification algorithms have also been reported in [95] where the results from network modeling using partial correlation are used as features for machine learning algorithms.

Among the many machine-learning algorithm, SVM is considered one of the simplest machine learning algorithm that identifies a separating hyper plane that maximizes the margin between the classes. Given a set of data points from two classes, the hyper plane that represents the largest separation between the classes or that maximizes the distance from it to the nearest data point on each side is called maximum-margin hyper plane.

SVM is a linear classifier that means it is capable of finding a hyperplane if the feature space is linearly separable. However most of the real world data is non linear in nature. In order to address this issue, an extension of SVM has been proposed in the literature with the use of kernels to find the separating hyper-plane. The idea of the kernels is to apply a nonlinear transformation to the input non linear feature space that may identify a hyper plane that linearly separates the classes in the transformed space.

Random Forest (RF) is a decision tree based classifier that performs random selection of features and uses bagging to construct decision trees with controlled variance. Random Forest algorithm construct a multitude of decision trees during training and predicts the label of the output class by counting the votes of the individual decision trees. This gives a probabilistic result unlike SVM that is gives binary classification output without any probability. Random Forests have been extensively used for classifying biological data, and it can be argued that it performs better than SVM for classification problems with large feature vectors. However Random Forest also tends to over fit the training data, therefore the validation procedure should be performed properly to avoid any inflated accuracies.

Neural networks are brain inspired machine learning algorithms, which aim to model mathematical representations of information processing via neurons. There are different layers defined in a neural network, where in the simplest case, first layer in the input layer, the second layer is made of activation units, outputting non-linear transformations and the third layer produces the classification prediction. Figure 8 shows the schematic of a simple two layer neural network.
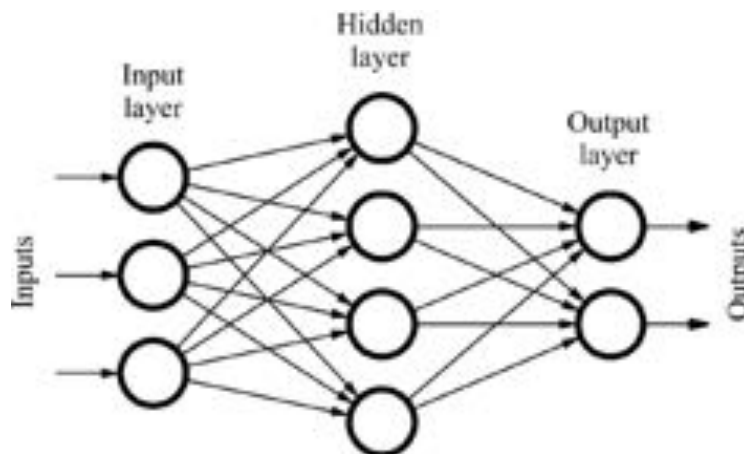


Figure 8: A simple feed forward neural network is shown. The first layer is the input layer, also called the visible layer, followed by the middle layer also called as the hidden layer. Inputs are transformed to the hidden using a weighting function W. Neural network output is generated at the output layer. From Quiza and Davim [96]. Reproduced with permission.

Recently Hinton and colleagues tested neural networks with several layers, also called as deep learning that has outperformed state of the arts methods to learn complex data

structures. However the proper use of deep learning requires optimization of several parameters, like number of units, layers and types of activation functions etc., which makes it non trivial. Recent developments using deep learning have been described in [97].

## 1.4. fMRI for small animals

### 1.4.1 Increasing sensitivity in small animal imaging

MRI offers high spatial resolution (of the order of the 10 um) however its temporal resolution is low (of the order of 1 sec.) relative to other techniques used in neuroscience for example EEG, which has a high temporal resolution (of the order of 1 ms) and low spatial resolution (of the order of 1 cm). Furthermore MRI has inherently low signal to noise ratio (SNR), which is dependent on the low degree of polarization and hence the low number of nuclei contributing to the signal. This is of particular relevance when imaging small animals, e.g. rodents and mice, as demands on spatial resolution are high and the signal is proportional to the dimension of the volume element. SNR can be increased by either increasing the signal, by decreasing the noise, or both [98]. High field magnets provide a solution to this by increasing the maximum achievable signal, however it comes with its own challenges. At high magnetic field strengths, images acquired contain artifacts due to magnetic inhomogeneity, forming another challenge to increase SNR. An alternative is the use of cryogenic coils operated at 20K to reduce the thermal noise component from the image, thus increasing the SNR. However this technique is only meaningful for small volumes since in large volumes sample noise dominates over electronic noise, while in small volumes both noise components are comparable. Therefore high field strength magnets combined with cryogenic coils, provide a solution for acquiring high resolution MR images in small animals, however this technique cannot be used in human MRI because in human MRI sample noise is larger than the electronic noise due to large size of the sample.

**1.4.2 Effects of anesthesia**

fMRI in small animal imaging plays an important role in evaluating the mechanistic aspects underlying neurovascular coupling and the BOLD signal. It is used as a non-invasive and hence translatable tool for characterization of the brain functional reorganization in response to a pharmacological challenge or to a pathological condition. Similar to humans, rs-fMRI has been extensively used to characterize functional networks in the rodent brain. Reproducible resting state networks have been demonstrated that include brain regions involved in auditory processing, motor function, and executive functioning etc. [99; 100]. Default mode network (DMN) has also been detected in rats [101; 102] similar to the human DMN [25; 26; 103].

While there are some reports of awake rat studies, the vast majority of rodent fMRI studies involve the use of anaesthesia. The use of anaesthetics in fMRI studies is even more common in mice than in rats, since it is more difficult to train mice for lying still in the magnet while being awake than rats. It is well known that anesthesia alters responses to stimuli or functional networks at rest and despite the frequent use of anaesthetics in small animal fMRI studies; the effects of anaesthetics have not yet been completely understood. The effects of anaesthesia on functional connectivity in the rat and mouse brain have been investigated by several groups [31; 70; 104; 105; 106; 107; 108; 109]. In 2015 Grandjean et. al. [110] compared the effect of four different anaesthetics on functional connectivity in the mouse brain using seed-based analysis and reported characteristic changes depending on the type of anaesthetic used. They found that functional connectivity patterns under medetomidine differed from those observed for animals under isoflurane, propofol or urethane anaesthesia.

Anaesthesia adds to the bias in the experimental measurements since the anaesthetics are known to affect the underlying neuronal signals, the physiological baseline, a critical determinant when measuring the hemodynamic response, as well as the efficiency of neurovascular coupling.

In order to evaluate the effects of anaesthetics on both the neuronal and neurovascular signal, electrophysiological recordings were measured in parallel with fMRI in rats under anaesthesia [105; 111; 112; 113; 114]. The results indicate that the local BOLD fluctuations and electrophysiological recordings show similar signals under several

anaesthetics, including isoflurane. This indicates that though anaesthetics may affect the neurovascular coupling, functional connectivity can still be reliably inferred. However since different anaesthetic regimens are known to affect different networks, it is important to select the anaesthetics based on the networks to be studied in order to minimize their interference with the underlying local neuronal signals.

## 1.5 References

[1] B. Milner, and D. Klein, Loss of recent memory after bilateral hippocampal lesions: memory and memories-looking back and looking forward. J Neurol Neurosurg Psychiatry 87 (2016) 230.

[2] W.B. Scoville, and B. Milner, Loss of recent memory after bilateral hippocampal lesions. 1957. J Neuropsychiatry Clin Neurosci 12 (2000) 103-13.

[3] L.R. Squire, The legacy of patient H.M. for neuroscience. Neuron 61 (2009) 6-9.

[4] R.J. Joynt, Centenary of patient "Tan". His contribution to the problem of aphasia. Arch Intern Med 108 (1961) 953-6.

[5] O.A. Selnes, and A. Hillis, Patient Tan revisited: a case of atypical global aphasia? J Hist Neurosci 9 (2000) 233-7.

[6] P.L. Nunez, R.B. Silberstein, P.J. Cadusch, R.S. Wijesinghe, A.F. Westdorp, and R. Srinivasan, A theoretical and experimental study of high resolution EEG based on surface Laplacians and cortical imaging. Electroencephalogr Clin Neurophysiol 90 (1994) 40-57.

[7] F. Babiloni, F. Cincotti, F. Carducci, P.M. Rossini, and C. Babiloni, Spatial enhancement of EEG data by surface Laplacian estimation: the use of magnetic resonance imaging-based head models. Clin Neurophysiol 112 (2001) 724-7.

[8] B. Burle, L. Spieser, C. Roger, L. Casini, T. Hasbroucq, and F. Vidal, Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. Int J Psychophysiol 97 (2015) 210-20.

[9] M.E. Raichle, A brief history of human brain mapping. Trends Neurosci 32 (2009) 118-26.

[10] S.S. Kety, and C.F. Schmidt, The Nitrous Oxide Method for the Quantitative Determination of Cerebral Blood Flow in Man: Theory, Procedure and Normal Values. J Clin Invest 27 (1948) 476-83.

[11] S.S. Kety, and C.F. Schmidt, The Effects of Active and Passive Hyperventilation on Cerebral Blood Flow, Cerebral Oxygen Consumption, Cardiac Output, and Blood Pressure of Normal Young Men. J Clin Invest 25 (1946) 107-19.

[12] W.H. Sweet, The uses of nuclear disintegration in the diagnosis and treatment of brain tumor. N Engl J Med 245 (1951) 875-8.

[13] M.E. Phelps, E.J. Hoffman, N.A. Mullani, and M.M. Ter-Pogossian, Application of annihilation coincidence detection to transaxial reconstruction tomography. J Nucl Med 16 (1975) 210-24.

[14] L. Sokoloff, M. Reivich, C. Kennedy, M.H. Des Rosiers, C.S. Patlak, K.D. Pettigrew, O. Sakurada, and M. Shinohara, The [14C]deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat. J Neurochem 28 (1977) 897-916.

[15] L. Sokoloff, Metabolic probes for localization of functional activity in the central nervous system. Int J Neurol 18 (1984) 40-8.

[16] S. Ogawa, T.M. Lee, A.R. Kay, and D.W. Tank, Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci U S A 87 (1990) 9868-72.

[17] K.K. Kwong, J.W. Belliveau, D.A. Chesler, I.E. Goldberg, R.M. Weisskoff, B.P. Poncelet, D.N. Kennedy, B.E. Hoppel, M.S. Cohen, R. Turner, and et al., Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci U S A 89 (1992) 5675-9.

[18] P.A. Bandettini, E.C. Wong, R.S. Hinks, R.S. Tikofsky, and J.S. Hyde, Time course EPI of human brain function during task activation. Magn Reson Med 25 (1992) 390-7.

[19] R.B. Buxton, E.C. Wong, and L.R. Frank, Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn Reson Med 39 (1998) 855-64.

[20] J.B. Mandeville, J.J. Marota, C. Ayata, G. Zaharchuk, M.A. Moskowitz, B.R. Rosen, and R.M. Weisskoff, Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. J Cereb Blood Flow Metab 19 (1999) 679-89.

[21] E.L. Hahn, Spin Echoes. Physical Review 80 (1950) 580-594.

[22] M. P., Multi-planar imaging formation using NMR spin-echo. Journal of Physics C: Solid State Physics 10 (1977).

[23] B. Biswal, F.Z. Yetkin, V.M. Haughton, and J.S. Hyde, Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med 34 (1995) 537-41.

[24] B.B. Biswal, Resting state fMRI: a personal history. Neuroimage 62 (2012) 938-44.

[25] M.E. Raichle, A.M. MacLeod, A.Z. Snyder, W.J. Powers, D.A. Gusnard, and G.L. Shulman, A default mode of brain function. Proc Natl Acad Sci U S A 98 (2001) 676-82.

[26] M.D. Fox, and M.E. Raichle, Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci 8 (2007) 700-11.

[27] S.A. Rombouts, C.J. Stam, J.P. Kuijer, P. Scheltens, and F. Barkhof, Identifying confounds to increase specificity during a "no task condition". Evidence for hippocampal connectivity using fMRI. Neuroimage 20 (2003) 1236-45.

[28] G. Deshpande, S. Laconte, S. Peltier, and X. Hu, Tissue specificity of nonlinear dynamics in baseline fMRI. Magn Reson Med 55 (2006) 626-32.

[29] R.M. Birn, J.B. Diamond, M.A. Smith, and P.A. Bandettini, Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. Neuroimage 31 (2006) 1536-48.

[30] J.E. Galvin, J.L. Price, Z. Yan, J.C. Morris, and Y.I. Sheline, Resting bold fMRI differentiates dementia with Lewy bodies vs Alzheimer disease. Neurology 76 (2011) 1797-803.

[31] J. Grandjean, R. Derungs, L. Kulic, T. Welt, M. Henkelman, R.M. Nitsch, and M. Rudin, Complex interplay between brain function and structure during cerebral amyloidosis in APP transgenic mouse strains revealed by multi-parametric MRI comparison. Neuroimage (2016).

[32] P. Vemuri, D.T. Jones, and C.R. Jack, Jr., Resting state functional MRI in Alzheimer's Disease. Alzheimers Res Ther 4 (2012) 2.

[33] Z. Wang, C. Yan, C. Zhao, Z. Qi, W. Zhou, J. Lu, Y. He, and K. Li, Spatial patterns of intrinsic brain activity in mild cognitive impairment and Alzheimer's disease: a resting-state functional MRI study. Hum Brain Mapp 32 (2011) 1720-40.

[34] M. Greicius, Resting-state functional connectivity in neuropsychiatric disorders. Curr Opin Neurol 21 (2008) 424-30.

[35] N.D. Woodward, and C.J. Cascio, Resting-State Functional Connectivity in Psychiatric Disorders. JAMA Psychiatry 72 (2015) 743-4.

[36] M.N. Baliki, A.R. Mansour, A.T. Baria, and A.V. Apkarian, Functional reorganization of the default mode network across chronic pain conditions. PLoS One 9 (2014) e106133.

[37] K.T. Martucci, W.R. Shirer, E. Bagarinao, K.A. Johnson, M.A. Farmer, J.S. Labus, A.V. Apkarian, G. Deutsch, R.E. Harris, E.A. Mayer, D.J. Clauw, M.D. Greicius, and S.C. Mackey, The posterior medial cortex in urologic chronic

pelvic pain syndrome: detachment from default mode network-a resting-state study from the MAPP Research Network. Pain 156 (2015) 1755-64.

[38] T. Wen, and S. Hsieh, Network-Based Analysis Reveals Functional Connectivity Related to Internet Addiction Tendency. Front Hum Neurosci 10 (2016) 6.

[39] M. Gottlich, N.M. Jandl, J.F. Wojak, A. Sprenger, J. von der Gablentz, T.F. Munte, U.M. Kramer, and C. Helmchen, Altered resting-state functional connectivity in patients with chronic bilateral vestibular failure. Neuroimage Clin 4 (2014) 488-99.

[40] I.M. Veer, C.F. Beckmann, M.J. van Tol, L. Ferrarini, J. Milles, D.J. Veltman, A. Aleman, M.A. van Buchem, N.J. van der Wee, and S.A. Rombouts, Whole brain resting-state analysis reveals decreased functional connectivity in major depression. Front Syst Neurosci 4 (2010).

[41] Y. Zhu, Y. Tang, T. Zhang, H. Li, Y. Tang, C. Li, X. Luo, Y. He, Z. Lu, and J. Wang, Reduced functional connectivity between bilateral precuneus and contralateral parahippocampus in schizotypal personality disorder. BMC Psychiatry 17 (2017) 48.

[42] C. Tang, Z. Zhao, C. Chen, X. Zheng, F. Sun, X. Zhang, J. Tian, M. Fan, Y. Wu, and J. Jia, Decreased Functional Connectivity of Homotopic Brain Regions in Chronic Stroke Patients: A Resting State fMRI Study. PLoS One 11 (2016) e0152875.

[43] C.F. Beckmann, M. DeLuca, J.T. Devlin, and S.M. Smith, Investigations into resting-state connectivity using independent component analysis. Philos Trans R Soc Lond B Biol Sci 360 (2005) 1001-13.

[44] C.F. Beckmann, and S.M. Smith, Tensorial extensions of independent component analysis for multisubject FMRI analysis. Neuroimage 25 (2005) 294-311.

[45] R.M. Hutchison, S.M. Mirsattari, C.K. Jones, J.S. Gati, and L.S. Leung, Functional networks in the anesthetized rat brain revealed by independent component analysis of resting-state FMRI. J Neurophysiol 103 (2010) 3398-406.

[46] M.J. McKeown, L.K. Hansen, and T.J. Sejnowsk, Independent component analysis of functional MRI: what is signal and what is noise? Curr Opin Neurobiol 13 (2003) 620-9.

[47] J.V. Stone, Independent component analysis: an introduction. Trends Cogn Sci 6 (2002) 59-64.

[48] V.D. Calhoun, T. Adali, G.D. Pearlson, and J.J. Pekar, Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Hum Brain Mapp 13 (2001) 43-53.

[49] K.J. Friston, Modes or models: a critique on independent component analysis for fMRI. Trends Cogn Sci 2 (1998) 373-5.

[50] L. Tian, Y. Kong, J. Ren, G. Varoquaux, Y. Zang, and S.M. Smith, Spatial vs. Temporal Features in ICA of Resting-State fMRI - A Quantitative and Qualitative Investigation in the Context of Response Inhibition. PLoS One 8 (2013) e66572.

[51] L. Griffanti, G. Salimi-Khorshidi, C.F. Beckmann, E.J. Auerbach, G. Douaud, C.E. Sexton, E. Zsoldos, K.P. Ebmeier, N. Filippini, C.E. Mackay, S. Moeller, J. Xu, E. Yacoub, G. Baselli, K. Ugurbil, K.L. Miller, and S.M. Smith, ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. Neuroimage 95 (2014) 232-47.

[52] G. Salimi-Khorshidi, G. Douaud, C.F. Beckmann, M.F. Glasser, L. Griffanti, and S.M. Smith, Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90 (2014) 449-68.

[53] C.F. Beckmann, and S.M. Smith, Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imaging 23 (2004) 137-52.

[54] N. Filippini, B.J. MacIntosh, M.G. Hough, G.M. Goodwin, G.B. Frisoni, S.M. Smith, P.M. Matthews, C.F. Beckmann, and C.E. Mackay, Distinct patterns of brain activity in young carriers of the APOE-epsilon4 allele. Proc Natl Acad Sci U S A 106 (2009) 7209-14.

[55] D.V. Smith, A.V. Utevsky, A.R. Bland, N. Clement, J.A. Clithero, A.E. Harsch, R. McKell Carter, and S.A. Huettel, Characterizing individual differences in functional connectivity using dual-regression and seed-based approaches. Neuroimage 95 (2014) 1-12.

[56] X.N. Zuo, C. Kelly, J.S. Adelstein, D.F. Klein, F.X. Castellanos, and M.P. Milham, Reliable intrinsic connectivity networks: test-retest evaluation using ICA and dual regression approach. Neuroimage 49 (2010) 2163-77.

[57] T.E. Nichols, and A.P. Holmes, Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15 (2002) 1-25.

[58] T. Koenig, L. Prichep, D. Lehmann, P.V. Sosa, E. Braeker, H. Kleinlogel, R. Isenhart, and E.R. John, Millisecond by millisecond, year by year: normative EEG microstates and developmental stages. Neuroimage 16 (2002) 41-8.

[59] D. Lehmann, P.L. Faber, S. Galderisi, W.M. Herrmann, T. Kinoshita, M. Koukkou, A. Mucci, R.D. Pascual-Marqui, N. Saito, J. Wackermann, G. Winterer, and T. Koenig, EEG microstate duration and syntax in acute, medication-naive, first-episode schizophrenia: a multi-center study. Psychiatry Res 138 (2005) 141-56.

[60] A.A. Ioannides, Dynamic functional connectivity. Curr Opin Neurobiol 17 (2007) 161-70.

[61] E.A. Allen, E. Damaraju, S.M. Plis, E.B. Erhardt, T. Eichele, and V.D. Calhoun, Tracking whole-brain connectivity dynamics in the resting state. Cereb Cortex 24 (2014) 663-76.

[62] U. Sakoglu, G.D. Pearlson, K.A. Kiehl, Y.M. Wang, A.M. Michael, and V.D. Calhoun, A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. MAGMA 23 (2010) 351-66.

[63] N. Leonardi, J. Richiardi, M. Gschwind, S. Simioni, J.M. Annoni, M. Schluep, P. Vuilleumier, and D. Van De Ville, Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest. Neuroimage 83 (2013) 937-50.

[64] N. Leonardi, W.R. Shirer, M.D. Greicius, and D. Van De Ville, Disentangling dynamic networks: Separated and joint expressions of functional connectivity patterns in time. Hum Brain Mapp 35 (2014) 5984-95.

[65] N. Leonardi, and D. Van De Ville, On spurious and real fluctuations of dynamic functional connectivity during rest. Neuroimage 104 (2015) 430-6.

[66] J. Mantilla, J. Paredes, J.J. Bellanger, E. Donal, C. Leclercq, R. Medina, and M. Garreau, Classification of LV wall motion in cardiac MRI using kernel Dictionary Learning with a parametric approach. Conf Proc IEEE Eng Med Biol Soc 2015 (2015) 7292-5.

[67] E. Varol, and C. Davatzikos, Supervised block sparse dictionary learning for simultaneous clustering and classification in computational anatomy. Med Image Comput Comput Assist Interv 17 (2014) 446-53.

[68] S.M. Smith, K.L. Miller, G. Salimi-Khorshidi, M. Webster, C.F. Beckmann, T.E. Nichols, J.D. Ramsey, and M.W. Woolrich, Network modelling methods for FMRI. Neuroimage 54 (2011) 875-91.

[69] J. Wang, X. Wang, M. Xia, X. Liao, A. Evans, and Y. He, GRETNA: a graph theoretical network analysis toolbox for imaging connectomics. Front Hum Neurosci 9 (2015) 386.

[70] D.V. D'Souza, E. Jonckers, A. Bruns, B. Kunnecke, M. von Kienlin, A. Van der Linden, T. Mueggler, and M. Verhoye, Preserved modular network organization in the sedated rat brain. PLoS One 9 (2014) e106156.

[71] K.J. Friston, B. Li, J. Daunizeau, and K.E. Stephan, Network discovery with DCM. Neuroimage 56 (2011) 1202-21.

[72] H. Onias, A. Viol, F. Palhano-Fontes, K.C. Andrade, M. Sturzbecher, G. Viswanathan, and D.B. de Araujo, Brain complex network analysis by means of resting state fMRI and graph analysis: will it be helpful in clinical epilepsy? Epilepsy Behav 38 (2014) 71-80.

[73] T. Xu, K.R. Cullen, B. Mueller, M.W. Schreiner, K.O. Lim, S.C. Schulz, and K.K. Parhi, Network analysis of functional brain connectivity in borderline personality disorder using resting-state fMRI. Neuroimage Clin 11 (2016) 302-15.

[74] A.M. Winkler, G.R. Ridgway, M.A. Webster, S.M. Smith, and T.E. Nichols, Permutation inference for the general linear model. Neuroimage 92 (2014) 381-97.

[75] K.J. Friston, L. Harrison, and W. Penny, Dynamic causal modelling. Neuroimage 19 (2003) 1273-302.

[76] J. Kahan, and T. Foltynie, Understanding DCM: ten simple rules for the clinician. Neuroimage 83 (2013) 542-9.

[77] K.E. Stephan, W.D. Penny, R.J. Moran, H.E. den Ouden, J. Daunizeau, and K.J. Friston, Ten simple rules for dynamic causal modeling. Neuroimage 49 (2010) 3099-109.

[78] A.M. Bastos, V. Litvak, R. Moran, C.A. Bosman, P. Fries, and K.J. Friston, A DCM study of spectral asymmetries in feedforward and feedback connections between visual areas V1 and V4 in the monkey. Neuroimage 108 (2015) 460-75.

[79] R. Moran, D.A. Pinotsis, and K. Friston, Neural masses and fields in dynamic causal modeling. Front Comput Neurosci 7 (2013) 57.

[80] H.R. Brown, and K.J. Friston, The functional anatomy of attention: a DCM study. Front Hum Neurosci 7 (2013) 784.

[81] J. Daunizeau, L. Lemieux, A.E. Vaudano, K.J. Friston, and K.E. Stephan, An electrophysiological validation of stochastic DCM for fMRI. Front Comput Neurosci 6 (2012) 103.

[82] K.J. Friston, J. Kahan, B. Biswal, and A. Razi, A DCM for resting state fMRI. Neuroimage 94 (2014) 396-407.

[83] M.I. Garrido, K.J. Friston, S.J. Kiebel, K.E. Stephan, T. Baldeweg, and J.M. Kilner, The functional anatomy of the MMN: a DCM study of the roving paradigm. Neuroimage 42 (2008) 936-44.

[84] D.A. Pinotsis, N. Brunet, A. Bastos, C.A. Bosman, V. Litvak, P. Fries, and K.J. Friston, Contrast gain control and horizontal interactions in V1: a DCM study. Neuroimage 92 (2014) 143-55.

[85] A. Razi, J. Kahan, G. Rees, and K.J. Friston, Construct validation of a DCM for resting state fMRI. Neuroimage 106 (2015) 1-14.

[86] S.E. Viswanath, P. Tiwari, G. Lee, A. Madabhushi, and I. Alzheimer's Disease Neuroimaging, Dimensionality reduction-based fusion approaches for imaging and non-imaging biomedical data: concepts, workflow, and use-cases. BMC Med Imaging 17 (2017) 2.

[87] J. Ye, Dimension reduction algorithms in data mining, with applications, University of Minnesota, 2005, pp. 91.

[88] R. Pang, B.J. Lansdell, and A.L. Fairhall, Dimensionality reduction in neuroscience. Curr Biol 26 (2016) R656-60.

[89] N. Kriegeskorte, R. Goebel, and P. Bandettini, Information-based functional brain mapping. Proc Natl Acad Sci U S A 103 (2006) 3863-8.

[90] Z. Wang, A.R. Childress, J. Wang, and J.A. Detre, Support vector machine learning-based fMRI data group analysis. Neuroimage 36 (2007) 1139-51.

[91] A. Mahmoudi, S. Takerkart, F. Regragui, D. Boussaoud, and A. Brovelli, Multivoxel pattern analysis for FMRI data: a review. Comput Math Methods Med 2012 (2012) 961257.

[92] Q. Bukhari, D. Borsook, M. Rudin, and L. Becerra, Random Forest Segregation of Drug Responses May Define Regions of Biological Significance. Front Comput Neurosci 10 (2016) 21.

[93] K.H. Brodersen, L. Deserno, F. Schlagenhauf, Z. Lin, W.D. Penny, J.M. Buhmann, and K.E. Stephan, Dissecting psychiatric spectrum disorders by generative embedding. Neuroimage Clin 4 (2014) 98-111.

[94] K.H. Brodersen, T.M. Schofield, A.P. Leff, C.S. Ong, E.I. Lomakina, J.M. Buhmann, and K.E. Stephan, Generative embedding for model-based classification of fMRI data. PLoS Comput Biol 7 (2011) e1002079.

[95] A. Khazaee, A. Ebrahimzadeh, and A. Babajani-Feremi, Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease. Brain Imaging Behav 10 (2016) 799-817.

[96] R. Quiza, and J.P. Davim, Computational Methods and Optimization. in: J.P. Davim, (Ed.), Machining of Hard Materials, Springer London, London, 2011, pp. 177-208.

[97] I. Arel, D.C. Rose, and T.P. Karnowski, Research frontier: deep machine learning--a new frontier in artificial intelligence research. Comp. Intell. Mag. 5 (2010) 13-18.

[98] J. Grandjean, and M. Rudin, What can functional connectivity in mice tell us about Alzheimer\U+2019\s disease? An investigation with resting-state fMRI at 9.4T, ETH-Zürich, Zürich, 2014, pp. 1 Band.

[99] J.S. Damoiseaux, S.A. Rombouts, F. Barkhof, P. Scheltens, C.J. Stam, S.M. Smith, and C.F. Beckmann, Consistent resting-state networks across healthy subjects. Proc Natl Acad Sci U S A 103 (2006) 13848-53.

[100] V. Zerbi, J. Grandjean, M. Rudin, and N. Wenderoth, Mapping the mouse brain with rs-fMRI: An optimized pipeline for functional network identification. Neuroimage 123 (2015) 11-21.

[101] F. Sforazzini, A.J. Schwarz, A. Galbusera, A. Bifone, and A. Gozzi, Distributed BOLD and CBV-weighted resting-state networks in the mouse brain. Neuroimage 87 (2014) 403-15.

[102] H. Lu, Q. Zou, H. Gu, M.E. Raichle, E.A. Stein, and Y. Yang, Rat brains also have a default mode network. Proc Natl Acad Sci U S A 109 (2012) 3979-84.

[103] P. Fransson, Spontaneous low-frequency BOLD signal fluctuations: an fMRI investigation of the resting-state default mode of brain function hypothesis. Hum Brain Mapp 26 (2005) 15-29.

[104] I. Nasu, N. Yokoo, S. Takaoka, K. Takata, T. Hoshikawa, M. Okada, and Y. Miura, The dose-dependent effects of isoflurane on outcome from severe

forebrain ischemia in the rat. Anesth Analg 103 (2006) 413-8, table of contents.

[105] K. Masamoto, M. Fukuda, A. Vazquez, and S.G. Kim, Dose-dependent effect of isoflurane on neurovascular coupling in rat cerebral cortex. Eur J Neurosci 30 (2009) 242-50.

[106] W.J. Pan, J.C. Billings, J.K. Grooms, S. Shakil, and S.D. Keilholz, Considerations for resting state functional MRI and functional connectivity studies in rodents. Front Neurosci 9 (2015) 269.

[107] E. Jonckers, D. Shah, J. Hamaide, M. Verhoye, and A. Van der Linden, The power of using functional fMRI on small rodents to study brain pharmacology and disease. Front Pharmacol 6 (2015) 231.

[108] E. Jonckers, J. Van Audekerke, G. De Visscher, A. Van der Linden, and M. Verhoye, Functional connectivity fMRI of the rodent brain: comparison of functional connectivity networks in rat and mouse. PLoS One 6 (2011) e18876.

[109] A. Schroeter, F. Schlegel, A. Seuwen, J. Grandjean, and M. Rudin, Specificity of stimulus-evoked fMRI responses in the mouse: the influence of systemic physiological changes associated with innocuous stimulation under four different anesthetics. Neuroimage 94 (2014) 372-84.

[110] J. Grandjean, A. Schroeter, I. Batata, and M. Rudin, Optimization of anesthesia protocol for resting-state fMRI in mice based on differential effects of anesthetics on functional connectivity patterns. Neuroimage 102 Pt 2 (2014) 838-47.

[111] K. Masamoto, and I. Kanno, Anesthesia and the quantitative evaluation of neurovascular coupling. J Cereb Blood Flow Metab 32 (2012) 1233-47.

[112] B.J. Palanca, G.A. Mashour, and M.S. Avidan, Processed electroencephalogram in depth of anesthesia monitoring. Curr Opin Anaesthesiol 22 (2009) 553-9.

[113] C.J. Pomfrett, J.R. Sneyd, J.R. Barrie, and T.E. Healy, Respiratory sinus arrhythmia: comparison with EEG indices during isoflurane anaesthesia at 0.65 and 1.2 MAC. Br J Anaesth 72 (1994) 397-402.

[114] P.L. Purdon, E.T. Pierce, G. Bonmassar, J. Walsh, P.G. Harrell, J. Kwo, D. Deschler, M. Barlow, R.C. Merhar, C. Lamus, C.M. Mullaly, M. Sullivan, S. Maginnis, D. Skoniecki, H.A. Higgins, and E.N. Brown, Simultaneous electroencephalography and functional magnetic resonance imaging of general anesthesia. Ann N Y Acad Sci 1157 (2009) 61-70.

## 2. Resting state fMRI in mice reveals anesthesia specific signatures of brain functional networks and their interactions

**Bukhari Q, Schroeter A, Cole D and Rudin M**

**Abstract**

fMRI studies in mice typically require the use of anesthetics. Yet, it is known that anesthesia alters responses to stimuli or functional networks at rest. In this work, we have used Dual Regression analysis Network Modeling to investigate the effects of two commonly used anesthetics, isoflurane and medetomidine, on rs-fMRI derived functional networks, and in particular to what extent anesthesia affected the interaction within and between these networks.

Experimental data have been used from a previous study [1]. We applied multivariate ICA analysis and Dual Regression to infer the differences in functional connectivity between isoflurane and medetomidine. Further network analysis was performed to investigate within- and between-network connectivity differences between these anesthetic regimens. The results revealed five major networks in the mouse brain: lateral cortical, associative cortical, default mode, subcortical, and thalamic network. The anesthesia regime had a profound effect both on within- and between-network interactions. Under isoflurane anesthesia predominantly intra- and inter-cortical interactions have been observed, with only minor interactions involving subcortical structures and in particular attenuated cortico—thalamic connectivity. In contrast, medetomidine anesthetized mice displayed subcortical functional connectivity including interactions between cortical and thalamic ICA components. Combining the two anesthetics at low dose resulted in network interaction that constituted the superposition of the interaction observed for each anesthetic alone.

The study demonstrated that network modeling is a promising tool for analyzing the brain functional architecture in mice and comparing alterations therein caused by

different physiological or pathological states. Understanding the differential effects of anesthetics on brain networks and their interaction is essential when interpreting fMRI data recorded under specific physiological and pathological conditions.

**2.1 Introduction**

Analyzing adaptations of brain networks is becoming increasingly important for characterizing physiological or pathological states, or evaluating responses to therapeutic interventions. Parallel to the Human Connectome Project [2] there are considerable efforts to elucidate structural and functional connectivity also in rodents triggered by the expectation that rodent studies might provide valuable translational insight into mechanisms underlying FC and how these are altered during pathology. In functional magnetic resonance imaging (fMRI), functional connectivity across brain regions can be inferred from the temporal correlation of fluctuations in the baseline fMRI signal, i.e. under stimulus-free conditions. Several techniques have been suggested for analyzing resting state fMRI (rs-fMRI) data including seed-based analysis or dual regression. Seed-based analysis is straightforward, yet as univariate method considers each voxel independently, which implies that the approach considers only one effect at a time [3]. Seed-based analysis faces concerns related to the inherent biases of experimenter selection of seed regions [4]. In addition, any network not associated to these seeds cannot be identified. The quality of seed-based analysis depends critically on the seed selection, which should be optimally adapted to the anatomical or functional brain areas. Therefore, regions are typically derived from a neuroanatomical atlas, which however may not be optimally adapted to the structural/functional unit for a specific subject due to anatomical variability and/or imperfect registration of the image data set to the atlas-based template. ICA and seed-based are complementary approaches that have pros and cons. Seed based analysis allows selection of fine-grained functional units; however, the use of fine-grained seeds is susceptible to errors related to registration. Also, it has been shown that biases inherent in the seed selection can result in a large variability in the results [4; 5], which becomes crucial during network estimation. Nevertheless, if there is a strong prior hypothesis with regard to the involvement of specific brain regions, seed based analysis using adapted seed masks are of great value. In contrast, regions derived from the ICA analysis are not subject to any anatomical constraints, which imply that given

the limitations in sensitivity and intrinsic spatial resolution, they may not be ideally matched to structural/functional units as derived from a high-resolution brain atlas. Careful inspection of ICA results and removal of noise components becomes essential. Typically, ICA components are allocated to structural units in a post-hoc manner according to the best fit. ICA analysis does typically not produce fine-grained functional maps rendering them more robust against registration errors, at the expense of fine structure information. In the absence of a hypothesis, ICA analysis appears appropriate as it is purely data driven.

Dual Regression (DR) in combination with probabilistic independent component analysis (ICA) constitutes a multivariate approach for analyzing rs-fMRI data [6] with spatial ICA maps being fed as input in to the DR pipeline. The DR approach [5; 6] first regresses the z-normalized group-IC spatial maps against the subject-specific 4D resampled datasets to give a set of subject-specific, variance normalized time courses for each component separately, and then – at a second-level of regression – these time-courses are regressed against the same 4D dataset to calculate a subject-specific set of spatial maps. The use of multivariate methods allows the simultaneous consideration of effects from all brain regions, i.e., the brain is treated as a fully connected network, or set of networks. ICA has been shown to produce reliable and comparable results both at the individual subject and the group level [7; 8]. For in-depth network analysis, between-network interactions can also be considered by a comprehensive analysis of all ICA components[9]. Graph theoretical approaches could be used to further analyze DR derived network information in order to structure them according to clusters (sub networks), nodes and edges. While frequently applied to human fMRI data, use of such approaches in small animal fMRI is still rather limited [10; 11]. Network-based approaches have also been recently applied to mouse and rat rs-fMRI data, including prior parcellation into ICA components [12; 13]

Both medetomidine and isoflurane have been used in longitudinal fMRI experiments in rodents yielding robust BOLD response to external stimuli [14; 15; 16; 17]. As these agents involve different modes of actions, which affect both central and peripheral responses it is not surprising that fMRI responses were found to depend on the specific anesthetic used [17; 18]. Their differential effect is also reflected by anesthetic specific functional connectivity patterns [1; 18]. In particular, it has been reported that medetomidine, while yielding rather stable results in rats, decreases inter-

hemispheric FC in mice [19; 20; 21]. In general, the optimal choice of anesthetic will depend on the specific problem to be addressed, i.e. should have minimal interference with the processes to be studied. The combination of complementary anesthetics may have synergistic effects and allow reducing the dose of the individual agents and thereby unwanted biochemical/physiological side effects [1; 16].

In this study we evaluated the use of DR followed by graph theory based network analysis for detecting differences in mouse functional networks with respect to anesthesia-induced differences in physiological state. We analyzed the effects of two commonly used anesthetics and their combination, which have been shown to affect functional connectivity patterns in a drug-dependent manner [1]. In particular, we focused on obtaining detailed interactions among networks and sub-networks of mouse brain functional architecture. We analyzed blood-oxygen level-dependent (BOLD) rs-fMRI data in terms of interacting fMRI networks by using partial correlation, which is thought to more closely represent brain functional principles than simple correlation of time courses extracted from individual seeds [5; 9].


## 2.2 Materials and methods

### 2.2.1 Imaging

**Animals, preparation and anesthesia**

The analysis is based on rs-fMRI data collected in an earlier study; we refer to Grandjean et al. [1], where experimental details have been described. In brief, female C57BL/6 mice of 10 to 15 weeks of age have been used for the study. For the rs-fMRI data collection mice had been intubated and artificially ventilated with an 80% air 20% oxygen mixture using a small animal ventilator (CWE, Ardmore, USA). Three groups of mice subject to different anesthesia protocols were studied: group 1 (N=11) received 1% isoflurane administered via the ventilation mixture; group 2 (N=13) an initial i.v bolus injection of 0.1 mg/kg medetomidine hydrochloride followed by a continuous infusion at a rate of 0.2mg/kg/h of the drug; and group 3 (N=8) received the combination of isoflurane and medetomidine with half the doses administered in groups 1 and 2, respectively. Rs-fMRI data have been acquired using a Bruker Biospec 94/30 small animal MR system (Bruker BioSpin MRI, Ettlingen, Germany) operating at 400 MHz (9.4 T) equipped with a four-element receive-only cryogenic phased array

coil (Bruker BioSpin AG, Fällanden, Switzerland). Detailed acquisition parameters are given in Grandjean et. al. [1]. For the current study, data have been downloaded from the central.xnat.org repository (Project ID: fMRI_ane_mouse;[1]). BOLD fMRI experimental data were acquired using a gradient-echo echo-planar imaging (GE-EPI) sequence: FOV = 23.7 x14 mm2, MD = 90 x 60, yielding an in-plane voxel dimension of 263 x 233 μm, flip angle (FA) = 90°, bandwidth = 300 kHz, TR =1000 ms, TE = 10 ms, NA = 1, yielding a temporal resolution of 1 s, with interleaved acquisition of slices. The duration of the image time series was 6 min. Mice were paralyzed in order to facilitate artificial ventilation and to eliminate motion artifacts during data acquisition. We analyzed reflexes and flinching behavior in non-paralyzed animals and did not detect any differences between animals anesthetized with either anesthetic nor did we detect any indication of pain.

### 2.2.2 Data processing and statistical analysis

### Preprocessing

All the preprocessing was performed using FSL's recommended preprocessing pipeline from FMRIB's Software Library (FSL version 5). Preprocessing included motion correction, removal of non-brain structures, high pass temporal filtering with sigma = 75.0s, pre-whitening and global spatial smoothing using a filter with a 0.2 mm kernel. After the pre-processing the functional scans were aligned to the high-resolution template EPI scan using non-linear registration with 7 degrees of freedom as implemented in FLIRT, followed by nonlinear (FNIRT) warping [22; 23]

### a) ICA Analysis and Dual Regression

We used FSL's MELODIC software for probabilistic independent component analysis [24]. The multi-session temporal ICA concatenated (Concat-ICA) approach, as recommended for resting state data analysis [25; 26], allowed the inputting of all subjects from all the groups in a temporally concatenated fashion for the ICA analysis. Concat-ICA yielded different components without the need for specifying any explicit time series model.

A total of 70 independent components (IC maps) were extracted from each analysis group. A mixture model approach was used to perform the inference on estimated maps. An alternative hypothesis test based on fitting a Gaussian/gamma mixture

model to the distribution of voxel intensities within spatial maps [26] was used to threshold the IC maps. Out of the 70 independent components (IC maps) in each group, only 17 components on average were selected for each comparison, while the components that overlapped with vascular structures and ventricles were excluded from further analysis, however these components were still included as the regressors of no interest in the DR analysis. Similarly, components concentrated within the regions at the brain surface, which are prone to be affected by motion-related artifacts, were also excluded. Supplementary Figure 1 shows the removed ICA components.

We used DR (FSL 5.0.2.2) for between-subject analysis allowing for voxel-wise comparisons of rs-fMRI data [6; 27]. We used unpaired t-tests to test for differences between anesthetic regimen conditions. Specifically, the design matrix was subject to [1 -1] contrasts to identify brain regions and networks displaying greater FC in one anesthetic condition relative to another.

Non-parametric permutation based inference analysis [28] was performed with subject-specific component spatial maps concatenated across subjects and submitted to voxel-wise between-subject analysis testing for effects of anesthetics on FC using FSL-randomise *[29]*. FSL's general linear model (GLM) was used to define contrasts based on unpaired t-test, testing for anesthesia effects among different groups. For each analysis we ran 5000 randomized permutations in line with the FSL default recommendations, while threshold-free cluster enhancement [30] was used for statistical inference to validate the likelihood of extended areas of signal, which also takes in to account information from neighboring voxels. TFCE enhances cluster-like structures but the image remains fundamentally voxelized. This cluster enhancement renders TFCE more sensitive than voxel-wise thresholding [30]. Correction for multiple comparisons across space was applied assuming an overall significance of a ($p < 0.05$) using permutation testing and TFCE. Bonferroni correction ($p \leq 0.05/17$) was applied separately to each analysis depending on the number of components of interest[31].

**b) Network modeling**

FSLNets (FSL, 5.0.2.2) has been used for network modeling of rs-fMRI data. The data processing pipeline is depicted in Suppl. Figure 2. Different network matrix calculation methods have been applied. Full correlation (FC) estimates both direct and indirect connections, while partial correlation (PC) only estimates direct connections.

We used L1 partial correlation method for Partial Correlation (PC) analysis, which yielded direct connections only [9]. The PC matrices of the BOLD time courses of each component from dual regression were then clustered to form a dendrogram. These clusters were then used as input in to the GLM analysis and run through FSL-randomise [29]to perform 5000 permutations to test for statistical significance. Edges, i.e. connections between network nodes showing statistically significant differences between the groups under consideration were obtained from GLM analysis. These significant network edges were then used to calculate network box plots (Suppl. Table 1 summarizes the values obtained through box plots.) that take into account each edge and provide more information on differences in connectivity values between the groups. FSLNets was corrected for multiple comparisons with false discovery rate (FDR) using the same unpaired t-test design matrix as used previously for DR analysis.

## 2.3 Results

### 2.3.1 Dual Regression confirmed results of seed-based analysis and identified additional components

Of the 70 components derived from ICA, an average of 17 components (range 16-18) were retained for further analyses after discarding components at the brain surface and those involving vascular structures or ventricles. The number varied across individual analyses as a different number of components had to be discarded according to our selection criteria. Apart from auditory cortices, which appeared strictly lateralized, ICA derived components typically comprised bilateral homotopic brain areas (Figure 1, Table 1).
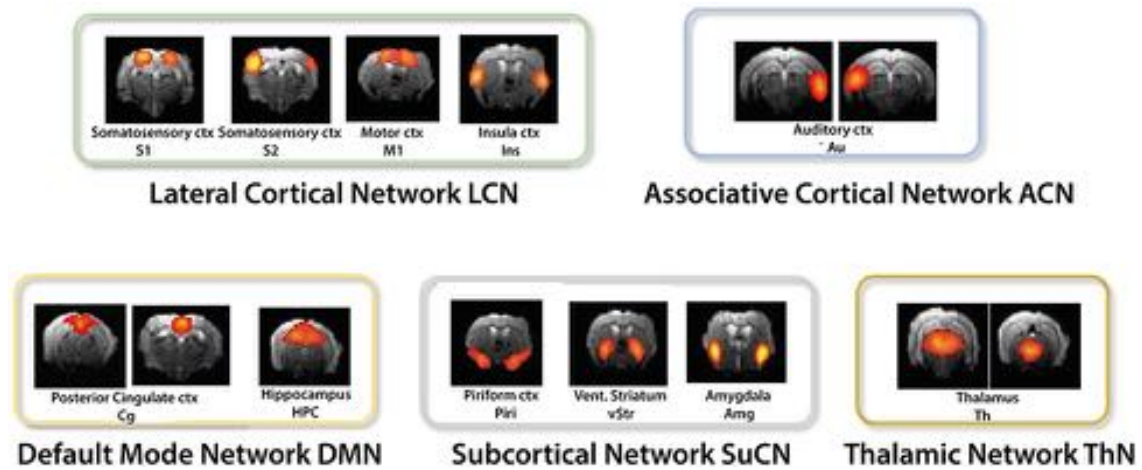
**Figure 1**: ICA derived components were grouped into the lateral cortical network (LCN), associative cortical network (ACN), default mode network (DMN), subcortical network (SuCN), and the thalamic network (ThN).

| | | Isoflurane | Medetomidine | med/iso |
|---|---|:---:|:---:|:---:|
| Lateral Cortical Network (LCN) | Somatosensory cortex S1/S2 | X | X | X |
| | Motor cortex M1 | X | X | X |
| | Suppl motor cortex M2 | X | X | X |
| | Insular cortex Ins | X | X | X |
| Associative Cortical Network (ACN) | Limbic cortex Lim | X | X | X |
| | Visual cortex Vis | | X | |
| | Auditory cortex Au | X | X | X |
| Default Mode Network (DMN) | Prefrontal cortex PFC | | X | |
| | Cingulate cortex Cg | X | X | X |
| | Dorsal hippocampus dHPC | X | | |
| | Ventral hippocampus vHPC | X | X | X |
| Subcortical Network (SuCN) | Piriform cortex Piri | X | X | X |
| | Dorsal striatum dStr | X | | |
| | Lateral striatum lStr | | | X |
| | Ventral striatum vStr | X | X | X |
| | Amygdala Amg | X | X | X |
| Thalamic Network (THN) | Dorsal thalamus dTh | X | X | X |
| | Ventral thalamus vTh | X | X | X |
| Extended SuCN | Hypothalamus HTH | X | X | X |
| | Globus pallidus GP | X | X | |
| | Olfactory tubercle OT | X | | X |

Allocation of ICA components identified to major brain networks.

**Table 1**: Regions identified in selected ICA components.

DR revealed differences in functional connectivity between isoflurane- and medetomidine-anestethised mice (Figure 2). Cortical areas display major difference between the two anesthetics, with mice under medetomidine anesthesia displaying only very weak intercortical functional connectivity.
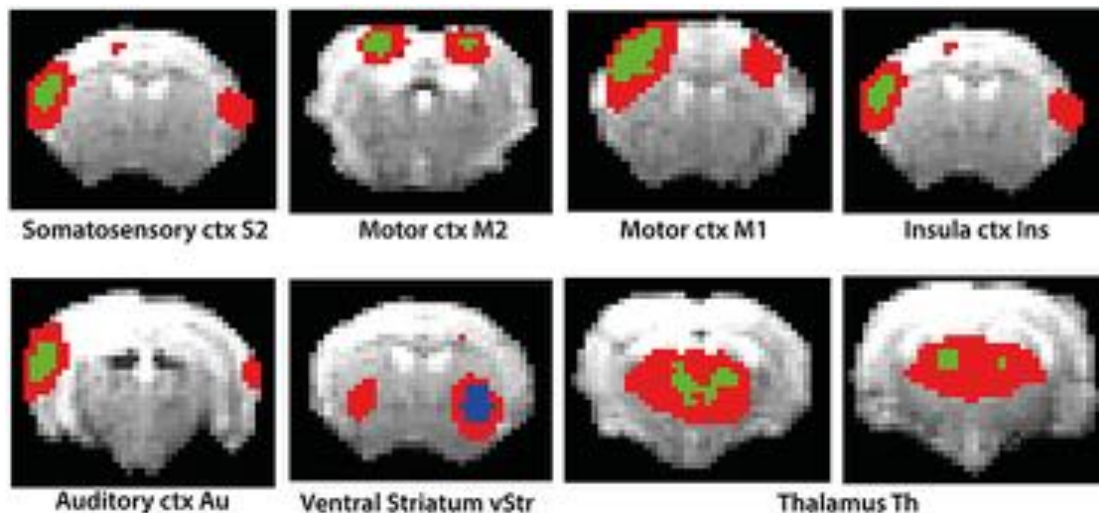


**Figure** 2: Results of Dual Regression analysis for eight components derived from ICA (red color overlay). The reference of the components to anatomical structures is indicated in the figure. Bonferroni corrected DR results showing the regions whose co-activation with the ICA components (as shown in the figure) was significantly higher (green color) or lower (blue color) in isoflurane as compared to medetomidine-anesthetized mice.

The comparison of DR using probabilistic ICA based temporal concatenation with seed-based analysis using ICA informed seed selection revealed obvious similarities. Supplementary Fig. 3 shows the comparison of seed based with the ICA derived components. By definition the outcome of seed-based analysis is confined to network components associated with the selected seed region, typically revealing functional connectivity between homotopic regions. It is therefore not surprising that the model-free DR approach revealed additional anatomical regions as part of networks that displayed profound group differences when compared to the results of seed-based analysis. The previous study reports seed selection to include three sensory regions, the anterior (ant-), medial (med-), and posterior (post-) parietal cortex in addition to

components in the cingulate cortex, ventral and dorsal striatum, and limbic areas [1]. Data driven ICA also identified these regions as relevant components, but also additional regions including the olfactory tubercle, globus pallidus and amygdala, though among additional regions found only components involving the amygdala reached the significance level to be included in the results of the DR analysis.

**2.3.2 Between-network connectivity analysis using Network Modeling**

Performing a between-network analysis on the basis of the DR results as described in FSLNets implies comparing consecutively the time series of network X as derived from ICA with the averaged time series of each of the other networks. In contrast, the seed-based approach compares a single time series signal from region X separately with the temporal signals from other regions. Hence, seed-based analysis is limited to the analysis of 'within-network' connectivity, while DR-based FSLNets allows modeling of connectivity between network components in addition. This allowed grouping of individual components into functional networks (Table 1). ICAs were grouped together based on the FSLNets derived hierarchical clustering. We used a similar nomenclature as used by [12] in order to keep the uniformity of reported networks. Five major networks have been identified: the default mode network (DMN) comprising cingulate cortex and hippocampus, the lateral cortical network (LCN) with somatosensory, secondary somatosensory, motor, and insular cortices, the associative cortical network (ACN) including auditory cortex, the subcortical network (SuCN) with piriform cortex, ventral striatum, and amygdala, and the thalamic network (ThN) comprising dorsal and ventral thalami.

The strengths of the connectivity between different network components was analyzed using network box plots as illustrated for the connectivity between S1-vTh and S2-vTh under isoflurane- vs. medetomidine-anesthesized mice (Figure 3). Thalamocortical interaction is completely suppressed in isoflurane-anesthetized animals, whereas under medetomidine anesthesia a weak negative thalamocortical correlation is observed. Analogous analyses have been carried out for all possible interactions among ICA components and for all anesthetic regimens, and the results displayed in the form of interaction matrices highlighting interactions found to be significant (Figure 4a; Suppl. Table 1 shows the connectivity values under different anesthetics between the selected ICA components obtained through boxplots).
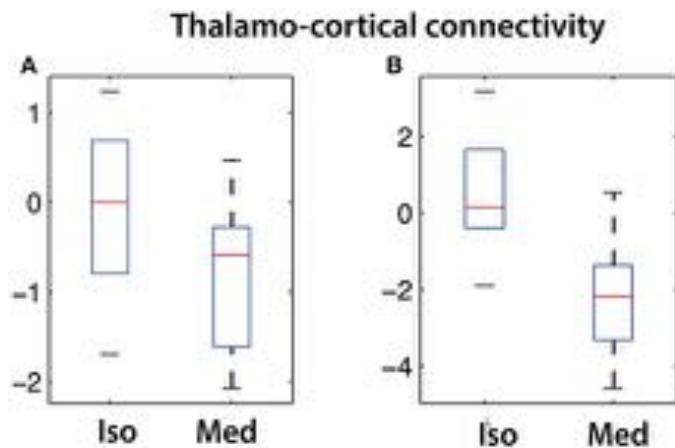
**Figure 3**: Cortico-thalamic connectivity differences in isoflurane vs. medetomidine anesthetized mice. Connectivity strength is shown for the network between (a) somatosensory cortex (S1) and ventral thalamus (vTh) and (b) between secondary somatosensory cortex (S2) and ventral thalamus (vTh). Under isoflurane anesthesia these connectivities are largely suppressed, while a significant negative correlation has been found for medetomidine anesthetized mice.
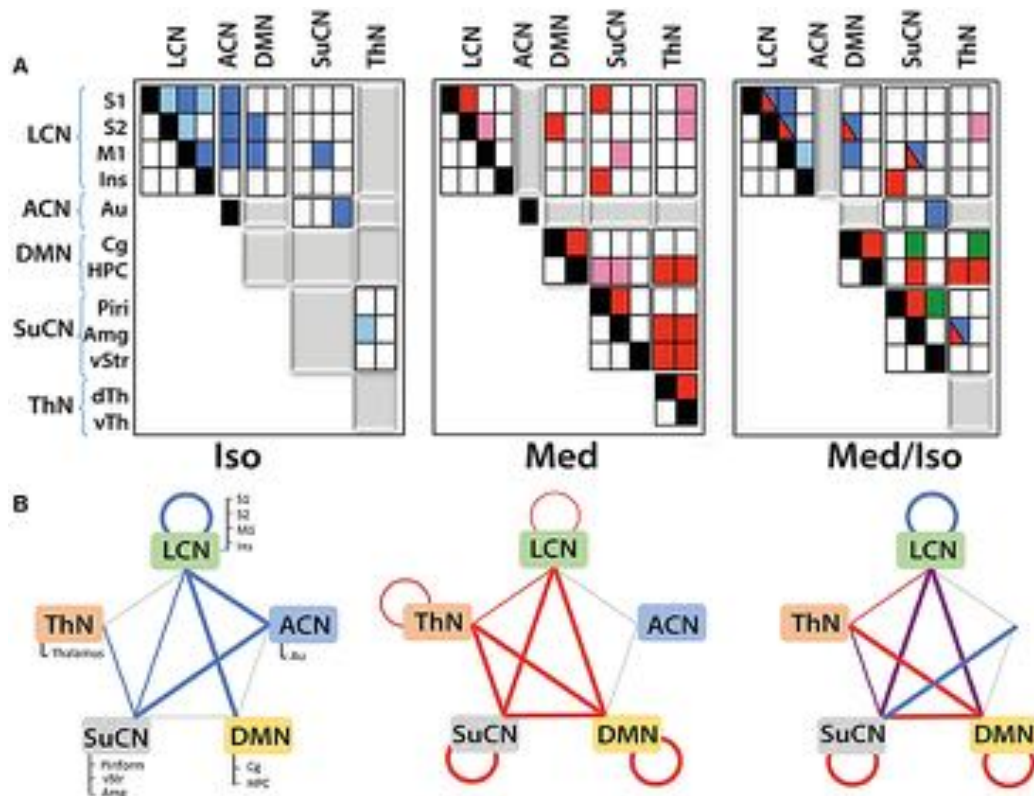


**Figure 4:** Functional networks and their interaction as derived from DR analysis: (a) Matrices displaying within- and between-network interactions under the various

anesthesia regimens. Matrices have been structured according to the five functional networks identified. Colors indicate significant interactions observed under isoflurane (blue) or medetomidine anesthesia (red). For the group receiving the combination anesthesia, elements are indicated in different colors depending on whether they apparently arise from the isoflurane group (blue), from the medetomidine group (red), from both groups (blue/red) or were observed in the combination group exclusively (green). Positive correlations are indicated by dark colors, while light colors represent negative correlations. Gray blocks indicate the absence of any significant interaction. Significant group differences (with Bonferroni correction) in connectivity have been observed for all the interactions displayed. (b) Within- and between-network interactions detected in isoflurane (blue lines), medetomidine (red lines) and medetomidine/ isoflurane anesthetized mice. In isoflurane/medetomidine anesthetized mice network interactions, blue lines represent interactions observed in mice under isoflurane anesthesia only, red under medetomidine only, and violet represents interactions observed for both anesthesia regimes. The width of the lines in the figure indicates the strength of the connectivity.

Mice under isoflurane anesthesia displayed strong connectivity among LCN components, which was found to be less pronounced in medetomidine-anesthetized animals. In contrast the occurrence of interactions involving subcortical structures was a characteristic for medetomidine anesthesia. This included inter-thalamic connectivity as well as networks involving thalamocortical connections, both of which did not display significant functional connectivity under isoflurane anesthesia (Figure 4b). On the other hand, DMN-cortical connectivity was preserved under isoflurane anesthesia but not under medetomidine (Figure 4a, middle panel). Given the complementary nature of networks observed under these two anesthetics, Grandjean et al. [1] suggested the combination of the two as potentially attractive regimen.

The network interaction patterns observed when using the combination medetomidine/isoflurane can be largely represented as the superposition of the network interaction patterns obtained for each anesthetic alone (Figure 4a, 4b), with some deviations. All connections between cortical and subcortical structures observed under medetomidine were preserved for the combination regime except some interactions within the thalamic and the subcortical network. Also, the combination anesthesia displayed the strong intra-cortical networks observed for isoflurane but not for medetomidine. As Grandjean et al. [1] included propofol and urethane anesthetized

mice in their study, the corresponding results of the network analysis for these two anesthetics have been compared to that obtained for isoflurane anesthetized animals (Supplementary Figure 4).

The schemes described in Figure 4b capture the differential nature of network interaction only at a high level. When analyzing the interactions within and between modules in more detail, additional differences among the various anesthesia regimens become apparent (Figure 4a). For example, while both isoflurane and medetomidine anesthetized mice display within-LCN interactions the nature of these interactions is different: for the isoflurane group significant interactions between ICA components S1-S2, S1-M1, S1-Ins, S2-M1, M1-Ins were observed, while for medetomidine only two of these interactions (S1-S2, S2-M1) were found to be significant. For the combination regime medetomidine/isoflurane all interactions observed under isoflurane alone were found significant with the exception of S1-Ins. The connectivity within SuCN was found absent for isoflurane anesthetized mice, but was found under medetomidine or medetomidine/isoflurane anesthesia. Inter-thalamic network interactions have been observed in medetomidine but not in isoflurane and medetomidine/isoflurane anesthetized animals. Apart from inter-thalamic network connectivity, all the other interactions involving the thalamic network found under medetomidine were preserved in the combination regime. Similarly, the DMN was also found to be more functionally connected to other networks in medetomidine and medetomidine/isoflurane combination regime than in isoflurane only. Some connectivity patterns have been observed exclusively in the medetomidine/isoflurane combination regime, such as the connection between ventral striatum and piriform cortex.

## 2.4 Discussion

The vast majority of rodent fMRI studies involve the use of anesthesia, which inevitably interferes with brain function and may confound effects of interest unrelated to the anesthetic effects. Hence, understanding the effects imposed by the anesthetic regimen is essential for proper analysis of fMRI data, in particular information on functional connectivity across the brain. It may allow identifying anesthesia-specific network signatures, which might then be accounted for during further data analysis. The effect of anesthesia on functional connectivity in the rat and mouse brain has been investigated by several groups [10; 17; 19; 20; 32; 33; 34; 35]. Grandjean et. al. [1]

compared the effect of four different anesthetics on functional connectivity in the mouse brain using seed-based analysis and reported characteristic changes depending on the type of anesthetic used. In particular, functional connectivity patterns recorded under medetomidine differed from those observed for animals under isoflurane, propofol or urethane anesthesia. The current study using DR analysis confirmed these results for isoflurane and medetomidine and identified additional nodes/ brain regions to be included in the anesthesia-specific signature. DR based network analysis including the analysis of between-network interactions is arguably a more comprehensive depiction of 'systems-level' activity/connectivity in the brain, in particular with networks derived from a data-driven approach such as ICA, and thus might unveil new knowledge about brain systems where prior hypotheses are unclear. Functional connectivity between nodes could be either direct or indirect, i.e. relayed via another node. While correlation is a mere measure of functional connectivity, irrespective of its nature, PC analysis reveals direct connectivity exclusively [9]. While Grandjean et al. [1] have used FC analysis to identify regions displaying temporal signal profiles with high correlation to a seed region, in this work we employed PC measures to focus primarily on networks based on 'direct' functional connectivity. An objective of network analysis is to identify nodes connected by direct connectivity (edges) and eliminate spurious edge effects due to an indirect third region in-between. It is important to note that direct connectivity does not imply monosynaptic connections. In fact, the structural correlate for direct FC can be both monosynaptic and polysynaptic. Our results reveal interesting within- and between-network interactions showing preserved intra- and inter-cortical interactions under isoflurane, subcortical interactions under medetomidine and superposition of these interactions under the combined anesthetics regimen.

DR and network analysis have mostly been applied to human studies so far with an exception of few recent reports in small animal rsfMRI [10; 11]. On the other hand, the mouse brain and in particular its cortical organization is considerably simpler and less subject to inter-individual variability, which should add consistency to the data. In fact, DR yielded reasonable, neurobiologically plausible results for mouse rs-fMRI data. For example, this is illustrated by the fact that even for deep-lying small structures such as the amygdala or the ventral striatum (nucleus accumbens), statistically significant results have been obtained across groups.

The network interactions observed in mice receiving the medetomidine/isoflurane combination anesthesia can be largely composed as a superposition of networks found under the isoflurane or medetomidine alone, with some deviations. There are, nevertheless, a few aspects of the different group results that deserve special attention. DMN-ThN connectivity has been described as a structural connection in mice [36]. In this study, we observed the functional links DMN-ThN and LCN-ThN in medetomidine, but not in isoflurane anesthetized mice (Figure 3), which clearly highlights the potential confounds linked to the use of anesthesia in functional brain imaging studies. The anesthesia-specific connectivity pattern might arise from the different molecular modes of action of the two anesthetics or differential effects on the cerebrovasculature. Isoflurane is an anesthetic, while medetomidine is a sedative with analgetic activity. The two compounds have different molecular modes of action interacting with either the GABAergic (isoflurane) or the a2 adrenergic system (medetomidine). A striking observation is the loss of cortico thalamic FC in isoflurane and to a certain degree also in medetomidine anesthetized mice. This may reflect the anesthetic efficacy of these drugs as loss of frontal-thalamic connectivity has been associated with loss of consciousness in humans [37] and rats [38]. The latter study demonstrated decreasing strength of this connection upon increasing the dose of the anesthetic. Along these lines, it has been demonstrated that light sedation with halothane [39] or medetomidine [21] preserved cortico-thalamic functional connectivity to some extent. Hence the observed differences in isoflurane and medetomidine anesthetized mice may reflect differences in anesthesia depth, i.e. the brain state. On the other hand, differences in the pharmacological mode of action and physiological activity (e.g. effects on the vascular tone) of isoflurane on medetomidine are likely to contribute to the differential responses. In addition, the two compounds exert rather opposing effects on the cerebrovascular system, isoflurane acting as vasodilator and medetomidine as vasoconstrictor, which may affect the translation of spontaneous neuronal activity into the BOLD signal assessed by fMRI. Interestingly, combining the two anesthetics at a low dose retained the interactions between DMN-ThN and LCN-ThN observed with medetomidine along with the intercortical interactions observed with isoflurane anesthesia, and thus constitutes an attractive anesthesia regimen for fMRI investigations in mice. Along similar lines, interactions within the SuCN between piriform cortex and amygdala were reliably detected in medetomidine and medetomidine/isoflurane anaesthetized mice, but not under isoflurane only. This functional connectivity pattern is supported by the observation of

structural connections between these regions [36]. The existence of a structural connectivity does not warrant functional connectivity as illustrated by the differential functional connectivity patterns for the different anesthetics. For example intracortical connectivity was found to be profoundly reduced in medetomidine as compared to isoflurane-anesthetized mice. This connectivity was found to be in part recovered when using the combination anesthesia, though the interaction remained weaker, in that a significant connectivity between LCN and ACN could not be detected anymore.

Mice under medetomidine anesthesia displayed anti-correlated functional connectivity between cortical structures and thalamus. There have been mixed reports on cortico-thalamic functional connectivity under anesthesia and it is widely debated in the literature. The preservation of thalamocortical activity under anesthesia has been reported previously in animals as well in humans [40; 41]. Boveroux et al. [42] reported similar anti-correlation during their study of propofol-induced unconsciousness in humans. An anti-correlated pattern of thalamus and cortical was also found in a study of rats with limbic seizures [43], in line with our results. On the other hand, some studies have shown no cortico-thalamic interaction during sedation [1; 44; 45; 46; 47; 48], while other studies have revealed diminished but detectable thalamo-cortical connectivity [49; 50]. It appears that the cortico-thalamic interaction is modulated by the type and depth of anesthesia, similar to response in other brain regions such as the frontal cortex displaying decreased activity in propofol and sevoflurane anesthesia in humans [51], a region differentially affected by anesthetics also in our study. Furthermore in [5], authors reported that DR outperforms seed based analysis. Despite putting the seeds in the same areas identified by ICA, the authors were not able to replicate the results from DR, while their results from DR analysis had been independently verified in a separate group of subjects. This might explain the inability to detect statistically significant cortico-thalamic interaction in medetomidine anesthetized mice in the previous study using seed based analysis.

The results get even more complex when analyzing the interactions at the level of the individual ICAs that constitute a network. While for all anesthetic regimen tested, connections between the major networks have been observed – with the exception of ACN, for which interactions have been only detected under isoflurane anesthesia – there is considerable variability regarding the network components responsible for these interactions. These differences may again reflect anesthesia specific connectivity patterns. Alternatively, the differences found for the various anesthesia regimens may

also arise from limitations in the statistical approaches, which – due to the small dimensions and correspondingly low SNR typically encountered in mouse fMRI – may lead to a significant finding for one but not for another anesthesia regime. An important limitation in fMRI studies assessing functional connectivity in anesthetized rodents is that data cannot be referred to the conscious baseline state. As a results, anesthesia induced changes in functional connectivity cannot be characterized. Nonetheless, FC pattern observed in anesthetized rodents have been found to correspond to patterns observed in awake humans [35]. In addition, analysis of functional connectivity patterns under different anesthetics may help to identify anesthesia induced alterations in rs-fMRI patterns in mice.

Despite these limitations, the results obtained in this study are consistent with previous findings in humans and other species [52; 53]. The modulatory effects of anesthetics on functional connectivity between the brain regions highlights the importance of analyzing fMRI responses to pharmacological or physiological intervention at the level of brain networks rather than analyzing changes in isolated brain regions, a holistic approach that is gaining increasing attention in the neuroscience community.
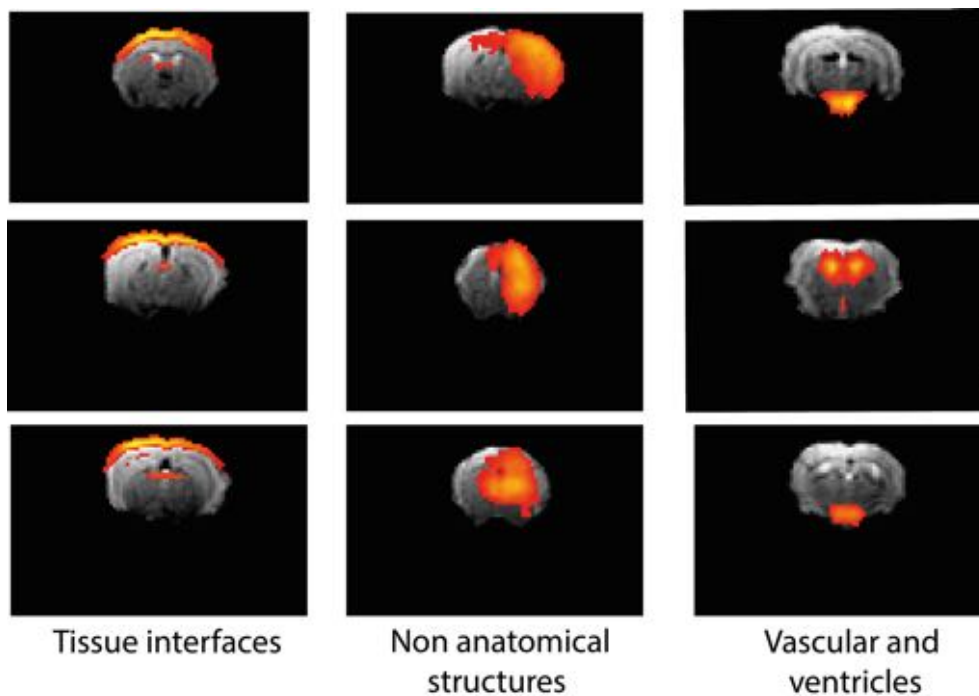
In conclusion, we have used DR in combination with data-driven ICA analysis to study the effects of different anesthetic regimen on brain functional networks in the mouse. Five basic networks have been identified, which display within- and between-network interactions that depend on the anesthetic used. While medetomidine preserves most of the intra- and inter-network connectivities, except those involving the ACN, the intra- and inter-cortical network interactions (LCN-LCN, LCN-ACN) are better retained in isoflurane-anesthetized mice. An important result is that the network interactions observed under the combination anesthesia medetomidine/isoflurane largely constitute the superposition of the interactions found for each anesthetic alone. Understanding the differential effects of anesthetics on brain functional networks in animals is relevant when analyzing changes induced by physiological stress or pathological conditions. Deeper understanding of the effect of an anesthetic on large-scale brain networks is also relevant for clinical research, as it may help with achieving safer yet maximally effective anesthetic protocols with minimum side effects.

## 2.5 Supplementary Material

**Suppl. Table 1**: Connectivity values between selected ICA components under isoflurane, medetomidine and medetomidine/isoflurane combined anesthetic regimens. The values were obtained from network box plots using unpaired t-test design matrix. Figures 4(a) and 4(b) displaying the network interaction matrix and graphs, respectively, is based on the quantitative information from these box plots.

|  | isoflurane | medetomidine | med/iso |
|---|---|---|---|
| S1 – S2 | -3.0 | 0.2 | 0.6 |
| S1 – M1 | 0.3 | 0 | 2.4 |
| S1 – Ins | -1.2 | 0 | 0 |
| S2 – M1 | -1.1 | -2 | 0.8 |
| M1 – Ins | 1.0 | 0 | -2 |
| Au – S1 | 0.5 | 0 | 0 |
| Au – S2 | 0.4 | 0 | 0 |
| Au – M1 | 0.8 | 0 | 0 |
| Cg – S2 | 1.5 | 1.3 | 1.1 |
| Cg – M1 | 1.2 | 0 | 0.7 |
| Amg – M1 | 1.4 | -1.3 | 0.5 |
| Piri – S1 | 0 | 0.4 | 0 |
| Piri – Ins | 0 | 1.8 | 0.1 |
| vTh – S1 | 0 | -0.7 | 0 |
| vTh – S2 | 0 | -2.1 | -0.4 |
| Au-vStr | 0.6 | 0 | 0.5 |
| Cg – HPC | 0 | 0.9 | 0.7 |
| HPC – Piri | 0 | -1.2 | 0 |
| HPC - Amg | 0 | -1.5 | 1.4 |

| | | | |
|---|---|---|---|
| HPC – dTh | 0 | 2.5 | 0.3 |
| HPC - vTh | 0 | 1.9 | 0.2 |
| Cg - Amg | 0 | 0 | 0.5 |
| Cg – vTh | 0 | 0 | 0.7 |
| Amg – dTh | -1.1 | 0.1 | 1.3 |
| Piri – Amg | 0 | 2.7 | 0.5 |
| Amg – vTh | 0 | 2.3 | 0 |
| vStr – dTh | 0 | 2.8 | 0 |
| vStr – vTh | 0 | 2.6 | 0 |
| Piri – vStr | 0 | 0 | 0.8 |
| dTh – vTh | 0 | 2 | 0 |



Tissue interfaces    Non anatomical    Vascular and
                     structures        ventricles

**Supplementary Figure 1:** Examples of ICA components removed from further analysis as they mainly represent vascular structures, ventricles, tissue interfaces or could not be clearly assigned to anatomical structures (see labels in figure).
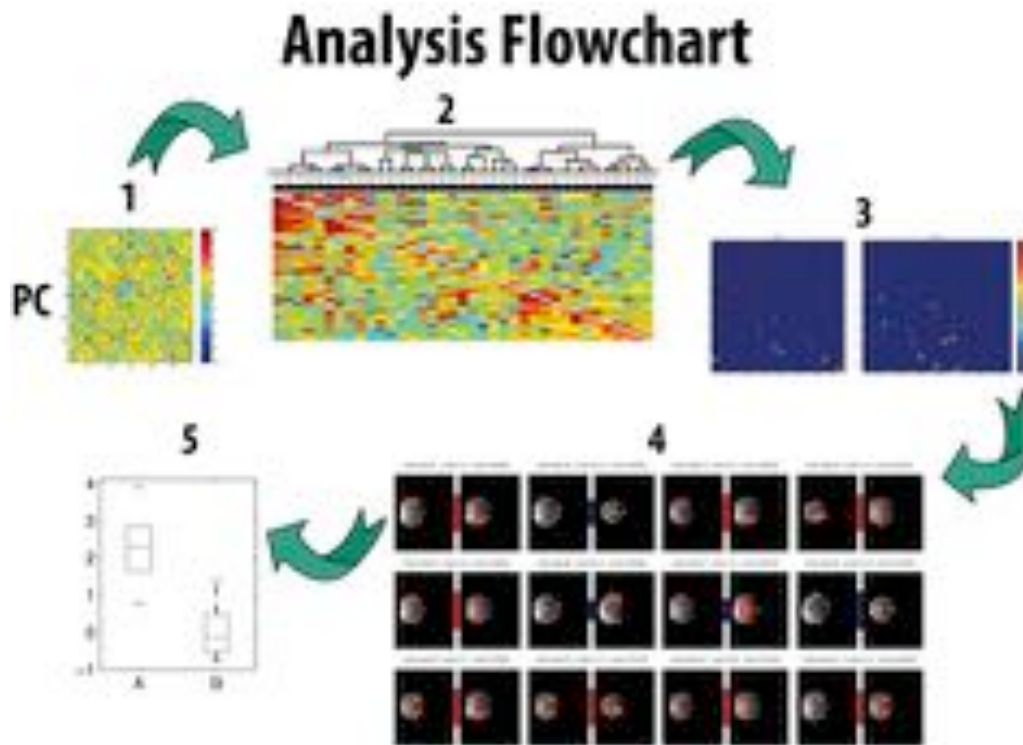
**Supplementary Figure 2**: Analysis Flowchart. The partial correlation matrix (1) is clustered in a subsequent step using hierarchical clustering algorithm (2) The input for hierarchical clustering were the time series in to hierarchical clustering, i.e. clustering is according to the functional similarity. Application of unpaired t-test yields enabled the identification of networks that displayed significant differences under the two anesthetic conditions A and B (3) as represented by network matrices containing corrected p-values for each contrast (A>B, A<B). For convenience, 1-p values are displayed. Elements below the diagonal are the 1-p values, elements above the diagonal indicate significance of the two-group t-test at corrected-p<0.05. The left matrix represents interaction with A>B and right matrix with B<A. The anatomical location of interacting networks components displaying significant differences is shown in (4), with red bars indicating positive and blue bars indicating negative correlation. The width of the bar indicates the strength of the correlation. The box plot (5) indicates the correlation coefficient for one specific network in condition A and B showing that the specific interaction was significant under A but not B.
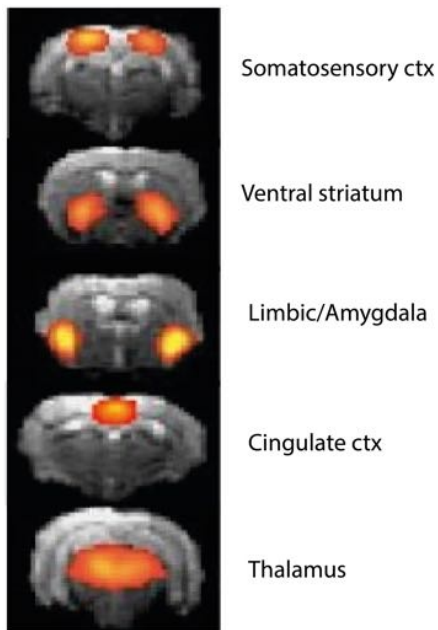
**Supplementary Figure 3:** ICA derived components illustrating somatosensory cortex, ventral striatum, limbic areas, cingulate cortex and thalamus. For comparison with analogous seed-based regions the reader is referred to the Grandjean et. al. [1], Fig1.
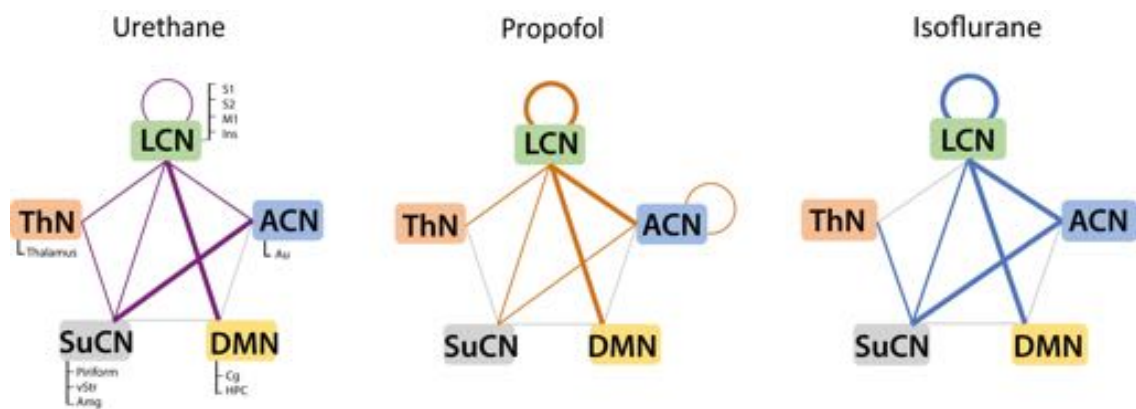


**Supplementary Figure 4:** Comparison of functional networks found for urethane- and propofol-anesthetized mice compared to that of isoflurane-anesthetized animals. There is a striking similarity of the network maps for the three anesthetics. Nevertheless, there are few differences, e.g. some of the interactions appear weakened in urethane- and propofol- compared to isoflurane-anesthetized mice.

## 2.6 References

[1] J. Grandjean, A. Schroeter, I. Batata, and M. Rudin, Optimization of anesthesia protocol for resting-state fMRI in mice based on differential effects of anesthetics on functional connectivity patterns. Neuroimage 102 (2014) 838-47.

[2] D.C. Van Essen, S.M. Smith, D.M. Barch, T.E. Behrens, E. Yacoub, K. Ugurbil, and W.U.-M.H. Consortium, The WU-Minn Human Connectome Project: an overview. Neuroimage 80 (2013) 62-79.

[3] M. Hampson, B.S. Peterson, P. Skudlarski, J.C. Gatenby, and J.C. Gore, Detection of functional connectivity using temporal correlations in MR images. Hum Brain Mapp 15 (2002) 247-62.

[4] D.M. Cole, S.M. Smith, and C.F. Beckmann, Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. Front Syst Neurosci 4 (2010) 8.

[5] D.V. Smith, A.V. Utevsky, A.R. Bland, N. Clement, J.A. Clithero, A.E. Harsch, R. McKell Carter, and S.A. Huettel, Characterizing individual differences in functional connectivity using dual-regression and seed-based approaches. Neuroimage 95 (2014) 1-12.

[6] N. Filippini, B.J. MacIntosh, M.G. Hough, G.M. Goodwin, G.B. Frisoni, S.M. Smith, P.M. Matthews, C.F. Beckmann, and C.E. Mackay, Distinct patterns of brain activity in young carriers of the APOE-epsilon4 allele. Proc Natl Acad Sci U S A 106 (2009) 7209-14.

[7] J.S. Damoiseaux, S.A. Rombouts, F. Barkhof, P. Scheltens, C.J. Stam, S.M. Smith, and C.F. Beckmann, Consistent resting-state networks across healthy subjects. Proc Natl Acad Sci U S A 103 (2006) 13848-53.

[8] X.N. Zuo, C. Kelly, J.S. Adelstein, D.F. Klein, F.X. Castellanos, and M.P. Milham, Reliable intrinsic connectivity networks: test-retest evaluation using ICA and dual regression approach. Neuroimage 49 (2010) 2163-77.

[9] S.M. Smith, K.L. Miller, G. Salimi-Khorshidi, M. Webster, C.F. Beckmann, T.E. Nichols, J.D. Ramsey, and M.W. Woolrich, Network modelling methods for FMRI. Neuroimage 54 (2011) 875-91.

[10] J. Grandjean, R. Derungs, L. Kulic, T. Welt, M. Henkelman, R.M. Nitsch, and M. Rudin, Complex interplay between brain function and structure during cerebral amyloidosis in APP transgenic mouse strains revealed by multi-parametric MRI comparison. Neuroimage (2016).

[11] M.J. Henckens, K. van der Marel, A. van der Toorn, A.G. Pillai, G. Fernandez, R.M. Dijkhuizen, and M. Joels, Stress-induced alterations in large-scale functional networks of the rodent brain. Neuroimage 105 (2015) 312-22.

[12] A. Liska, A. Galbusera, A.J. Schwarz, and A. Gozzi, Functional connectivity hubs of the mouse brain. Neuroimage 115 (2015) 281-91.

[13] A.E. Mechling, N.S. Hubner, H.L. Lee, J. Hennig, D. von Elverfeldt, and L.A. Harsan, Fine-grained mapping of mouse brain functional connectivity with resting-state fMRI. Neuroimage 96 (2014) 203-15.

[14] V.C. Austin, A.M. Blamire, K.A. Allers, T. Sharp, P. Styles, P.M. Matthews, and N.R. Sibson, Confounding effects of anesthesia on functional activation in rodent brain: a study of halothane and alpha-chloralose anesthesia. Neuroimage 24 (2005) 92-100.

[15] J.M. Adamczak, T.D. Farr, J.U. Seehafer, D. Kalthoff, and M. Hoehn, High field BOLD response to forepaw stimulation in the mouse. Neuroimage 51 (2010) 704-12.

[16] M. Fukuda, A.L. Vazquez, X. Zong, and S.G. Kim, Effects of the alpha(2)-adrenergic receptor agonist dexmedetomidine on neural, vascular and BOLD fMRI responses in the somatosensory cortex. Eur J Neurosci 37 (2013) 80-95.

[17] A. Schroeter, F. Schlegel, A. Seuwen, J. Grandjean, and M. Rudin, Specificity of stimulus-evoked fMRI responses in the mouse: the influence of systemic physiological changes associated with innocuous stimulation under four different anesthetics. Neuroimage 94 (2014) 372-84.

[18] K.A. Williams, M. Magnuson, W. Majeed, S.M. LaConte, S.J. Peltier, X. Hu, and S.D. Keilholz, Comparison of alpha-chloralose, medetomidine and isoflurane anesthesia for functional connectivity mapping in the rat. Magn Reson Imaging 28 (2010) 995-1003.

[19] E. Jonckers, D. Shah, J. Hamaide, M. Verhoye, and A. Van der Linden, The power of using functional fMRI on small rodents to study brain pharmacology and disease. Front Pharmacol 6 (2015) 231.

[20] E. Jonckers, J. Van Audekerke, G. De Visscher, A. Van der Linden, and M. Verhoye, Functional connectivity fMRI of the rodent brain: comparison of functional connectivity networks in rat and mouse. PLoS One 6 (2011) e18876.

[21] F.A. Nasrallah, S.K. Lew, A.S. Low, and K.H. Chuang, Neural correlate of resting-state functional connectivity under alpha2 adrenergic receptor agonist, medetomidine. Neuroimage 84 (2014) 27-34.

[22] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17 (2002) 825-41.

[23] M. Jenkinson, and S. Smith, A global optimisation method for robust affine registration of brain images. Med Image Anal 5 (2001) 143-56.

[24] C.F. Beckmann, and S.M. Smith, Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imaging 23 (2004) 137-52.

[25] C.F. Beckmann, M. DeLuca, J.T. Devlin, and S.M. Smith, Investigations into resting-state connectivity using independent component analysis. Philos Trans R Soc Lond B Biol Sci 360 (2005) 1001-13.

[26] C.F. Beckmann, and S.M. Smith, Tensorial extensions of independent component analysis for multisubject FMRI analysis. Neuroimage 25 (2005) 294-311.

[27] I.M. Veer, C.F. Beckmann, M.J. van Tol, L. Ferrarini, J. Milles, D.J. Veltman, A. Aleman, M.A. van Buchem, N.J. van der Wee, and S.A. Rombouts, Whole brain resting-state analysis reveals decreased functional connectivity in major depression. Front Syst Neurosci 4 (2010).

[28] T.E. Nichols, and A.P. Holmes, Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15 (2002) 1-25.

[29] A.M. Winkler, G.R. Ridgway, M.A. Webster, S.M. Smith, and T.E. Nichols, Permutation inference for the general linear model. Neuroimage 92 (2014) 381-97.

[30] S.M. Smith, and T.E. Nichols, Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44 (2009) 83-98.

[31] L. Tian, Y. Kong, J. Ren, G. Varoquaux, Y. Zang, and S.M. Smith, Spatial vs. Temporal Features in ICA of Resting-State fMRI - A Quantitative and Qualitative Investigation in the Context of Response Inhibition. PLoS One 8 (2013) e66572.

[32] I. Nasu, N. Yokoo, S. Takaoka, K. Takata, T. Hoshikawa, M. Okada, and Y. Miura, The dose-dependent effects of isoflurane on outcome from severe forebrain ischemia in the rat. Anesth Analg 103 (2006) 413-8, table of contents.

[33] K. Masamoto, M. Fukuda, A. Vazquez, and S.G. Kim, Dose-dependent effect of isoflurane on neurovascular coupling in rat cerebral cortex. Eur J Neurosci 30 (2009) 242-50.

[34] D.V. D'Souza, E. Jonckers, A. Bruns, B. Kunnecke, M. von Kienlin, A. Van der Linden, T. Mueggler, and M. Verhoye, Preserved modular network organization in the sedated rat brain. PLoS One 9 (2014) e106156.

[35] W.J. Pan, J.C. Billings, J.K. Grooms, S. Shakil, and S.D. Keilholz, Considerations for resting state functional MRI and functional connectivity studies in rodents. Front Neurosci 9 (2015) 269.

[36] S.W. Oh, J.A. Harris, L. Ng, B. Winslow, N. Cain, S. Mihalas, Q. Wang, C. Lau, L. Kuan, A.M. Henry, M.T. Mortrud, B. Ouellette, T.N. Nguyen, S.A. Sorensen, C.R. Slaughterbeck, W. Wakeman, Y. Li, D. Feng, A. Ho, E. Nicholas, K.E. Hirokawa, P. Bohn, K.M. Joines, H. Peng, M.J. Hawrylycz, J.W. Phillips, J.G. Hohmann, P. Wohnoutka, C.R. Gerfen, C. Koch, A. Bernard, C. Dang, A.R. Jones, and H. Zeng, A mesoscale connectome of the mouse brain. Nature 508 (2014) 207-14.

[37] O. Akeju, M.L. Loggia, C. Catana, K.J. Pavone, R. Vazquez, J. Rhee, V. Contreras Ramirez, D.B. Chonde, D. Izquierdo-Garcia, G. Arabasz, S. Hsu, K. Habeeb, J.M. Hooker, V. Napadow, E.N. Brown, and P.L. Purdon, Disruption of thalamic functional connectivity is a neural correlate of dexmedetomidine-induced unconsciousness. Elife 3 (2014) e04499.

[38] Z. Liang, X. Liu, and N. Zhang, Dynamic resting state functional connectivity in awake and anesthetized rodents. Neuroimage 104 (2015) 89-99.

[39] F. Sforazzini, A.J. Schwarz, A. Galbusera, A. Bifone, and A. Gozzi, Distributed BOLD and CBV-weighted resting-state networks in the mouse brain. Neuroimage 87 (2014) 403-15.

[40] A. Silva, H. Cardoso-Cruz, F. Silva, V. Galhardo, and L. Antunes, Comparison of anesthetic depth indexes based on thalamocortical local field potentials in rats. Anesthesiology 112 (2010) 355-63.

[41] R.N. Mhuircheartaigh, D. Rosenorn-Lanng, R. Wise, S. Jbabdi, R. Rogers, and I. Tracey, Cortical and subcortical connectivity changes during decreasing levels of consciousness in humans: a functional magnetic resonance imaging study using propofol. J Neurosci 30 (2010) 9095-102.

[42] P. Boveroux, A. Vanhaudenhuyse, M.A. Bruno, Q. Noirhomme, S. Lauwick, A. Luxen, C. Degueldre, A. Plenevaux, C. Schnakers, C. Phillips, J.F. Brichant,

V. Bonhomme, P. Maquet, M.D. Greicius, S. Laureys, and M. Boly, Breakdown of within- and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. Anesthesiology 113 (2010) 1038-53.

[43] D.J. Englot, B. Modi, A.M. Mishra, M. DeSalvo, F. Hyder, and H. Blumenfeld, Cortical deactivation induced by subcortical network dysfunction in limbic seizures. J Neurosci 29 (2009) 13006-18.

[44] F. Zhao, T. Zhao, L. Zhou, Q. Wu, and X. Hu, BOLD study of stimulation-induced neural activity and resting-state connectivity in medetomidine-sedated rat. Neuroimage 39 (2008) 248-60.

[45] N.S. White, and M.T. Alkire, Impaired thalamocortical connectivity in humans during general-anesthetic-induced unconsciousness. Neuroimage 19 (2003) 402-11.

[46] M.T. Alkire, R.J. Haier, and J.H. Fallon, Toward a unified theory of narcosis: brain imaging evidence for a thalamocortical switch as the neurophysiologic basis of anesthetic-induced unconsciousness. Conscious Cogn 9 (2000) 370-86.

[47] G.A. Mashour, and M.T. Alkire, Consciousness, anesthesia, and the thalamocortical system. Anesthesiology 118 (2013) 13-5.

[48] X. Liu, K.K. Lauer, B.D. Ward, S.J. Li, and A.G. Hudetz, Differential effects of deep sedation with propofol on the specific and nonspecific thalamocortical systems: a functional magnetic resonance imaging study. Anesthesiology 118 (2013) 59-69.

[49] Z. Liang, J. King, and N. Zhang, Intrinsic organization of the anesthetized brain. J Neurosci 32 (2012) 10183-91.

[50] S.P. Kim, E. Hwang, J.H. Kang, S. Kim, and J.H. Choi, Changes in the thalamocortical connectivity during anesthesia-induced transitions in consciousness. Neuroreport 23 (2012) 294-8.

[51] K.K. Kaisti, L. Metsahonkala, M. Teras, V. Oikonen, S. Aalto, S. Jaaskelainen, S. Hinkka, and H. Scheinin, Effects of surgical levels of propofol and sevoflurane anesthesia on cerebral blood flow in healthy subjects studied with positron emission tomography. Anesthesiology 96 (2002) 1358-70.

[52] W.D. Winters, Effects of drugs on the electrical activity of the brain: anesthetics. Annu Rev Pharmacol Toxicol 16 (1976) 413-26.

[53] A.G. Hudetz, General anesthesia and human brain connectivity. Brain Connect 2 (2012) 291-302.

## 3. Increasing isoflurane dose reduces homotopic correlation and functional segregation of brain networks in mice as revealed by resting-state fMRI

**Bukhari Q[1], Schroeter A[1] and Rudin M[1, 2] ***

[1] Institute for Biomedical Engineering, ETH Zurich and University of Zurich, Zurich, Switzerland

[2] Institute of Pharmacology and Toxicology, University of Zurich, Zurich, Switzerland

**Abstract**

Effects of anesthetics on brain functional networks are not fully understood. In this work, we investigated functional brain networks derived from resting-state fMRI data obtained under different doses of isoflurane in mice using stationary and dynamic functional connectivity (dFC) analysis. Stationary network analysis using FSL Nets revealed a clear modular structure of functional networks, which could be segregated into a lateral cortical, an associative cortical, the default mode network, a subcortical network, and a thalamic network. Increasing isoflurane dose led to a loss of functional connectivity between the bilateral cortical regions. In addition, dFC analysis revealed a dominance of dynamic functional states (dFS) exhibiting pronounced modular structure in mice anesthetized with a low dose of isoflurane, while at high isoflurane levels dFS showing widespread unstructured correlation displayed highest weights. This indicates that spatial segregation across brain functional networks is lost with increasing dose of the anesthetic drug used. This loss might be indicative of a state of deep anesthesia. Combining the results of stationary and dynamic FC analysis indicates that increasing isoflurane levels leads to loss of modular network organization, which includes loss of the strong bilateral interactions between homotopic brain areas.

### 3.1 Introduction

Characteristics of general anesthesia are loss of sensation and analgesia, loss of muscle control (muscle relaxation) and eventually loss of consciousness. These effects are caused by a temporary change in the brain activity state induced by the anesthetic drug, i.e. changes in local activity pattern and in the activity of functional networks that depend both on the type and the concentration of the anesthetic. How these effects

on the concerted neural activity across the brain translate into the physiological characteristics of anesthesia, e.g. what kind of network changes relate to analgesia or loss of consciousness, is currently not known. EEG has been extensively used for assessing direct neural effects of anesthetic drugs [1; 2; 3], important for guiding their administration. Yet, spatial resolution of EEG recordings is poor and information regarding the interference of pharmacological interventions including anesthesia with distributed cerebral processing (functional networks) is problematic to extract.

Analysis of brain function in the absence of a specific stimulation paradigm using functional magnetic resonance imaging (so-called resting-state fMRI, rs-fMRI) has gained tremendous momentum in recent years as it allows identifying functional networks on the basis of the temporal correlation of the signal across brain regions [4; 5]. Typically correlation analysis compares signal fluctuations over time intervals of several minutes, which allows identifying 'stationary' coherent clusters, which are considered to constitute functional networks. More recently, it has been found that these stationary networks represent temporal integrals of dynamic processes occurring at a much faster time scale. A number of approaches have been suggested to study the dynamic aspects of functional connectivity (dFC). For example, Leonardi et al. [6; 7] have described a dFC algorithm with time varying windows using a dictionary learning approach, which yielded a number of robust dynamic functional states (dFS), the contributions of which to the overall activity fluctuate in the course of a rs-fMRI time series. The method has been recently transferred to mouse rs-fMRI [8] These authors could identify highly structured dFS reflecting interactions within and across major cerebral networks in the mouse, which fluctuate over time warranting the involvement of many brain areas in information processing. dFC analysis has the potential to reveal important insights that might remain hidden when analyzing stationary FC.

Changes in functional network imposed by anesthesia have been investigated both in humans and animals [9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 20]. A recent study described dose dependent effects of isoflurane on the stationary cerebral cortical networks in rats [21]. The authors reported that at isoflurane doses higher than 1.5% interhemispheric cortical FC strength was found decreased or completely suppressed, though there was no information to what extent dFSs were affected. The latter point was addressed in a study using macaque monkeys [22], in which high isoflurane levels were found to decrease the number of dFSs. However, the authors did not investigate

the changes in the relative weights of dFS in response to alterations in the level of isoflurane.

In this paper, we investigated effects of increasing the isoflurane dose on the functional networks in mice considering both stationary as well as dynamic aspects. In particular, we were interested whether increasing the isoflurane dose affected the nature of the dFS, and whether the individual dFS showed a differential sensitivity to anesthesia depth. Mice are attractive in this context as the wide range of genetically engineered strains may allow investigating mechanistic aspects underlying the various physiological characteristics of anesthesia such as analgesia, muscle relaxation, loss of consciousness.

## 3.2 Methods

### 3.2.1 Animals, preparation, and anesthesia

The experiments were performed in compliance with Swiss laws on animal protection and approved by the Veterinary Office of the Canton of Zurich. Female C57BL/6 mice (Janvier, Le Genest-St Isle, France) between 10 and 15 weeks old were studied. All mice were initially anesthetized with isoflurane in a 20% $O_2$ / 80% air mixture: 3.5% for induction, 2% for endotracheal intubation and during set-up on the animal bed. Throughout the duration of the experiment, animals were mechanically ventilated using a small animal ventilator (CWE, Ardmore, USA) with a 20% $O_2$ / 80% air mixture at a rate of 80 breaths/min, a respiration cycle of 25% inhalation, 75% exhalation, and an inspiration volume of 1.8 ml/min. The head was placed with the animal's incisors secured over a bite bar and fixated by ear bars, ophthalmic ointment was applied to the eyes, and a rectal temperature probe was inserted to keep the animal at 36.5 ± 0.5 °C by means of a warm-water circuit integrated into the animal holder (Bruker Biospin GmbH, Ettlingen, Germany). The tail vein was cannulated for intravenous (i.v.) administration of anesthetics and the neuromuscular blocking agent pancuronium bromide (Sigma-Aldrich, Steinheim, Germany).

Two independent set of studies were performed to evaluate the effects of isoflurane in a dose dependent manner.

Group 1 (Dose-escalation in individual mice): Twelve animals were used in the experiment. Isoflurane (Abbott, Cham, Switzerland) was sequentially increased from 1.1% to 1.3%, 1.5% and 2.0% in a 20% $O_2$ / 80% air mixture for each individual mouse. After each incremental increase of isoflurane concentration, there was a 10

min interval for equilibration before the fMRI data acquisition was started. Mice remained in the MR scanner throughout the duration of the experiment.

Group 2 (Single dose for each mouse): Isoflurane was administered in a 20% $O_2$ / 80% air mixture at a single defined dose per mouse in order to avoid any accumulation effects. Doses used were 1.1% (N=10 mice), 1.2% (N=19), and 1.5% (N=18).

Each animal received an i.v. bolus injection of 0.5 mg/kg pancuronium bromide dissolved in saline (0.5 mg/3 ml) followed by a continuous infusion of 0.5 mg/kg/h of pancuronium bromide corresponding to an infusion rate of the solution of 3 ml/kg/h.

Animal preparation, anesthesia protocols, and the conditions during resting-state measurements were identical for the fMRI experiments and for the assessment of systemic physiological parameters. Approximately 20 min were used for animal preparation, and a further 20 min for preparatory MRI scans. Subsequently, rs-fMRI data sets of 6 min duration each were acquired. After the experiments, time for recovery from anesthesia and pancuronium bromide administration was provided for all the animals.

### 3.2.2 fMRI

MRI/fMRI experiments were carried out using a Bruker Biospec 94/30 small animal MR system (Bruker BioSpin MRI, Ettlingen, Germany) operating at 400 MHz (9.4 T). A four-element receive-only cryogenic phased array coil (Bruker BioSpin AG, Faellanden, Switzerland) was used in combination with a linearly polarized room temperature volume resonator for transmission (Bruker BioSpin MRI, Ettlingen, Germany). Anatomical images acquired in the sagittal and horizontal direction allowed exact positioning of 12 adjacent coronal slices of 0.5 mm slice thickness, which were used for the rs-fMRI scans. A gradient-echo echo-planar imaging (GE-EPI) sequence has been used for rs-fMRI data acquisition with field of view=16x7 mm2, matrix dimensions=80x35, TR=1 s, TE=12 ms, flip angle=60 degrees. The time series acquired was of 360 s length.

### 3.2.3 Measurement of systemic physiological parameters

For physiological parameter measurement, the left hind limb of the mouse was shaved and a fiberoptic pulse oximeter (MouseOx, STARR Life Science, Oakmont, USA) fixed to the flank in order to record heart rate (in beats per minute, bpm), pulse distention (in μm), and oxygen saturation (in %).

### 3.2.4 Data Processing

All the pre-processing was performed using tools from FMRIB's Software Library (FSL version 5). FSL's recommended pre processing pipeline was used. Motion correction, removal of non-brain structures, high pass temporal filtering with sigma = 75.0 s; pre-whitening and global spatial smoothing of 0.2 mm was applied as part of the pre-processing.

After the pre-processing, the functional scans were aligned to the high-resolution anatomical QBI (Queensland Brain Institute) template using linear affine and nonlinear diffeomorphic transformation registration as implemented in ANTs (ANTs. v 1.9; http://picsl.upenn.edu/ANTS/).

We used FSL's MELODIC for probabilistic independent component analysis [23]. The multi-session temporal ICA concatenated approach, as recommended for rs-fMRI data analysis, allowed to input all subjects from all the groups in a temporally concatenated fashion for the ICA analysis. ConcatICA yielded different activations and artifact components without the need of specifying any explicit time series model.

A total of 50 independent components (IC maps) were extracted and the mixture model approach was applied on these estimated maps to perform for inference analysis. An alternative hypothesis test based on fitting a Gaussian/gamma mixture model to the distribution of voxel intensities within spatial maps [24; 25] was used to threshold the IC maps. A threshold of 0.5 (p < 0.5) was selected for the alternative hypothesis in order to assign equal 'cost' to false-positives and false-negatives. Out of the 50 independent components (IC maps), only 25 numbers of components were selected, while the components that overlapped with vascular structures and ventricles were excluded from further analysis. Similarly, regions at the brain surface, which are prone to be affected by the motion artifacts due to e.g. breathing, were excluded.

Dual Regression (FSL 5.0.2.2) was used for between-subject analysis allowing for voxel-wise comparisons of rs-fMRI [26; 27]. Dual regression is a technique that first regresses z-score group-IC maps (group spatial maps) into subject specific 4D resampled datasets to give a set of subject-specific variance normalized time courses for each component separately, and then regresses these time-courses into the same 4D dataset to get a subject-specific set of spatial maps. We used dual regression to generate the subject specific set of spatial maps from IC components.

Non-parametric permutation based inference [28] was performed with the subject-

specific component spatial maps concatenated across subjects for each analysis and submitted to voxel-wise between-subject analysis to test for effects of anesthetics dose dependency on functional connectivity using FSL-randomise [29]. Contrasts were set up using FSL's general linear model (GLM) and 5000 randomized permutations were run as the FSL default setting. We used threshold-free cluster enhancement [30] for statistical inference to validate the likelihood of extended areas of signal, which also takes into account information from neighboring voxels and combines the quality of both conventional cluster-based as well as voxel-based thresholding [30]. Correction for multiple comparisons across space was applied assuming an overall significance of $\alpha$ (p<0.05) using permutation testing and TFCE. Bonferroni correction was applied separately to each analysis depending on the number of components of interest.

FSL Nets (FSLNets v0.6) was used for estimating the network model of rs-fMRI data. The partial correlation matrices of the BOLD signal time courses of each component from dual regression were then clustered to form a dendrogram. These clusters were used as input in to the GLM analysis and rsun through FSL-randomise [29] to perform 5000 permutations to test for statistical significance. Edges, i.e. connections between network nodes showing statistically significant differences between the groups under consideration were obtained from GLM analysis. These significant network edges were then used to calculate the network box plots that take into account each edge and provide more information on difference in connectivity values between the groups. We applied FDR with multiple corrections method using the same unpaired t-test design matrix as used previously for DR analysis.

### 3.2.5 Dynamic Functional Connectivity analysis

Dynamic functional connectivity networks were evaluated using the dictionary learning approach [6; 7; 31]. Time courses were extracted from the ICA maps and fed into the dFC algorithm. Simple building blocks (referred as 'atoms') of whole brain connectivity were estimated using dictionary learning algorithm [31] with 30 folds each made of 200 iterations. Building blocks (atoms) were estimated for all anesthetic doses and all grouped together in a concatenated fashion, similar to ICA concatenation approach, and then were further regressed from the building blocks estimated from each anesthetic dose, similar to the dual regression approach. Atoms estimated were energy bounded. The algorithm was run iteratively many folds and matched to the first fold using Hungarian algorithm with spatial correlation as similarity measure in order

to produce robust dictionary learning atoms or dynamic functional connectivity states (dFS). The patterns extracted from the algorithm explained more than 50% of the variance in the sliding window correlation matrix. The atoms generated by the algorithm are the transient states of functional connectivity and therefore it is important to perform statistical tests in order to find group differences. In order to compare the transient states across different doses, we fixed their sequence to assure comparability across the study. A two-sample t-test was performed to statistically validate the results of the dynamic functional connectivity analysis.

## 3.3 Results

### 3.3.1 Static functional connectivity

Dual regression and network analysis using FSL Nets revealed static functional connectivity between homotopic regions in the two hemispheres. Network analysis performed for selected ICs using FSL Nets revealed a significant loss of functional connectivity between the homotopic regions of the two hemispheres upon increasing the dose of isoflurane (Fig. 1).
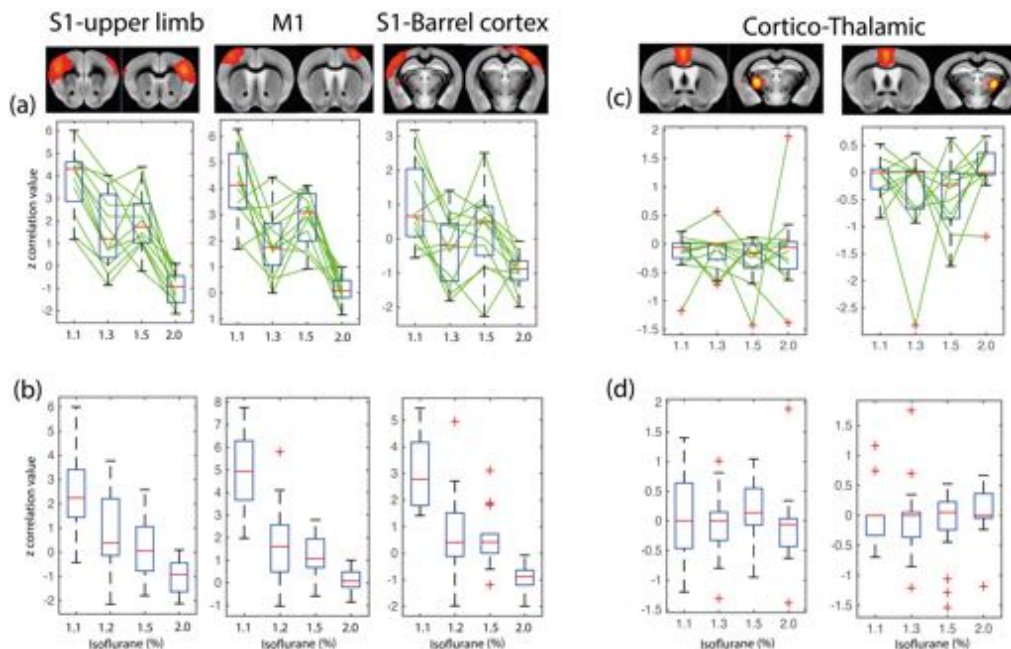


**Figure 1: Loss of FC between ICs upon increasing isoflurane dose.** (a) Loss of connectivity between bilateral homotopic regions in individual animals upon increasing dose of isoflurane (group 1) for three selected cortical ICs. Overlay images show cortical regions of interest; diagrams changes in z-values, with boxplots

75

indicating mean and the standard deviation (mean±SD) and green lines show changes in individual mice. The animals were kept in the MR scanner as isoflurane dose was increased. The p-values for each comparison are given in Suppl. Table 1. (b) Changes in individuals when exposing each animal to a single dose of isoflurane (group 2). (c) Absence of connectivity between cingulate cortex and thalamus upon increasing the isoflurane dose in individual mice of group 1. (d) Loss of connectivity between cingulate cortex and thalamus in individuals when exposing each animal to a single isoflurane dose (group 2). (a)-(d) Labels at abscissa indicate isoflurane concentration (in %). All results have been corrected statistically using randomized permutations.

While this decrease affected all cortical regions, there were quantitative differences depending on the homotopic regions evaluated. For example, the dynamic range for the decrease in z-transformed partial correlation values for the interhemispheric homotopic connectivity was higher for the somatosensory area representing the upper limb (S1-UL) as compared to forelimb motor M1 cortex (M1-FL), which again was higher than in the somatosensory barrel field (S1-BF). For the highest isoflurane dose, negative z-values have been obtained for S1-UL and S1-BF, but not for M1-FL. The z-transformed partial correlation values have been corrected using randomised permutations for multiple corrections. No significant cortico-thalamic interaction could be identified at any of the isoflurane doses (for details see Supplementary Material).(Fig. 1c,d).

Full stationary FCs using the full-length time series were computed for each isoflurane dose The resulting stationary FC matrix is organized according to brain networks as defined in [32; 33] with odd lines/columns comprising ICs located in the left hemisphere and even rows/columns comprising ICs in the right hemisphere. The matrix includes the lateral cortical network (LCN: including somatosensory S1, secondary somatosensory cortex S2 and motor M1 cortex), the associative cortical network (ACN: including limb cortex limb and auditory cortex Au), default-mode network (DMN: including prefrontal cortex PFC and cingulate cortex Cg), sub-cortical network (SuCN: including piriform cortex Piri, striatum Str, ventral amygdala vAmg, lateral amygdala lAmg and globus pallidus GP) and thalamic network (ThN: including dorsal thalamus dTh and ventral thalamus vTh). The brain regions corresponding to the ICs in the matrix are shown in Fig. 2. Significant within-network interactions for

LCN and ACN as well as between-network interactions LCN-ACN were observed for an anesthesia doses between 1.1 and 1.5% isoflurane (Figure 3). At the 2.0% isoflurane only minimal residual FC have been observed. At all isoflurane levels there was no obvious FC involving the thalamic networks.
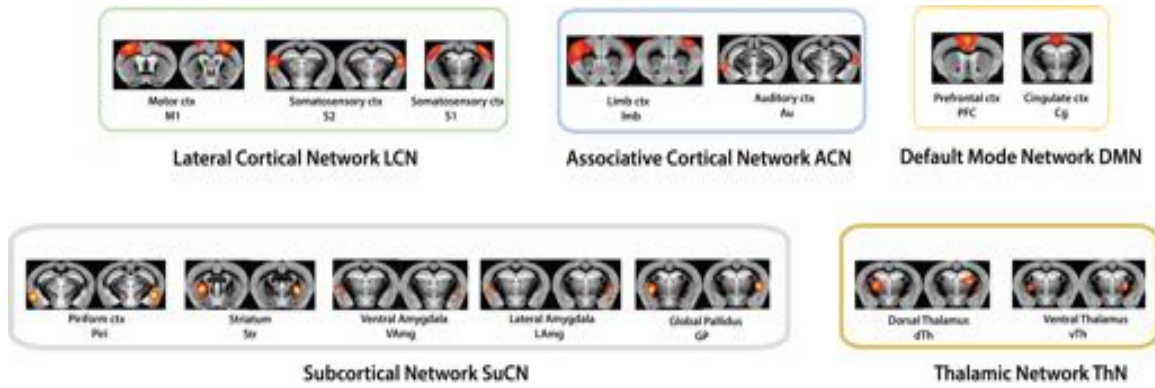


**Figure 2: Allocation of ICs to functional modules.** The 25 ICs that have been identified; have been attributed to the modules defined in [32; 33], i.e. lateral cortical network (LCN), associated cortical network (ACN), default mode network (DMN), subcortical network (SuCN) and thalamic network (ThN).
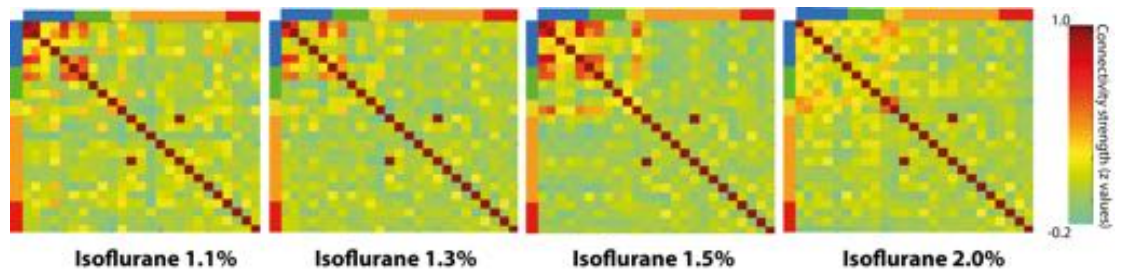


**Figure 3: Loss of connectivity between ICs upon increasing isoflurane dose.** Full correlations between the time series were computed for each isoflurane dose to obtain a stationary functional connectivity matrix. The matrix is organized according to brain networks as defined in [32; 33] with odd rows/columns displaying ICs in left and even rows/columns in right hemisphere. Significant within and between interactions are found for LCN-ACN for isoflurane levels between 1.1 and 1.5%. This modular structure is not seen for 2% isoflurane. Modules are color coded as (bars on left side

and top of correlation matrix): LCN = blue, ACN = green, DMN = yellow, SuCN = orange, thalamus = red. The colour bar indicates the z-transformed correlation values.

### *3.3.2 Dynamic functional connectivity analysis*

Dictionary learning was used [6; 7; 8; 31] for evaluating dynamic aspects in functional connectivity. The atoms (dFS) were estimated by concatenating the results obtained from all groups and then those atoms were used as regressor for analysing rs-fMRI data sets for each individual anesthesia dose, similar to the procedure applied in ICA analysis and dual regression. Window length was set at 40 seconds. Since TR = 1 sec for our data, that means window length = 40 TRs. The step size was set at 1 TR. We set the number of components to keep at K=20. Number of dictionary learning folds were set at 30. Maximum number of restarts allowed for getting convergence was set at 5. We used least square projection method of least fitting for the dictionary learning backfitting. Twenty atoms explaining approximately 50% of the variability have been considered for the analysis (Suppl. Fig. 1), some of them exhibiting remarkable structure reflecting functional modules (Suppl. Fig. 2). For example dFS #1 reflects the interaction of LCN with the SuCN, DMN and ThN as well as DMN with SuCN and ThN, while dFS #9 captures the connectivity between components of the thalamus and cortical and DMN components. Similarly, dFS #10 predominantly reflects FC between LCN/SuCN and the other networks identified including ACN, and dFS #14 captures the interaction of DMN with other networks. The structure of the several dFS reveals distinct interactions within and between the networks identified (Suppl. Fig. 2).

We then applied the dFC algorithm separately to analyse rs-fMRI data obtained for the highest and the lowest isoflurane dose to identify the dFSs accounting for most of the variability in each case (Fig. 4). While the most relevant dFSs obtained at 1.1% isoflurane exhibited some structure (ACN-ACN and ACN-LCN; ACN/LCN-DMN; ACN/LCN-DMN and ACN/LCN-SuCN; ACN/LCN-SuCN and ACN/LCN-Thal connectivity), this was not the case for dFSs obtained at 2.0% isoflurane, which were found to be largely unstructured.
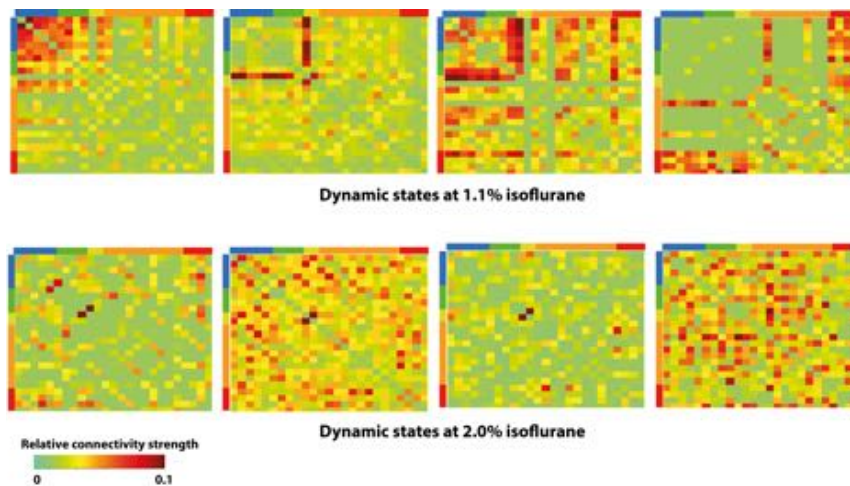
Dynamic states at 1.1% isoflurane

Dynamic states at 2.0% isoflurane

Relative connectivity strength

0          0.1

**Figure 4: Change in most relevant dynamic functional states upon increasing isoflurane dose from 1.1 to 2%**. The figure shows the four dFS obtained for mice anesthetized at 1.1% (upper row) and 2.0% isoflurane (lower row) associated with the highest weight. Modules are colour coded as: LCN = blue, ACN = green, DMN = yellow, SuCN = orange, thalamus = red. The colour bar indicates the z-transformed correlation values.

For analysis of dose-dependent effects the order of the atoms was kept fixed to allow for comparisons across groups. There was a significant decrease in weight for two of the atoms only when comparing results obtained at 1.1% isoflurane with all other doses individually: dFS #6 reflecting the interaction ACN/LCN with SuCN, and dFS #10 showing a decrease in weights of SuCN-ACN and SuCN-DMN networks with the increased dose of anesthetic (Fig. 5).



Relative connectivity strength

0          0.1
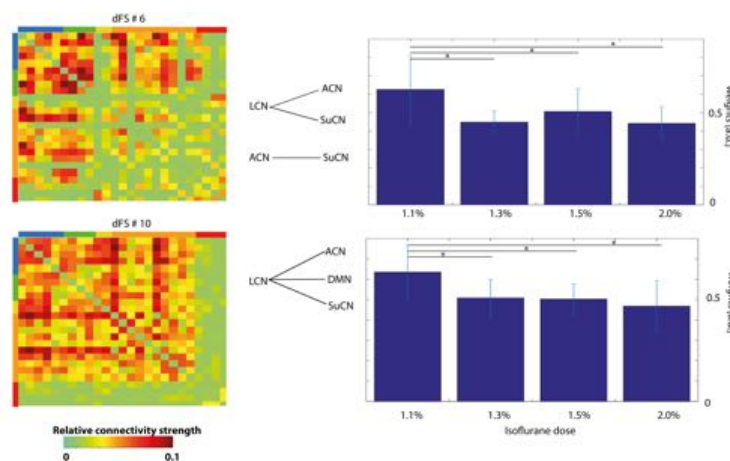
**Figure 5**: **Significant decrease in weight of dFSs as a function of isoflurane dose**. Results of dFC analysis for isoflurane doses of 1.1%, 1,3%, 1.5% and 2.0% (group 1).

dFS #6 and dFS #10 revealed significant decrease in weight between the lowest isoflurane dose (1.1%) and all other doses tested (FDR corrected, * p<0.01). These dFSs describe interactions between LCN-ACN, LCN-SuCN, LCN-DMN, LCN-ThN, SuCN-ACN and SuCN-DMN. The bar graphs show the mean absolute sum of dFS fluctuations.

Fig. 6 shows the significant dFSs when comparing mice anesthetized at 1.1% and 2.0%. Increasing the anesthesia dose led to decreased weights of dFS displaying modular structures as shown in Fig. 6.
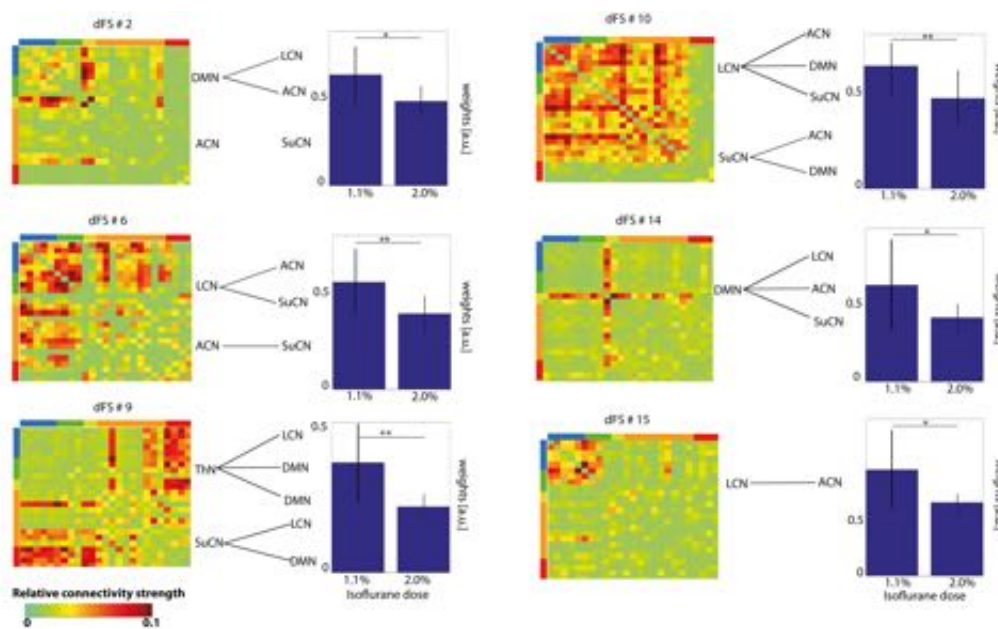


**Figure 6: Significant decrease in weight of dFSs comparing lowest and highest isoflurane dose.** dFC for isoflurane doses of 1.1% and 2.0% (group 1 ). dFS #2, dFS #6, dFS #9, dFS #10, dFS #14 and dFS #15 displayed a significant decrease upon increasing the isoflurane dose from 1.1% to 2.0% (* p<0.01, **p<0.001, FDR corrected). The bar graph shows mean absolute sum of dFS fluctuations. dFS #2 and dFS #14 represent the interactions of DMN with other networks, dFS #6 of both LCN and ACN with other networks, dFS #9 of SuCN and ThN with other networks, dFS #10 among all the networks except ThN, and finally dFS #15 within and between LCN and ACN.

### 3.3.3 Physiological measurements

Anesthesia has profound effects on respiration and cardiovascular output. As animals were mechanically ventilated respiratory depression upon increasing the anesthesia level could be avoided as reflected by the stability of the blood oxygen saturation (Fig. 7a). Nevertheless, there was a dose dependent effect on the cardiovascular parameters heart rate (Fig. 7b) and pulse distention (Fig. 7c), a measure of vessel pulsatility, and thus related to blood pressure. Both parameters significantly decreased as a function of the isoflurane dose.
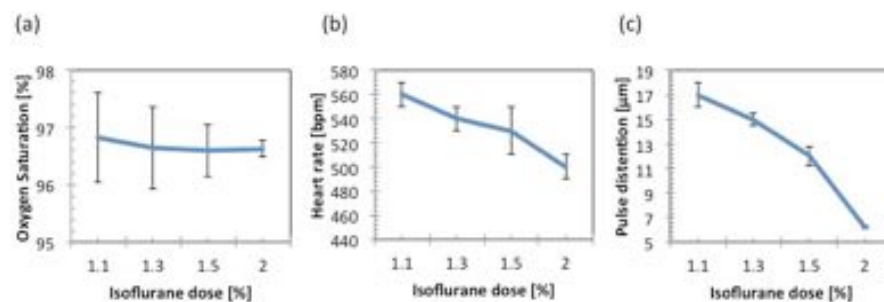


**Figure 7: Change in peripheral hemodynamic parameters as a function of isoflurane dose.** (a) Blood oxygen saturation stayed constant irrespective of isoflurane dose indicating stable respiratory condition in the artificially ventilated mice. (b) Heart rate dropped significantly upon increasing the isoflurane dose. (c) Similarly, there was a substantial decrease in pulse distention as the anesthesia level was increased. All values are given as mean±SD.

In order to identify potential effects of alterations in systemic physiological parameters on FC and dFS, we carried out a statistical analysis testing for significant group differences using heart rate or pulse distention values as correlator. We did not find any significant effects on FC. The detailed analysis has been described in Supplementary Material.

## 3.4 Discussion

Analysis of stationary FC, i.e. FC patterns integrated over time intervals of several minutes, in humans, monkeys, and rodents, has revealed significant anesthesia-induced alterations when compared to the awake state [34; 35; 36; 37; 38]. Nevertheless, this 'static' network view will be inherently incomplete as brain activity states are known to fluctuate at a much shorter time scale [39; 40] and the characterization of these dynamic functional states may constitute an important aspect for understanding anesthesia effects.

There are two major reasons for studying effects of anesthesia in mice. Firstly, the large majority of fMRI studies in rodents including mice use anesthesia in order to minimize confounding contributions due to physiological stress and motion. This leads to biased activity patterns as anesthesia inherently interferes with brain function, which is reflected by reports demonstrating anesthetic-specific effects on brain networks [14; 41; 42]. In addition, the effect on brain activity states does not just depend on the nature of the anesthetic drug but also on its dose, which determines anesthesia depth. Understanding these influences is essential for proper interpretation of results. Secondly, the possibility to modulate anesthesia effects on functional brain states by targeted pharmacological or genetic interventions may reveal mechanistic aspects on the anesthetics mode of action.

Isoflurane has emerged as attractive anesthetic for fMRI studies as it allows maintaining stable anesthesia condition suitable for functional imaging studies in rodents [35; 43; 44] and non-human primates [38; 45]. Nevertheless, isoflurane at higher doses leads to increases in cerebral blood flow thereby affecting neurovascular coupling as well as the hemodynamic baseline state, which will modulate fMRI readouts [46]. Also, its interference with the GABA neurotransmitter system may be limiting when studying GABAergic processing. Anesthetics such as medetomidine, an a-adrenergic agent, have been suggested as alternative [47]; however their suitability is subject to other limitations. Medetomidine is vasoconstrictive and may lead to side effects such as convulsion. The combined use of medetomidine and isoflurane at low dose allows controlling confounding side effects to a large extent, while combining beneficial aspects of the two drugs [14]. Though combination anesthesia is increasingly used in functional imaging studies, it is important to understand the effects of each compound individually in order to analyze potential synergistic or

antagonistic interactions. We therefore focused in the current work on characterizing effects of isoflurane on static and dynamic functional states in relation to its dose.

There are two hypotheses regarding anesthesia effects on brain function. The first hypothesis suggests anesthesia to disrupt functional networks of the brain such that individual brain regions become increasingly disconnected. Decreased functional connectivity within and across brain networks would impair information processing, which might be associated with loss of consciousness [10; 48]. The second hypothesis suggests anesthesia to synchronize activity across large brain networks. This loss of spatial segregation might prevent structured information processing and thus lead to loss of consciousness [35; 38; 49].

The result of stationary FC analysis reveals loss of connectivity between homotopic areas within the two hemispheres, which is in agreement with the first hypothesis. This is also revealed by dynamic network analysis, which indicates reduced weight of dFSs that exhibit pronounced modular structure upon increasing the isoflurane dose. On the other hand, the contributions of dFS that lack modular structure and rather display widespread synchronization across cortical and subcortical areas become dominant at an isoflurane dose of 2%. This is in line with the notion, that topological segregation is lost with increasing anesthesia depth, corresponding to the second hypothesis and in line with earlier studies in rats [42; 50] and mice [14]. Apparently, this widespread synchronization is averaged out when considering long integration intervals (stationary FC) and we observe loss of homotopic correlation as the prominent feature.

An interesting observation is that mice under all isoflurane levels displayed absence of cortico-thalamic FC. This is in contrast to several studies reporting absence of cortico-thalamic interaction during sedation [14; 17; 19; 20; 42; 50; 51], while other studies have revealed diminished but detectable thalamo-cortical FC [15; 16]. However there had been varying reports regarding the cortico-thalamic connectivity. For example, Boveroux et al. [13] found thalamo-cortical functional anti-correlation when studying propofol-induced unconsciousness in humans. Also in rats, anti-correlated FC between thalamus and cortex has been found [52]. These apparent discrepancies might indicate that the thalamo-cortical interaction is dependent on the type and depth of anesthesia.

An important aspect to be considered when analyzing the hemodynamic readout constituting the fMRI signal is the stability of the physiological state of the subject. Physiological parameter recordings revealed constant and physiological oxygen

saturation values throughout the experiment irrespective of the anesthetic doses used. This is to be expected as mice have been artificially ventilated at constant tidal volumes with a defined amount of oxygen provided via the respiration gas and allowed avoiding potential effects due to respiratory depression induced by anesthesia. Measurements of heart rate and pulse distention showed an overall decline with increasing dose of anesthesia indicative of reduced cardiac output. Since fMRI is an indirect measure of cerebral activation, which depends on the cerebrovascular baseline state, changes in blood pressure might affect the amplitude of the BOLD signal fluctuations and thereby potentially also the correlation analysis. This has also been reported previously in [53] and may be considered a drawback of studying anesthesia-related effects using fMRI.

Apart from potential impact due to alteration in the physiological state, fMRI is hampered by its inherently low temporal resolution due to (i) the temporal characteristic of the hemodynamic response acting as strong temporal low-pass filter, and (ii) the sequential nature of MRI data sampling. Sampling rates are of the order of seconds, which puts a lower limit to a sampling interval in dynamic FC analysis of the order of 15 to 20 s, which is long compared to changes in microstates as reported on the basis of EEG measurements [7; 31]. Changes occurring at sub-second time scales will be missed. Similarly the effect of the window length on dFS has also been widely discussed. Shakil et al. [54] showed that clustering based on the sliding window correlation did not reliably reflect the underlying state transitions unless the window length was comparable to the state duration. However, here we have utilized the algorithm that uses dictionary-learning approach, which has been shown to reliably estimate the dFS with changing window lengths within a range of 15 to 50 s [7; 31; 55; 56; 57]. We verified our results in two separate studies using a slightly different experimental design: one group with individual animals measured at a defined isoflurane dose and a second group with dose-escalation in individual animals. We hence refrained from combining them. Both studies yielded the same results demonstrating the robustness of the findings, though we are aware that larger group sizes might have revealed additional statistically changes in brain networks in response to increasing anesthesia depth. Our results refer strictly to isoflurane and it is not obvious to what extent the findings regarding specific networks can be generalized to other anesthetic drugs, which have different modes of action and thus affect different brain networks [14]. Nevertheless, there may be some general principles: Grandjean

et. al. [14] using stationary rs-fMRI demonstrated widespread cortical synchronization in mice exposed to high levels of urethane, i.e. loss of modular structure. This is analogous to the patterns observed under 2% of isoflurane and might indicate that loss of functional segregation is in fact an indicator of deep anesthesia, irrespective of the agent used.

In summary, we applied stationary and dynamic FC analysis to examine dose dependent effects of isoflurane anesthesia on functional networks in mice. We found that while stationary FC analysis with increasing dose of anesthetic revealed loss of functional connectivity between homotopic brain regions, dynamic FC analysis revealed loss of spatial segregation across some of the brain functional networks, which might reflect a state of deep anesthesia. Dynamic functional network analysis revealed significant interactions among functional networks that were not apparent from the conventional stationary analysis.
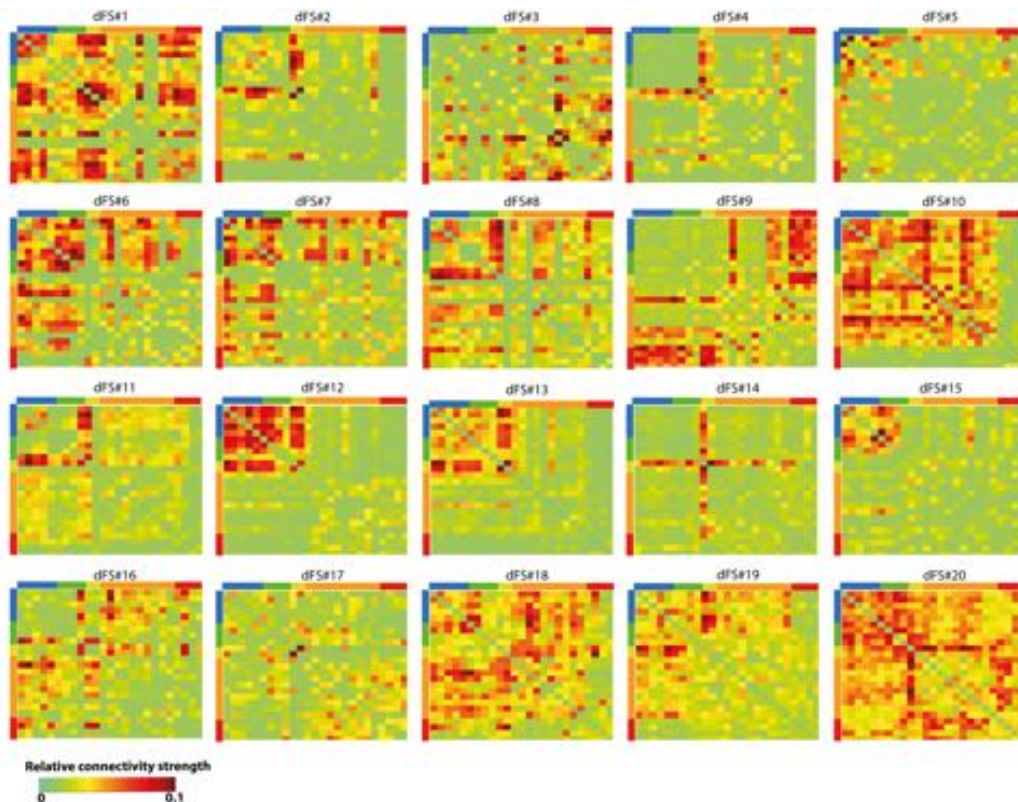
## 3.5 Supplementary Material

### 3.5.1 Supplementary Table

|  | S1B left - S1B right | M1 left - M1 right | S1 left - S1 right |
|---|---|---|---|
| **1.1 vs 1.3** | 0.972 | 0.998 | 0.91 |
| **1.1 vs 1.5** | 0.92 | 0.93 | 0.9 |
| **1.1 vs 2.0** | 0.999 | 0.999 | 0.999 |
| **1.3 vs 1.5** | 0.89 | 0.89 | 0.9 |
| **1.3 vs 2.0** | 0.998 | 0.999 | 0.92 |
| **1.5 vs 2.0** | 0.999 | 0.999 | 0.93 |

Suppl. Table 1: Corresponding p values of all the comparisons shown in Fig. 1a

### 3.5.2 Supplementary Figures



**Suppl. Fig. 1: Twenty dFS states derived from dictionary learning dFC analysis.** The majority of dFSs presents a high degree of (modular) structure and revealed network interactions that were masked in the stationary FC analysis. Modules are colour coded as (column at left and row at top of matrices): LCN = blue, ACN = green, DMN = yellow, SuCN = orange, thalamus = red.
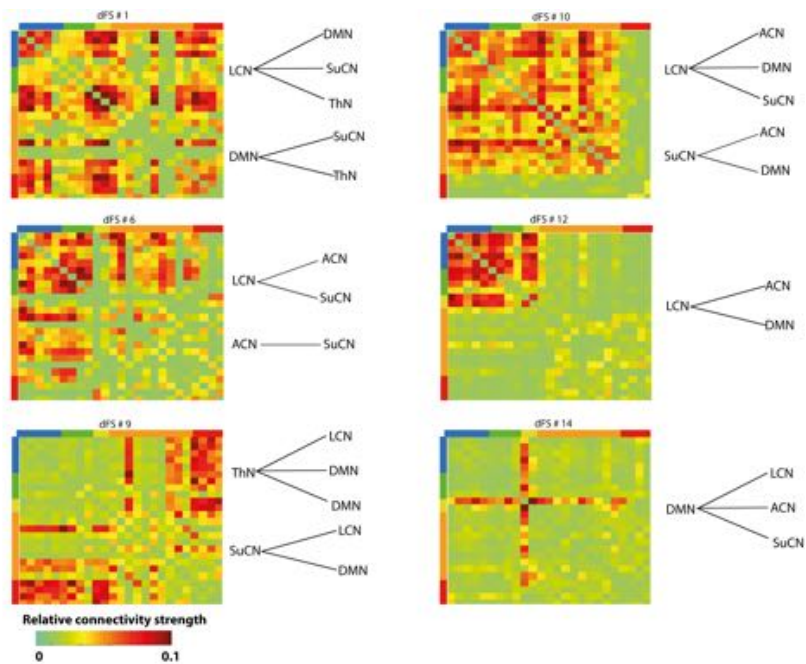
**Suppl. Fig. 2: Selected dFS of mouse brain obtained by concatenating data from all anaesthetic doses tested.** Atoms (dFSs) generated from the dictionary learning dFC analysis show a high degree of structure. Line drawings indicate dominant network interactions. Modules are colour coded as (column at left and row at top of matrices): LCN = blue, ACN = green, DMN = yellow, SuCN = orange, thalamus = red.

### 3.5.3 Supplementary Note

#### *3.5.3.1. Analysis of mean amplitude values of thalamus and cortex*

The connectivity between thalamus and cortex was found to be absent in all the doses of isoflurane
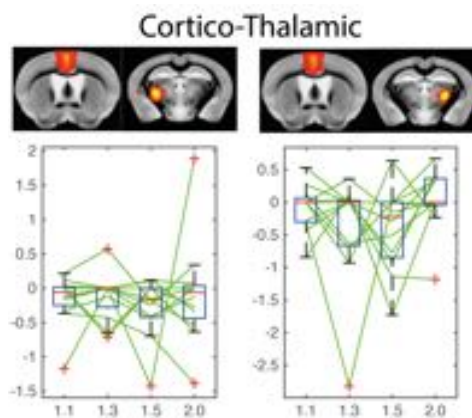
In order to determine the basis of low connectivity values between thalamus and cortex at higher doses of anesthesia (isoflurane), we analysed the mean amplitude values of thalamus and cortex at varying levels of isoflurane.

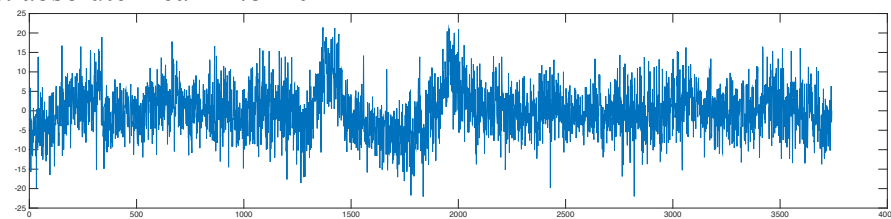The following regions-of-interest and their corresponding time series were used.

Component   #1                    # 52                    # 1                    # 61
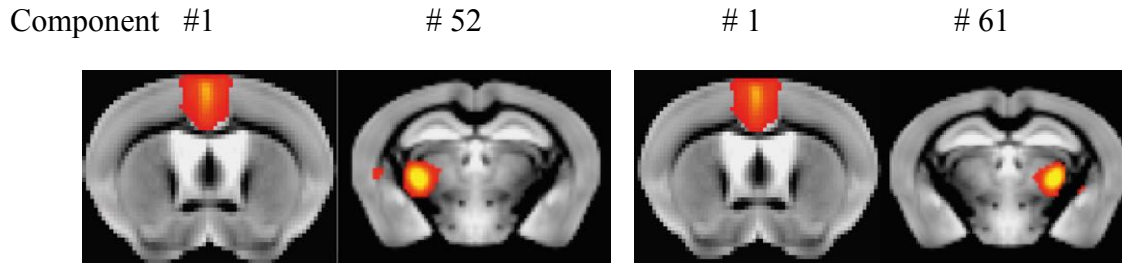


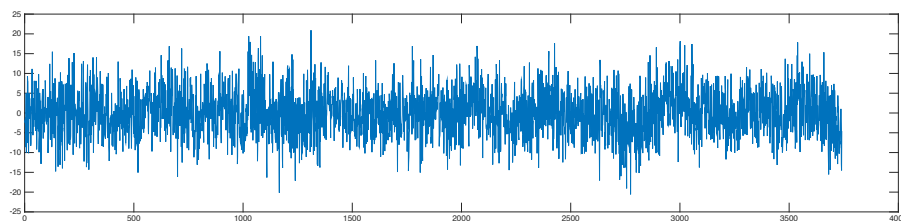All the time series were mean centered.

a) Component number 52

Left thalamus: Over all mean of all the 4 doses = 4.6956
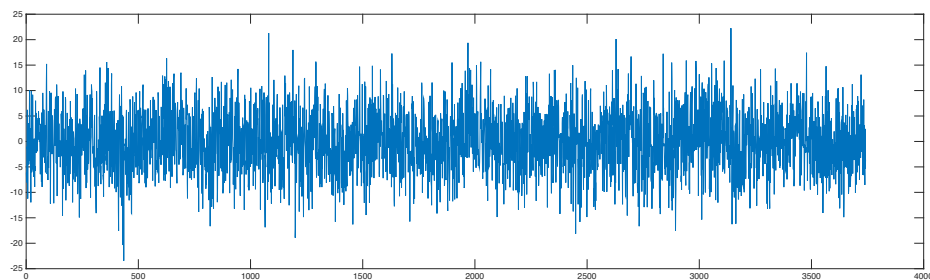
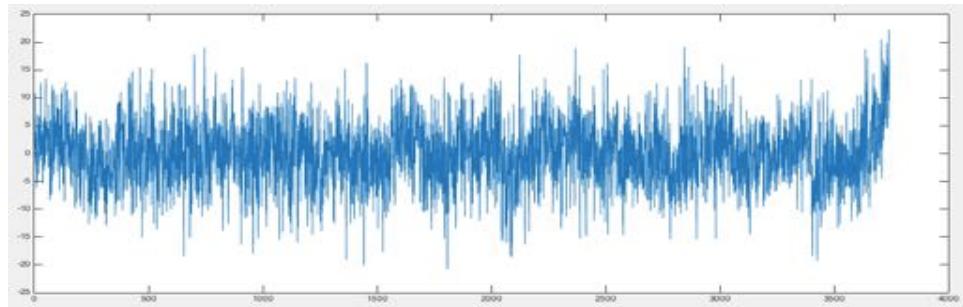Individual mean per dose

Dose 1.1%: absolute mean=4.8240



Dose 1.3%: absolute mean=4.6143



Dose 1.5%: absolute mean=4.7647
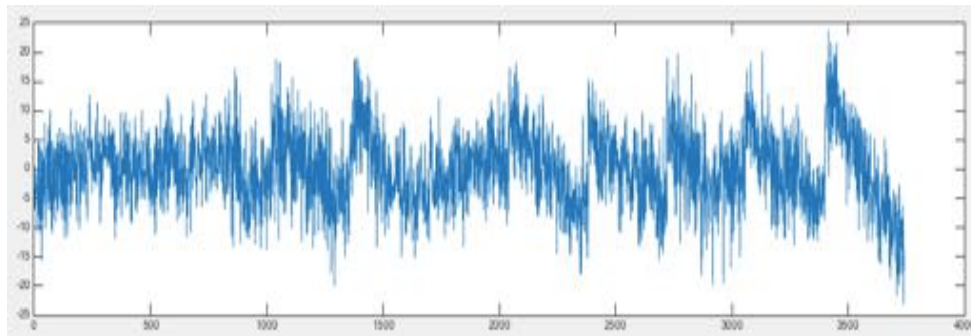
Dose 2.0%: absolute mean=4.5794
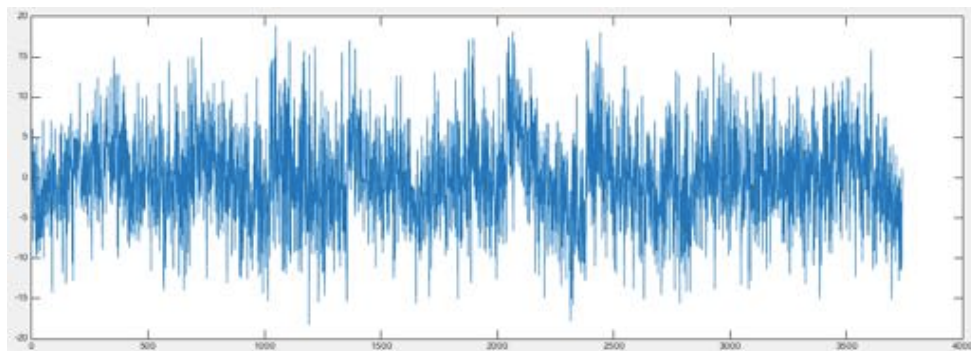


b) Component number 61

Right thalamus: overall mean of all the 4 doses: 4.4938
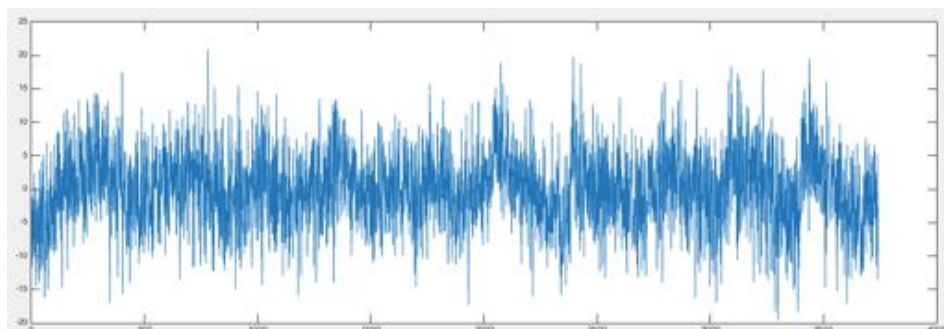

Individual mean per dose

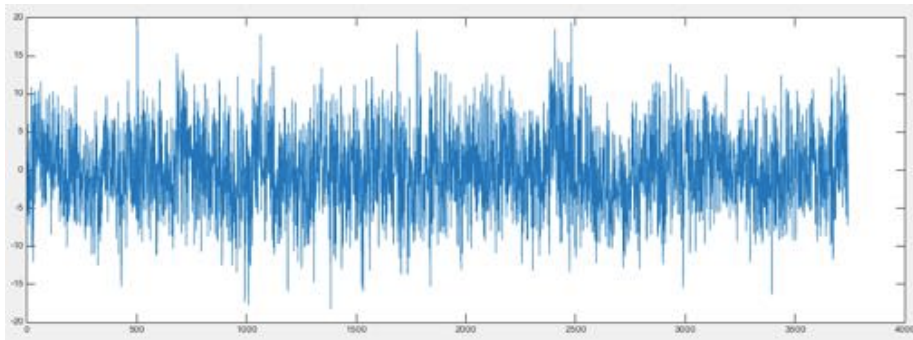Dose 1.1%: absolute mean=4.8862



Dose 1.3%: absolute mean=4.4130



Dose 1.5: absolute mean=4.4812

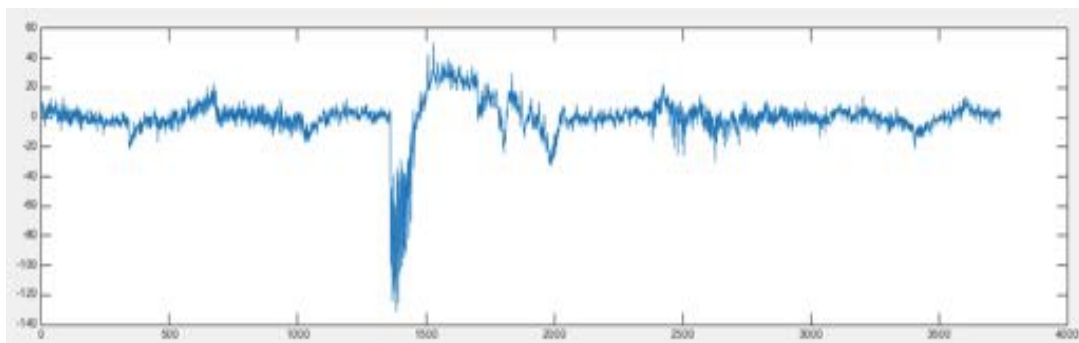Dose 2.0%: absolute mean=4.1948



c) Prefrontal cortex / Cingulate cortex

Component number 1

PFC/Cg: overall absolute mean of all the 4 doses: 6.0138

Individual mean per dose

Dose 1.1%: absolute mean= 7.2420



Dose 1.3%: absolute mean=3.9848
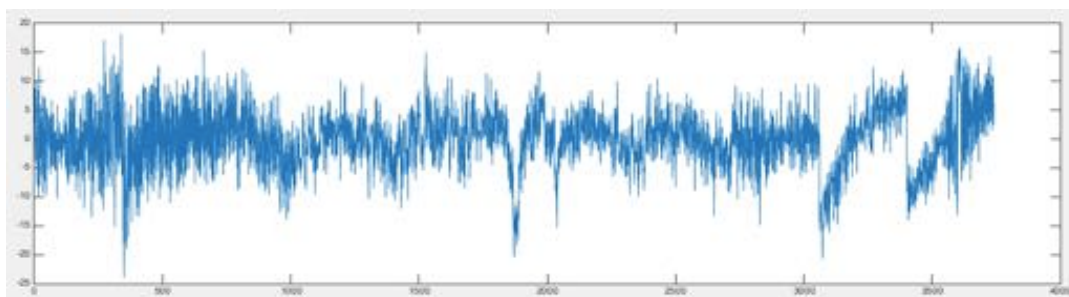
Dose 1.5%: absolute mean=5.9087
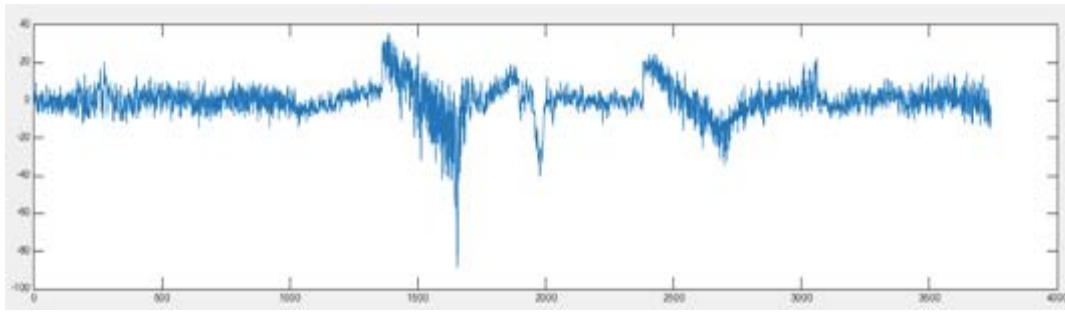


Dose 2.0%: absolute mean=6.9197
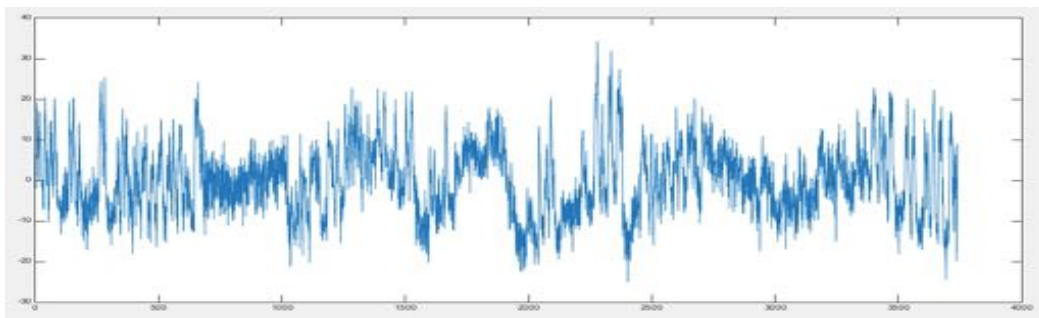


**Summarised results**

1. Left thalamus

| Dose | 1.1 | 1.3 | 1.5 | 2.0 |
|------|-----|-----|-----|-----|
| mean | 4.8240 | 4.6143 | 4.7647 | 4.5794 |
| Std | 3.7422 | 3.4224 | 3.5204 | 3.4871 |

2. Right thalamus

| Dose | 1.1 | 1.3 | 1.5 | 2.0 |
|------|-----|-----|-----|-----|
| mean | 4.8862 | 4.4130 | 4.4812 | 4.1948 |
| Std | 3.7583 | 3.4008 | 3.4880 | 3.1392 |

2. PFC/Cg

| Dose | 1.1 | 1.3 | 1.5 | 2.0 |
|------|-----|-----|-----|-----|
| mean | 7.2420 | 3.9848 | 5.9087 | 6.9197 |
| Std | 11.4408 | 3.2196 | 6.5113 | 4.9898 |

The data suggests that both thalamus and cingulate cortex showed consistent variations over time at all doses. *That implies that there was not enough evidence that the absence of correlation between thalamus and cortex is due to the inactivity of either of the regions. However we notice that the change in thalamus was smaller compared to the change in cortex with increase in doses.*

### 3.5.3.2. Analysing if there is any group effect based on pulse distention and heart rate

We used the dual regression dr2*.* files which were the group-concatenated maps per IC. We divided each of these IC files in to group wise pairs

*Group 1 = subject 1 – 12*

*Group 2 = subject 13 – 24*

*Group 3 = subject 25 – 36*

*Group 4 = subject 37 – 48*

*IC1 → Group 1, Group 2, Group 3, Group 4*

.

.

.

*IC100 → Group 1, Group 2, Group 3, Group 4*

**a) Pulse Distention (PD)**

| Dose (%) | 1.1 | 1.3 | 1.5 | 2 |
|---|---|---|---|---|
| PD | 17 | 15 | 12 | 6.2 |

The mean value is: 12.5

Based on this, we calculated the difference from mean to the individual values.

The design matrix that fits, approximates to

| 5 | 3 | -1 | -7 |
|---|---|---|---|

We multiplied each group with its corresponding multiplicative factor.

*IC1 → Group 1 x **5**, Group 2 x **3**, Group 3 x **-1**, Group 4 x **-7***

.

.

.

*IC100 → Group 1 x **5**, Group 2 x **3**, Group 3 x **-1**, Group 4 x **-7***

Then we concatenate the group-wise ICs

*Group 1 x 5 → IC1 . . . IC100*

*Group 2 x 3 → IC1 . . . IC100*

*Group 3 x -1 → IC1 . . . IC100*

*Group 4 x -7 → IC1 . . . IC100*

Finally we combined them according to their effect (positive or negative)

*Positive effect: Group 1 x 5 + Group 2 x 3 → IC1 . . . IC100*

*Negative effect: Group 3 x -1 + Group 4 x -7 → IC1 . . . IC100*

Then we use the design matrix of two-sample t test to estimate the significant differences using randomised permutation

Design matrix used was: 1 -1

Where 1 was for positive effect and -1 for negative effect

We ran 5000 randomised permutations to determine the significance. We did not find a significant effect of pulse distention on functional connectivity across the groups.

**b) Heart Rate (HR)**

| Dose (%) | 1.1 | 1.3 | 1.5 | 2 |
|---|---|---|---|---|
| HR(bpm) | 560 | 540 | 530 | 500 |

The mean value is: 532.5

Based on this, we calculated the difference from mean to the individual values.

The design matrix that fits, approximates to

| 16 | 4 | 1 | -21 |
|---|---|---|---|

We multiplied each group with its corresponding multiplicative factor.

*IC1 → Group 1 x **16**, Group 2 x **4**, Group 3 x **1**, Group 4 x **-21***

*.*

*.*

*.*

*IC100 → Group 1 x **16**, Group 2 x **4**, Group 3 x **1**, Group 4 x **-21***

Then we concatenate the group-wise ICs

*Group 1 x 16 → IC1 . . . IC100*

*Group 2 x 4 → IC1 . . . IC100*

*Group 3 x 1 → IC1 . . . IC100*

*Group 4 x -21 → IC1 . . . IC100*

Finally we combined them according to their effect (positive or negative)

*Positive effect: Group 1 x 16 + Group 2 x 4 Group 3 x 1 → IC1 . . . IC100*

*Negative effect: Group 4 x -21 → IC1 . . . IC100*

Then we use the design matrix of two-sample t test to estimate the significant differences using randomised permutation

Design matrix used was: 1 -1

Where 1 was for positive effect and -1 for negative effect

We ran 5000 randomised permutations to determine the significance. We did not find a significant effect of heart rate on functional connectivity across the groups.

**3.6 References**

[1] C.J. Pomfrett, J.R. Sneyd, J.R. Barrie, and T.E. Healy, Respiratory sinus arrhythmia: comparison with EEG indices during isoflurane anaesthesia at 0.65 and 1.2 MAC. Br J Anaesth 72 (1994) 397-402.

[2] E. Eriksson, Effects of local anaesthetics on the EEG. Acta Anaesthesiol Scand Suppl 25 (1966) 37-41.

[3] B.J. Palanca, G.A. Mashour, and M.S. Avidan, Processed electroencephalogram in depth of anesthesia monitoring. Curr Opin Anaesthesiol 22 (2009) 553-9.

[4] M.E. Raichle, A.M. MacLeod, A.Z. Snyder, W.J. Powers, D.A. Gusnard, and G.L. Shulman, A default mode of brain function. Proc Natl Acad Sci U S A 98 (2001) 676-82.

[5] B. Biswal, F.Z. Yetkin, V.M. Haughton, and J.S. Hyde, Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med 34 (1995) 537-41.

[6] N. Leonardi, J. Richiardi, M. Gschwind, S. Simioni, J.M. Annoni, M. Schluep, P. Vuilleumier, and D. Van De Ville, Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest. Neuroimage 83 (2013) 937-50.

[7] N. Leonardi, and D. Van De Ville, On spurious and real fluctuations of dynamic functional connectivity during rest. Neuroimage 104 (2015) 430-6.

[8] J. Grandjean, M.G. Preti, T.A.W. Bolton, M. Buerge, E. Seifritz, C.R. Pryce, D. Van De Ville, and M. Rudin, Dynamic reorganization of intrinsic functional networks in the mouse brain. Neuroimage 152 (2017) 497-508.

[9] V.C. Austin, A.M. Blamire, K.A. Allers, T. Sharp, P. Styles, P.M. Matthews, and N.R. Sibson, Confounding effects of anesthesia on functional activation in rodent brain: a study of halothane and alpha-chloralose anesthesia. Neuroimage 24 (2005) 92-100.

[10] A.G. Hudetz, General anesthesia and human brain connectivity. Brain Connect 2 (2012) 291-302.

[11] K.K. Kaisti, L. Metsahonkala, M. Teras, V. Oikonen, S. Aalto, S. Jaaskelainen, S. Hinkka, and H. Scheinin, Effects of surgical levels of propofol and sevoflurane anesthesia on cerebral blood flow in healthy subjects studied with positron emission tomography. Anesthesiology 96 (2002) 1358-70.

[12] A. Schroeter, F. Schlegel, A. Seuwen, J. Grandjean, and M. Rudin, Specificity of stimulus-evoked fMRI responses in the mouse: the influence of systemic physiological changes associated with innocuous stimulation under four different anesthetics. Neuroimage 94 (2014) 372-84.

[13] P. Boveroux, A. Vanhaudenhuyse, M.A. Bruno, Q. Noirhomme, S. Lauwick, A. Luxen, C. Degueldre, A. Plenevaux, C. Schnakers, C. Phillips, J.F. Brichant, V. Bonhomme, P. Maquet, M.D. Greicius, S. Laureys, and M. Boly, Breakdown of within- and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. Anesthesiology 113 (2010) 1038-53.

[14] J. Grandjean, A. Schroeter, I. Batata, and M. Rudin, Optimization of anesthesia protocol for resting-state fMRI in mice based on differential effects of anesthetics on functional connectivity patterns. Neuroimage 102 Pt 2 (2014) 838-47.

[15] S.P. Kim, E. Hwang, J.H. Kang, S. Kim, and J.H. Choi, Changes in the thalamocortical connectivity during anesthesia-induced transitions in consciousness. Neuroreport 23 (2012) 294-8.

[16] Z. Liang, J. King, and N. Zhang, Intrinsic organization of the anesthetized brain. J Neurosci 32 (2012) 10183-91.

[17] G.A. Mashour, and M.T. Alkire, Consciousness, anesthesia, and the thalamocortical system. Anesthesiology 118 (2013) 13-5.

[18] R.N. Mhuircheartaigh, D. Rosenorn-Lanng, R. Wise, S. Jbabdi, R. Rogers, and I. Tracey, Cortical and subcortical connectivity changes during decreasing levels of consciousness in humans: a functional magnetic resonance imaging study using propofol. J Neurosci 30 (2010) 9095-102.

[19] N.S. White, and M.T. Alkire, Impaired thalamocortical connectivity in humans during general-anesthetic-induced unconsciousness. Neuroimage 19 (2003) 402-11.

[20] F. Zhao, T. Zhao, L. Zhou, Q. Wu, and X. Hu, BOLD study of stimulation-induced neural activity and resting-state connectivity in medetomidine-sedated rat. Neuroimage 39 (2008) 248-60.

[21] K. Masamoto, M. Fukuda, A. Vazquez, and S.G. Kim, Dose-dependent effect of isoflurane on neurovascular coupling in rat cerebral cortex. Eur J Neurosci 30 (2009) 242-50.

[22] R.M. Hutchison, M. Hutchison, K.Y. Manning, R.S. Menon, and S. Everling, Isoflurane induces dose-dependent alterations in the cortical connectivity profiles and dynamic properties of the brain's functional architecture. Hum Brain Mapp 35 (2014) 5754-75.

[23] C.F. Beckmann, and S.M. Smith, Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imaging 23 (2004) 137-52.

[24] C.F. Beckmann, M. DeLuca, J.T. Devlin, and S.M. Smith, Investigations into resting-state connectivity using independent component analysis. Philos Trans R Soc Lond B Biol Sci 360 (2005) 1001-13.

[25] C.F. Beckmann, and S.M. Smith, Tensorial extensions of independent component analysis for multisubject FMRI analysis. Neuroimage 25 (2005) 294-311.

[26] N. Filippini, B.J. MacIntosh, M.G. Hough, G.M. Goodwin, G.B. Frisoni, S.M. Smith, P.M. Matthews, C.F. Beckmann, and C.E. Mackay, Distinct patterns of brain activity in young carriers of the APOE-epsilon4 allele. Proc Natl Acad Sci U S A 106 (2009) 7209-14.

[27] I.M. Veer, C.F. Beckmann, M.J. van Tol, L. Ferrarini, J. Milles, D.J. Veltman, A. Aleman, M.A. van Buchem, N.J. van der Wee, and S.A. Rombouts, Whole brain resting-state analysis reveals decreased functional connectivity in major depression. Front Syst Neurosci 4 (2010).

[28] T.E. Nichols, and A.P. Holmes, Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15 (2002) 1-25.

[29] A.M. Winkler, G.R. Ridgway, M.A. Webster, S.M. Smith, and T.E. Nichols, Permutation inference for the general linear model. Neuroimage 92 (2014) 381-97.

[30] S.M. Smith, and T.E. Nichols, Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44 (2009) 83-98.

[31] N. Leonardi, W.R. Shirer, M.D. Greicius, and D. Van De Ville, Disentangling dynamic networks: Separated and joint expressions of functional connectivity patterns in time. Hum Brain Mapp 35 (2014) 5984-95.

[32] A. Liska, A. Galbusera, A.J. Schwarz, and A. Gozzi, Functional connectivity hubs of the mouse brain. Neuroimage 115 (2015) 281-91.

[33] V. Zerbi, J. Grandjean, M. Rudin, and N. Wenderoth, Mapping the mouse brain with rs-fMRI: An optimized pipeline for functional network identification. Neuroimage 123 (2015) 11-21.

[34] G. Deshpande, C. Kerssens, P.S. Sebel, and X. Hu, Altered local coherence in the default mode network due to sevoflurane anesthesia. Brain Res 1318 (2010) 110-21.

[35] X. Liu, X.H. Zhu, Y. Zhang, and W. Chen, Neural origin of spontaneous hemodynamic fluctuations in rats under burst-suppression anesthesia condition. Cereb Cortex 21 (2011) 374-84.

[36] H. Lu, Y. Zuo, H. Gu, J.A. Waltz, W. Zhan, C.A. Scholl, W. Rea, Y. Yang, and E.A. Stein, Synchronized delta oscillations correlate with the resting-state functional MRI signal. Proc Natl Acad Sci U S A 104 (2007) 18265-9.

[37] S.J. Peltier, C. Kerssens, S.B. Hamann, P.S. Sebel, M. Byas-Smith, and X. Hu, Functional connectivity changes with concentration of sevoflurane anesthesia. Neuroreport 16 (2005) 285-8.

[38] J.L. Vincent, G.H. Patel, M.D. Fox, A.Z. Snyder, J.T. Baker, D.C. Van Essen, J.M. Zempel, L.H. Snyder, M. Corbetta, and M.E. Raichle, Intrinsic functional architecture in the anaesthetized monkey brain. Nature 447 (2007) 83-6.

[39] X. Di, and B.B. Biswal, Dynamic brain functional connectivity modulated by resting-state networks. Brain Struct Funct 220 (2015) 37-46.

[40] R. Hindriks, M.H. Adhikari, Y. Murayama, M. Ganzetti, D. Mantini, N.K. Logothetis, and G. Deco, Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI? Neuroimage 127 (2016) 242-56.

[41] D. Kalthoff, C. Po, D. Wiedermann, and M. Hoehn, Reliability and spatial specificity of rat brain sensorimotor functional connectivity networks are superior under sedation compared with general anesthesia. NMR Biomed 26 (2013) 638-50.

[42] X. Liu, X.H. Zhu, Y. Zhang, and W. Chen, The change of functional connectivity specificity in rats under various anesthesia levels and its neural origin. Brain Topogr 26 (2013) 363-77.

[43] R.M. Hutchison, S.M. Mirsattari, C.K. Jones, J.S. Gati, and L.S. Leung, Functional networks in the anesthetized rat brain revealed by independent component analysis of resting-state FMRI. J Neurophysiol 103 (2010) 3398-406.

[44] S.S. Kannurpatti, B.B. Biswal, Y.R. Kim, and B.R. Rosen, Spatio-temporal characteristics of low-frequency BOLD signal fluctuations in isoflurane-anesthetized rat brain. Neuroimage 40 (2008) 1738-47.

[45] R.M. Hutchison, L.S. Leung, S.M. Mirsattari, J.S. Gati, R.S. Menon, and S. Everling, Resting-state networks in the macaque at 7 T. Neuroimage 56 (2011) 1546-55.

[46] K. Masamoto, and I. Kanno, Anesthesia and the quantitative evaluation of neurovascular coupling. J Cereb Blood Flow Metab 32 (2012) 1233-47.

[47] M.D. Sinclair, A review of the physiological effects of alpha2-agonists related to the clinical use of medetomidine in small animal practice. Can Vet J 44 (2003) 885-97.

[48] T.L. Wu, A. Mishra, F. Wang, P.F. Yang, J.C. Gore, and L.M. Chen, Effects of isoflurane anesthesia on resting-state fMRI signals and functional connectivity within primary somatosensory cortex of monkeys. Brain Behav 6 (2016) e00591.

[49] P. Barttfeld, L. Uhrig, J.D. Sitt, M. Sigman, B. Jarraya, and S. Dehaene, Signature of consciousness in the dynamics of resting-state brain activity. Proc Natl Acad Sci U S A 112 (2015) 887-92.

[50] X. Liu, K.K. Lauer, B.D. Ward, S.J. Li, and A.G. Hudetz, Differential effects of deep sedation with propofol on the specific and nonspecific thalamocortical systems: a functional magnetic resonance imaging study. Anesthesiology 118 (2013) 59-69.

[51] M.T. Alkire, R.J. Haier, and J.H. Fallon, Toward a unified theory of narcosis: brain imaging evidence for a thalamocortical switch as the neurophysiologic basis of anesthetic-induced unconsciousness. Conscious Cogn 9 (2000) 370-86.

[52] D.J. Englot, B. Modi, A.M. Mishra, M. DeSalvo, F. Hyder, and H. Blumenfeld, Cortical deactivation induced by subcortical network dysfunction in limbic seizures. J Neurosci 29 (2009) 13006-18.

[53] R. Wang, T. Foniok, J.I. Wamsteeker, M. Qiao, B. Tomanek, R.A. Vivanco, and U.I. Tuor, Transient blood pressure changes affect the functional magnetic resonance imaging detection of cerebral activation. Neuroimage 31 (2006) 1-11.

[54] S. Shakil, C.H. Lee, and S.D. Keilholz, Evaluation of sliding window correlation performance for characterizing dynamic functional connectivity and brain states. Neuroimage 133 (2016) 111-28.

[55] S.D. Keilholz, M.E. Magnuson, W.J. Pan, M. Willis, and G.J. Thompson, Dynamic properties of functional connectivity in the rodent. Brain Connect 3 (2013) 31-40.

[56] C. Chang, and G.H. Glover, Time-frequency dynamics of resting-state brain connectivity measured with fMRI. Neuroimage 50 (2010) 81-98.

[57] R.M. Hutchison, T. Womelsdorf, J.S. Gati, S. Everling, and R.S. Menon, Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques. Hum Brain Mapp 34 (2013) 2154-77.

## 4. Dynamic effective connectivity using spectral DCM applied to resting state mice fMRI data

**Abstract**

DCM estimates effective connectivity [1; 2; 3] by fitting the neuronal model convolved with the hemodynamic response function to the observed fMRI data. Similar to dynamic functional connectivity [4; 5], the temporal variation of effective connectivity i.e. dynamic effective connectivity can be estimated. Knowing the temporal changes of the neuronal model using DCM might allow us inferring the changes in factors underlying functional connectivity, which we observe by correlating the time series extracted from different regions in the brain [6]. Effective connectivity estimates the interaction between the regions at the neuronal level. However traditionally, DCMs have only been estimated for static case so far. By estimating different DCMs at different time points of the time series, we could potentially model the temporal evolution of effective connectivity. In this study we applied a fully connected DCM model to the resting state fMRI data of mice. The time series data were then splitted in two halves to evaluate the effective connectivity for two time intervals. The results showed that at 2.0% isoflurane there are only few network connections between different brain regions. Furthermore dynamic effective connectivity revealed extra network connectivity and augmented strengths of the network that were not apparent from static effective connectivity analysis using full time series.

### 4.1 Introduction

Functional connectivity refers to the interactions between two spatially distant regions in the brain. This has been extensively studied in neuroimaging studies to understand the workings of the normal brain as well as to identify the changes in the brain due to a certain condition. In 1995, Biswal and colleagues reported the presence of functional connectivity among certain brain region even in the absence of any explicit task [7]. Analyses of the temporal coherence of these fluctuations allows identifying brain regions that display a high degree of synchronicity, which therefore are believed to be functionally connected. Since then there have been several analysis methods developed in order to extract more meaningful information of the brain networks from

these resting state fMRI signals. Resting state fMRI is also attractive for animal studies for both mechanistic purposes and as models of human disease. Since animals are typically anesthetized while imaged, rs-fMRI provides a natural way to explore the brain networks of these animals without disturbing the intrinsic functional connectivity. This method is also of great value to explore differences in brain pathologies and other cognitive disorders. rs-fMRI has become an established tool in clinical and experimental neuroscience and has been consistently applied across species for network identification. The approach is of particular interest for examining functional connectivity and brain networks in order to find alterations in brain connectivity in neurological or psychiatric disorders.

## 4.1.1 Dynamic functional connectivity

Typically when extracting information on FC on the basis of rs-fMRI data, the whole time series is used to compute correlation values, assuming that the functional networks are stationary in time. That essentially means that information on events/interactions that occur at a shorter time scale is averaged out. Yet, neural processing occurs at a much shorter time scale and EEG studies led to the identification of microstates that involve network rearrangements in the millisecond time domain [8; 9]. Dynamic functional connectivity (dFC) analysis [5] is a concept that aims at estimating the FC changes over relatively short time intervals constituting a small fraction of the full time series recorded. Limiting factors when using fMRI data are the low-pass temporal filtering imposed by the hemodynamic response function and the sequential nature of MRI data acquisition, which limits the time required to capture an image volume to one or a few seconds. Despite these limitations, dFC has been suggested as a more accurate representation of functional brain networks since brain networks are dynamic. There have been many algorithms proposed in the literature to estimate dFC from resting state fMRI data. Sliding window based correlation analysis is a simple method to define a window of duration $\tau$, for which a conventional FC analysis is carried out. The window must comprise a sufficient number of sample points to allow for a meaningful correlation analysis, yet should not be too long in order to still capture dynamic aspects. The window is then shifted by an increment $\Delta t$ (typically one sampling point), while maintaining its length $\tau$ and the FC analysis is carried out again. Repeating this procedure several times allows monitoring the changes in FC networks over time. These resulting correlation

matrices are then z-transformed and are called as dFC matrices $\mathbf{Z}_s(t_i; \tau)$. This procedure is illustrated in Figure 1.
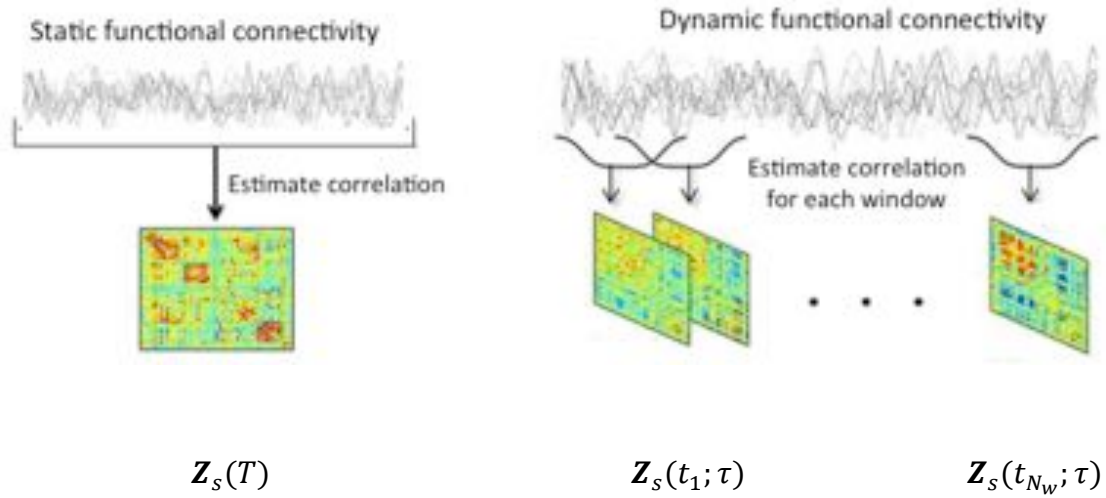


Static functional connectivity

Estimate correlation

Dynamic functional connectivity

Estimate correlation for each window

$$\mathbf{Z}_s(T) \qquad\qquad \mathbf{Z}_s(t_1; \tau) \qquad\qquad \mathbf{Z}_s(t_{N_w}; \tau)$$

Figure 1: Static FC matrix $\mathbf{Z}_s(T)$ is obtained for each subject $s$ by computing the coefficients of correlation between the whole time series of duration     for all pairs of regions (voxels) $i$ and $j$. dFC is estimated by carrying out the analogous analysis for time windows of duration $\tau$ separately, the starting point of which are incremented by $\Delta t$, yielding correlation matrices $\mathbf{Z}_s(t_i; \tau)$ with $1 \leq i \leq N_w$, $N_w = (T - \tau)/\Delta t$ being the number of windows. Adapted from [10].

## 4.1.2 Effective connectivity

Effective connectivity refers to the influence that one neural system exerts over another. That means, unlike functional connectivity which is based on the correlation between time series, and therefore explains the neurophysiological relationship between distant brain regions, effective connectivity aims to identify the relationship between the neurons, that means the relationships between distant brain regions at the neuronal level.

There have been several methods in the literature that's aims to effective connectivity. Granger causality [11; 12; 13; 14] and other Vector Autoregressive Modeling based methods, use temporal precedence for inferring causality in BOLD time series however it is regarded as directed functional connectivity rather than effective connectivity since model selection criterion in these methods make inferences about the autoregressive processes rather than causal neuronal coupling. Furthermore

Granger causality also ignores hemodynamic convolution [15; 16; 17; 18]. Similarly Structural Equation Modeling is another approach to infer effective connectivity that we can only apply to models of low complexity. SEM can only analyze steady-state brain connectivity patterns therefore it is not suitable for studying dynamic changes in the fMRI signal. Another method that has been consistently used in the literature is dynamic causal modeling, that infers causal neuronal relationship in the network and produces consistent results, some of which have been validated by simultaneously recording the local field potential (LFP) in animals.

### 4.1.3 Dynamic Causal Modeling (DCM) for estimating effective connectivity

In 2003 Friston et al. [19] proposed a method to reduce the "neuronal" evolution function to the most simple and generic form possible, i.e., a bilinear interaction between neuronal states $x$ and inputs $u$. They proposed to estimate effective connectivity [1; 2; 3] by fitting the neuronal model convolved with so-called "Balloon model", [20], an extension of hemodynamic response function, to the observed fMRI data. Then the deconvolution of the BOLD signal is performed to estimate neuronal response. This procedure is called dynamic causal modelling (DCM) as it attempts to infer causality between neuronal networks. Figure 2 shows the basic DCM model.

In order to estimate effective connectivity, several models are compared against each other to find the best fit to the experimental data (observed fMRI signal). Model comparison in effective connectivity plays an important role by selecting one of the several models, that best explains the cause of the observed fMRI data. Each alternative causal model represents distributed brain responses and therefore represents as many alternative hypotheses.
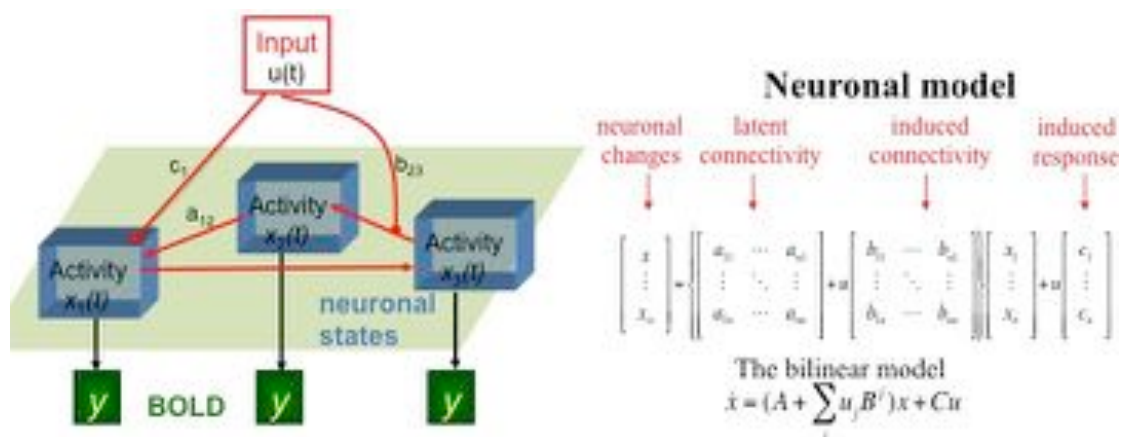
Figure 2: Figure shows a basic DCM model and its corresponding neuronal state equation.[1; 3; 19]

There have been numerous studies that use DCM to estimate the effective connectivity in humans as well as in other species [1; 21; 22; 23; 24; 25; 26; 27]. Due to the higher temporal resolution of local field potential (LFP) to infer on sequel of neuronal events, results of DCM have been validated by simultaneously recording the local LFP measurements that verified the ability of DCM to infer on synaptic processes. Razi et al [6] showed that functional connectivity can be derived from the effective connectivity, whereas the opposite is not true.

## 4.1.4 DCM for resting state fMRI

Typical DCM models cannot be specified without any driving inputs [1; 3; 25]. Resting state fMRI data has no driving inputs and therefore it had been challenging to apply DCM to resting state fMRI data. However recently there were two methods proposed to overcome this limitation. First approach is to include stochastic terms in the model [28; 29] and therefore is called stochastic DCM (sDCM). Stochastic DCM estimates time-dependent fluctuations in neuronal states producing observable fMRI data.

In case of stochastic DCM, a stochastic term is added to the ordinary differential equations used in standard DCM in order to model endogenous neuronal fluctuations. Therefore the stochastic generative model for the resting state fMRI time series can be written as:

$$\dot{x}(t) = f(x(t).u(t),\theta) + v(t) \tag{1}$$

where $\dot{x}$(t) is the rate of change of the neuronal states $x(t)$, $\theta$ are unknown parameters (i.e. the effective connectivity) and $v(t)$ is the stochastic process, modelling the random neuronal fluctuations that drive the resting state activity.

Using linearity assumptions, eq (1) can be written in generalised coordinates of motion, which can be further written in a simpler form for resting state activity as below

$$D \, \hat{x}(t) = A \, \hat{x}(t) + C \hat{u}(t) + \hat{v}(t) \tag{2}$$

where A is the Jacobian describing the behaviour – i.e. the effective connectivity – of the system near its stationary point in the absence of the fluctuations $\hat{v}(t)$. The linear dynamical system shown in eq (2) has quasi-deterministic behavior as shown by [28; 29] and is insensitive to initial conditions.

The second approach does not treat neuronal fluctuations as a stochastic noise rather evaluates the time-invariant parameters of their cross spectra, replacing the original time series with its second-order statistics (i.e., cross spectra), under stationarity assumptions [6; 25; 30]. Since this approach makes prediction in the frequency domain and is based on cross spectra features, it is called spectral DCM (spDCM) [6]. Spectral DCM has been found superior to stochastic DCM since spectral DCM was found more accurate and sensitive to group differences as well as it does not require estimation of hidden states as in the case of stochastic DCM.

Spectral DCM is based upon a deterministic model that estimates the time-invariant parameters of their cross spectra. In spectral DCM, the original time series is replaced by second order statistics (i.e., cross spectra), which means we estimate the covariance of the random fluctuations instead of estimating the time varying hidden states. This scale free form of the state noise has been suggested in the previous work on neuronal activity

$$g_v(\omega, \theta) = \alpha_v \, \omega^{-\beta_v} \tag{3}$$

$$g_e(\omega, \theta) = \alpha_e \, \omega^{-\beta_v} \tag{4}$$

where, $\{\alpha, \beta\} \subset \theta$ are the parameters that control the amplitudes and exponents of the spectral density of the neural fluctuations.

Using the model parameters, $\theta \supseteq \{A, C, \alpha, \beta\}$, the expected cross spectra is generated as follows

$$y(t) = \kappa(t) \otimes v(t) + e(t) \tag{5}$$

$$\kappa(t) = \partial_x g \, \exp(t \, \partial_x f) \tag{6}$$

$$g_y(\omega, \theta) = |K(\omega)|^2 \, g_v(\omega, \theta) + \, g_e(\omega, \theta) \hspace{3cm} (7)$$

where $K(\omega)$ is the Fourier transform of the system's (first order) Volterra kernels $\kappa(t)$, which are a function of the Jacobian or effective connectivity. Variational Laplace procedures can be used to estimate the unknown quantities $\psi = \{\varphi, \theta, \sigma\}$ of this equation [31; 32].

Given the consistency and validation of DCM results for estimating effective connectivity, we use DCM in this work. Moreover since spectral DCM has been shown to perform better than other DCM invariants for resting state data, we use spectral DCM for estimating effective connectivity in this work.

### 4.1.5 Dynamic effective connectivity

We propose to combine the two techniques by estimating effective connectivity over small windows of the resting state time series. Neuronal networks are dynamic in nature, and by estimating effective connectivity over short time windows, we can identify those changes at the neuronal network level that may have been averaged out in a typical estimation of effective connectivity through DCM using the whole resting state time series. The figure below presents the conceptual working of this technique.
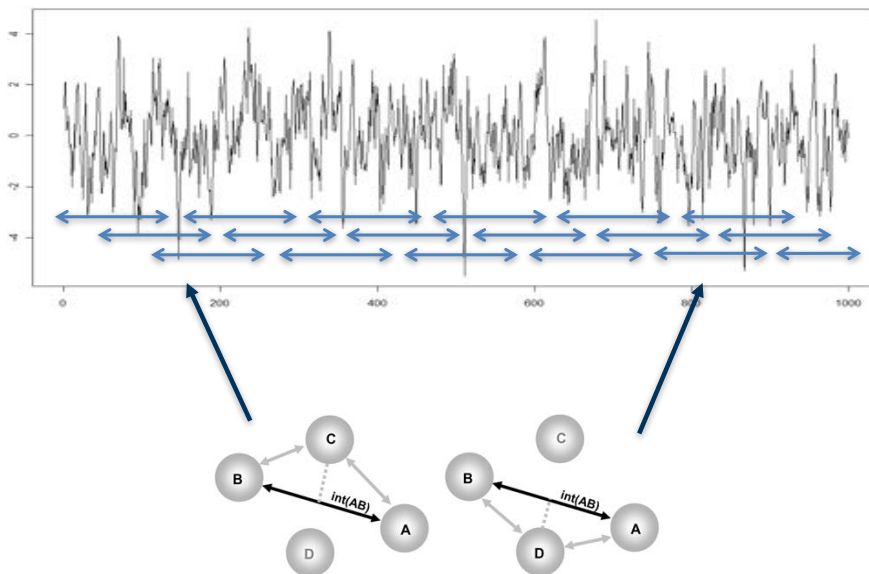


Figure 3: The estimation of dynamic effective connectivity over sliding time windows has been shown. The underlying neuronal model may change over the sliding windows..

## 4.2 Materials and method

### 4.2.1 Animals, preparation, and anesthesia

The experiments were performed in compliance with Swiss laws on animal protection and approved by the Veterinary Office of the Canton of Zurich. Female C57BL/6 mice (Janvier, Le Genest-St Isle, France) between 10 and 15 weeks old were studied. All mice were initially anesthetized with isoflurane in a 20% $O_2$ / 80% air mixture: 3.5% for induction, 2% for endotracheal intubation and during set-up on the animal bed. Throughout the duration of the experiment, animals were mechanically ventilated using a small animal ventilator (CWE, Ardmore, USA) with a 20% $O_2$ / 80% air mixture at a rate of 80 breaths/min, a respiration cycle of 25% inhalation, 75% exhalation, and an inspiration volume of 1.8 ml/min. The head was placed with the animal's incisors secured over a bite bar and fixated by ear bars, ophthalmic ointment was applied to the eyes, and a rectal temperature probe was inserted to keep the animal at 36.5 ± 0.5 °C by means of a warm-water circuit integrated into the animal holder (Bruker Biospin GmbH, Ettlingen, Germany). The tail vein was cannulated for intravenous (i.v.) administration of anesthetics and the neuromuscular blocking agent pancuronium bromide (Sigma-Aldrich, Steinheim, Germany).

Twelve animals were used in the experiment. Isoflurane (Abbott, Cham, Switzerland) dose of 2.0% was used in a 20% $O_2$ / 80% air mixture for each individual mouse. After giving isoflurane, there was a 10 min interval for equilibration before the fMRI data acquisition was started. Mice remained in the MR scanner throughout the duration of the experiment. For this work, we have only considered the neuroimaging data of the mice under 2.0% isoflurane.

### 4.2.2 fMRI

MRI/fMRI experiments were carried out using a Bruker Biospec 94/30 small animal MR system (Bruker BioSpin MRI, Ettlingen, Germany) operating at 400 MHz (9.4 T). A four-element receive-only cryogenic phased array coil (Bruker BioSpin AG, Faellanden, Switzerland) was used in combination with a linearly polarized room temperature volume resonator for transmission (Bruker BioSpin MRI, Ettlingen, Germany). Anatomical images acquired in the sagittal and horizontal direction allowed exact positioning of 12 adjacent coronal slices of 0.5 mm slice thickness, which were used for the rs-fMRI scans. A gradient-echo echo-planar imaging (GE-EPI) sequence has been used for rs-fMRI data acquisition with field of view=16x7 mm2, matrix

dimensions=80x35, TR=1 s, TE=12 ms, flip angle=60 degrees. The time series acquired was of 360 s length.

### 4.2.3 Data Pre-processing

All the pre-processing was performed using tools from FMRIB's Software Library (FSL version 5). FSL's recommended pre processing pipeline was used. Motion correction, removal of non-brain structures, high pass temporal filtering with sigma = 75.0 s; pre-whitening and global spatial smoothing of 0.2 mm was applied as part of the pre-processing. After the pre-processing, the functional scans were aligned to the high-resolution anatomical QBI (Queensland Brain Institute) template using linear affine and nonlinear diffeomorphic transformation registration as implemented in ANTs (ANTs. v 1.9; http://picsl.upenn.edu/ANTS/).

We used FSL's MELODIC for probabilistic independent component analysis [33]. The multi-session temporal ICA concatenated approach, as recommended for rs-fMRI data analysis, allowed to input all subjects from all the groups in a temporally concatenated fashion for the ICA analysis. ConcatICA yielded different activations and artifact components without the need of specifying any explicit time series model.

These results were used for Dual Regression and Network Modeling [34] of the brain networks using FSLNets. Based on the networks that were found significant, we identified 8 ICA components as shown in Figure 1. These ICs were then fed in to the dynamic causal modeling (DCM) framework in the SPM toolbox. In order to estimate the temporal evolution of effective connectivity networks, we divided the time series of 360s (360 volumes with 1 volume/s) in two blocks (0-180s, 161-340s), and ran DCM framework together with nested Bayesian Model Selection separately for each half of the time series for all subjects. We then applied a fully connected spectral DCM [6; 25] to the resting state fMRI data, and compared it against a nested models using parametric empirical bayes (PEB) framework [35; 36]. Connection strengths were identified by inspecting the parameter estimates and posterior expectations as stored in DCM.Ep.

### 4.2.4 Dynamic Effectiive Connectivity analysis

Dynamic effective connectivity analysis was performed by estimating effective connectivity over short time windows.

**4.2.4.1 Generation of sliding window signals**

We applied an square wave to the original time series, resulting in shortening the time series with keeping the data where square window was at its peak and removing the data where the square window was at its bottom. This method truncates the original time series in to smaller size. Since for DCM estimation, a larger time series is recommended, we used square wave of size 180 seconds. It can we described mathematically as follows

$$x(t) = \begin{cases} 1, & |t| < T \\ 0, & T < |t| \le \frac{1}{2}T \end{cases} \qquad (8)$$

This rectangular window is then multiplied with the experimental data time series to generate windowing effects. The same procedure was repeated for the other half of the rectangular window, which were generated by moving the rectangular window in x-axis (time axis) by 170 seconds, keeping the length of the window constant.

Figure 4 shows the rectangular window and after it is multiplied to the original fMRI time series. The result is a truncated time series of length equal to the rectangular window, that is 180 seconds.
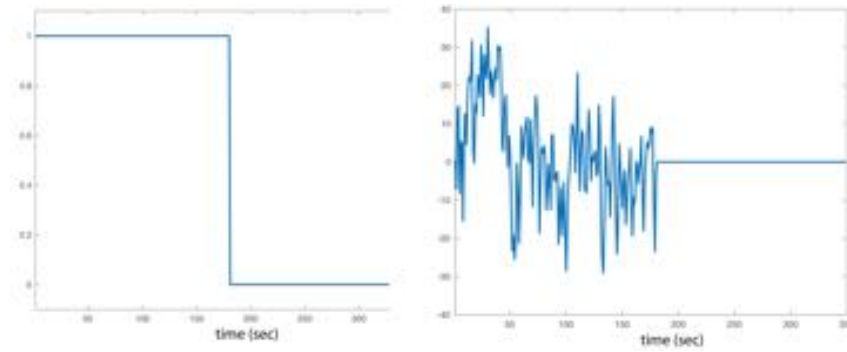


Figure 4: Rectangular window signal and its multiplication to the original fMRI time series.

For future work, we also suggest applying Gaussian window function to generate sliding windows. A Gaussian window may be implemented by calculating the coefficients of a Gaussian window.

$$w(n) = e^{-\frac{1}{2}\left(\alpha \frac{n}{(N-1)/2}\right)^2} = e^{\frac{-n^2}{2\sigma^2}}$$

where $-(N-1)/2 \le n \le (N-1)/2$ and $\alpha$ is inversely proportional to the standard deviation, $\sigma$, of a Gaussian random variable. The exact correspondence with the standard deviation of a Gaussian probability density function is $\sigma = (N-1)/(2\alpha)$.

The width of the Gaussian window can be set to a pre determined length from the literature, for example 50 sec, and an offset value equal to the last value on the discreet Gaussian curve can be used at the starting and ending points. The Gaussian curves thus generated would be padded with these offset values to make the length of this time series equal to that of the experimental data time series. This Gaussian window time series will then be multiplied with the experimental data time series to generate windowing effects. The same procedure will be repeated for multiple Gaussian windows which were generated by moving the Gaussian window in x-axis (time axis), keeping the length of the window constant

Supplementary figure 1 shows the generation of Gaussian windows and its application to the experimental fMRI signal.

### 4.2.4.2 Estimation of dynamic effective connectivity

Time courses were extracted from the selected ICA maps (n=8) and then fed in to the window generation functions, discussed above, to estimate input signals for the DCM analysis[6], which was used to estimate effective connectivity.

Fully connected DCMs of 8 regions of interest were estimated separately for each time window, as well as for full time series. All these DCMs were then used in the PEB framekwork for estimating group level connectivity strengths.

We applied nested PEB[35; 36] models, which compared all the possible models from a fully connected network.

### 4.3 Results

We used the following brain regions as selected from IC components for further analysis



Prefrontal ctx      Somatosensory ctx      Piriform ctx
PFC                 S1                     Piri

Amygdala                    Thalamus
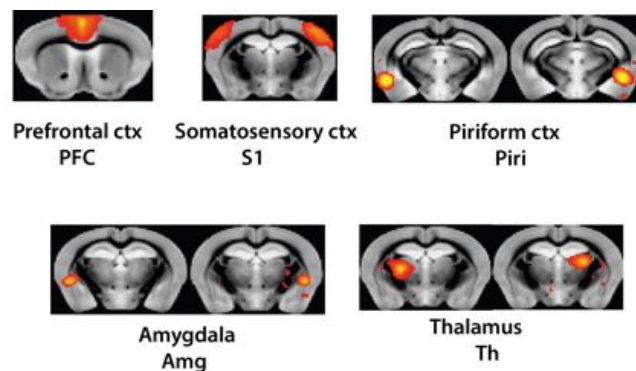Amg                         Th

Figure 5: ICs of brain regions selected through FSL Nets analysis

The DCM was used to estimate nested models that mean to test all possible models by switching on and off different connections and run under Bayesian model selection to select the best model. The connection strengths were found by estimating the values of matrix A. Supplementary figure 2 shows the model space and Bayesian model reduction and supplementary figure 3 shows the estimation of DCM parameters with strong evidence.

The DCM estimations of the parameter A with strong evidence (P>=0.95) is summarized in the table below for all windows.

|  | 1- 340 sec | 1 - 180 sec | 160 - 340 sec |
|---|---|---|---|
| Ins1 - Ins2 | 0.038 | -- | -- |
| Ins1 - PFC | 0.038 | -- | 0.045 |
| Ins1 - Amg1 | -- | 0.038 | -- |
| PFC - PFC | -0.085 | -0.09 | -0.087 |
| S1-S1 | -0.088 | -0.1 | -- |
| Th1 - Th1 | -0.08 | -0.083 | -0.086 |
| Th2 - Th2 | -0.083 | -0.1 | -0.093 |
| Amg1 - Amg1 | -0.081 | -0.102 | -0.09 |
| Amg2 - Amg2 | -0.096 | -0.11 | -0.11 |
| Ins1 - Ins1 | -0.085 | -0.105 | -0.1 |
| Ins2 - Ins2 | -0.086 | -0.1 | -0.11 |

Table 1: The table shows the parameter estimations of DCM connection strengths with P>=0.95 for different windows of the time series. The negative values are showing the inhibitory self-connections.

**4.4 Discussion**

Time varying DCMs may allow us to capture the effect of transient network patterns under different brain states. In comparison to dynamic functional connectivity analysis, dynamic effective connectivity yields information on the directionality of connections within the network. Dynamic DCMs, as shown in this work, may help identifying group differences based on the estimated time-varying network patterns.

We estimated temporal variations of DCMs by estimating effective connectivity at different time points. Table 1 shows the brain networks through final models selected under normal procedure (using full time series length) as well as during estimating the temporal evolution of DCMs. The changes in the estimated model parameters reflect the change in dynamic effective connectivity over time. The results show differences in effective connectivity networks because of changes in posterior expectations as calculated by DCM.Ep. In this case, the DCMs estimated from time point 181-360 seconds had a lower excitatory influence between insula (Ins) and amygdala (Amg) as selected by the random effects analysis using Bayesian model selection with $p > 0.95$.

The results showed that the neuronal connection strengths between different brain regions as estimated from DCM were low when considering the full time series. Some connections were also missing in the full time series analysis for example between Ins1 and Amg1 which was present in the time window of 1-180 sec however was missing from 160-340 sec as well when considering the full time series 1-340 sec. The inhibitory self-connections, negative values in the table, ensure that this connection decays and the system reaches stability.

**4.5 Supplementary material**
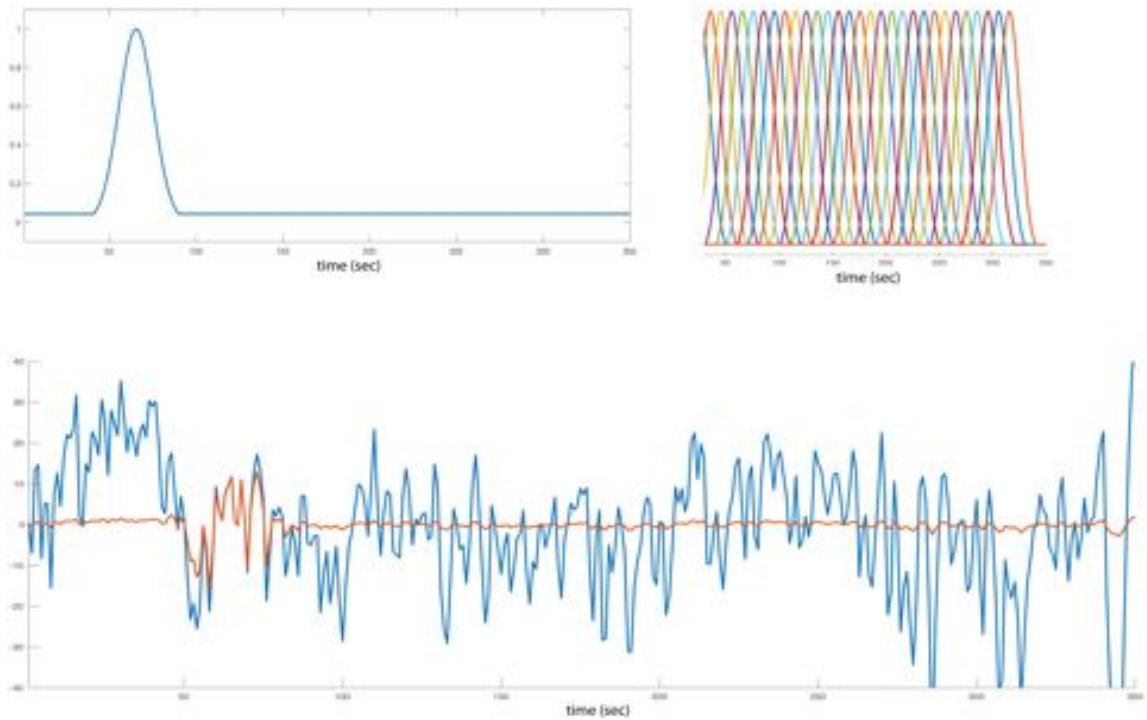


Supplementary Figure 1

    (a) Generation of Gaussian window with shift 40 seconds and length 50 seconds, as well as multiple Gaussian windows that were generated each with a shift of 10 seconds. (b) Application of the Gaussian window of length 50 second and shift 40 second as shown above, to the original fMRI time series. Signal in blue is the original fMRI time series, and the signal in red is the final multiplied signal that is obtained by multiplying the Gaussian window with the original fMRI time series.
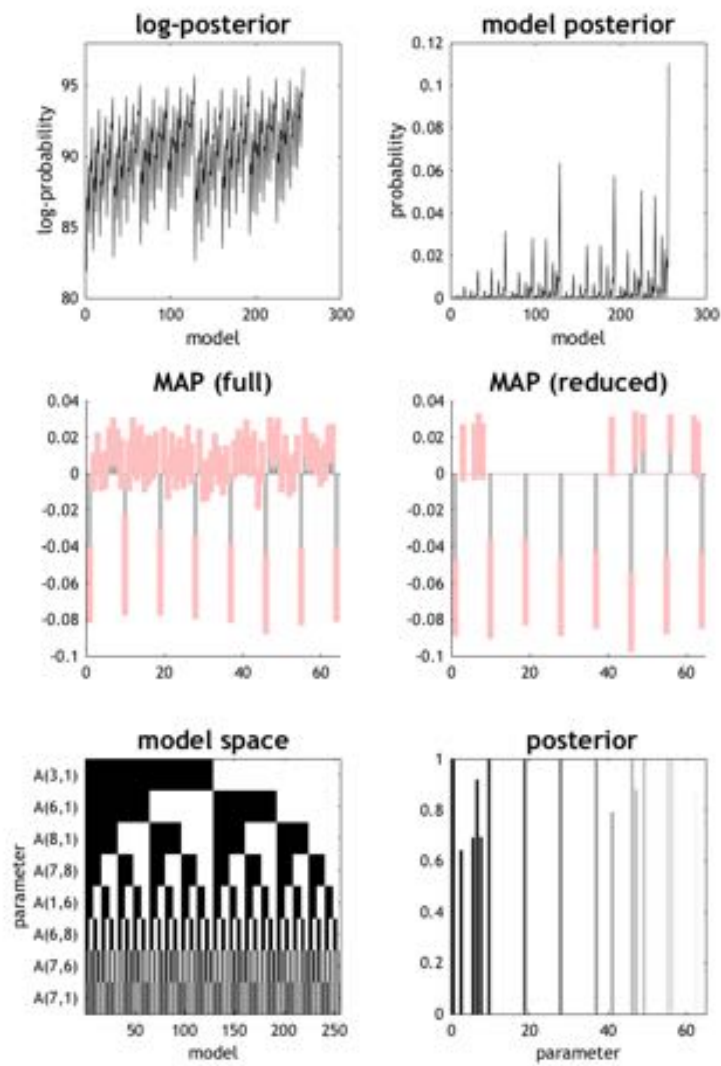
Supplementary Figure 2: Figure showing the model space and bayesian model reduction to estimate DCM parameters at group level.
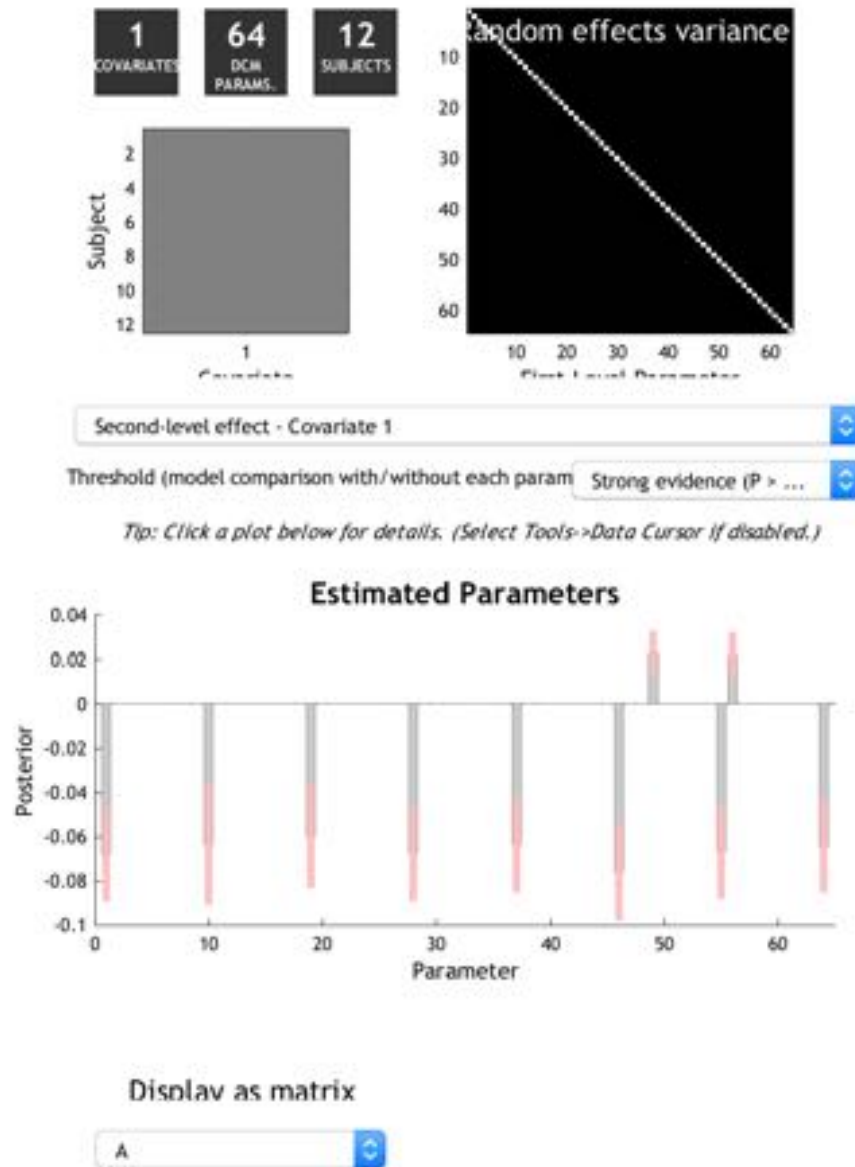
Supplementary Figure 3: Estimated DCM parameters with strong evidence using the whole time series (1-340 sec.)

## 4.6 References

[1] K.J. Friston, B. Li, J. Daunizeau, and K.E. Stephan, Network discovery with DCM. Neuroimage 56 (2011) 1202-21.

[2] J. Kahan, and T. Foltynie, Understanding DCM: ten simple rules for the clinician. Neuroimage 83 (2013) 542-9.

[3] K.E. Stephan, W.D. Penny, R.J. Moran, H.E. den Ouden, J. Daunizeau, and K.J. Friston, Ten simple rules for dynamic causal modeling. Neuroimage 49 (2010) 3099-109.

[4] X. Di, and B.B. Biswal, Dynamic brain functional connectivity modulated by resting-state networks. Brain Struct Funct 220 (2015) 37-46.

[5] A.A. Ioannides, Dynamic functional connectivity. Curr Opin Neurobiol 17 (2007) 161-70.

[6] A. Razi, J. Kahan, G. Rees, and K.J. Friston, Construct validation of a DCM for resting state fMRI. Neuroimage 106 (2015) 1-14.

[7] B. Biswal, F.Z. Yetkin, V.M. Haughton, and J.S. Hyde, Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med 34 (1995) 537-41.

[8] T. Koenig, L. Prichep, D. Lehmann, P.V. Sosa, E. Braeker, H. Kleinlogel, R. Isenhart, and E.R. John, Millisecond by millisecond, year by year: normative EEG microstates and developmental stages. Neuroimage 16 (2002) 41-8.

[9] D. Lehmann, P.L. Faber, S. Galderisi, W.M. Herrmann, T. Kinoshita, M. Koukkou, A. Mucci, R.D. Pascual-Marqui, N. Saito, J. Wackermann, G. Winterer, and T. Koenig, EEG microstate duration and syntax in acute, medication-naive, first-episode schizophrenia: a multi-center study. Psychiatry Res 138 (2005) 141-56.

[10] E.A. Allen, E. Damaraju, S.M. Plis, E.B. Erhardt, T. Eichele, and V.D. Calhoun, Tracking whole-brain connectivity dynamics in the resting state. Cereb Cortex 24 (2014) 663-76.

[11] J.R. Sato, A. Fujita, E.F. Cardoso, C.E. Thomaz, M.J. Brammer, and E. Amaro, Jr., Analyzing the connectivity between regions of interest: an approach based on cluster Granger causality for fMRI data analysis. Neuroimage 52 (2010) 1444-55.

[12] K.M. Gates, P.C. Molenaar, F.G. Hillary, N. Ram, and M.J. Rovine, Automatic search for fMRI connectivity mapping: an alternative to Granger causality

testing using formal equivalences among SEM path modeling, VAR, and unified SEM. Neuroimage 50 (2010) 1118-25.

[13] R. Goebel, A. Roebroeck, D.S. Kim, and E. Formisano, Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. Magn Reson Imaging 21 (2003) 1251-61.

[14] Z.X. Zang, C.G. Yan, Z.Y. Dong, J. Huang, and Y.F. Zang, Granger causality analysis implementation on MATLAB: a graphic user interface toolkit for fMRI data processing. J Neurosci Methods 203 (2012) 418-26.

[15] K. Friston, Dynamic causal modeling and Granger causality Comments on: the identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. Neuroimage 58 (2011) 303-5; author reply 310-1.

[16] G. Deshpande, K. Sathian, and X. Hu, Effect of hemodynamic variability on Granger causality analysis of fMRI. Neuroimage 52 (2010) 884-96.

[17] A.K. Seth, P. Chorley, and L.C. Barnett, Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling. Neuroimage 65 (2013) 540-55.

[18] X. Wen, G. Rangarajan, and M. Ding, Is Granger causality a viable technique for analyzing fMRI data? PLoS One 8 (2013) e67428.

[19] K.J. Friston, L. Harrison, and W. Penny, Dynamic causal modelling. Neuroimage 19 (2003) 1273-302.

[20] R.B. Buxton, E.C. Wong, and L.R. Frank, Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn Reson Med 39 (1998) 855-64.

[21] A.M. Bastos, V. Litvak, R. Moran, C.A. Bosman, P. Fries, and K.J. Friston, A DCM study of spectral asymmetries in feedforward and feedback connections between visual areas V1 and V4 in the monkey. Neuroimage 108 (2015) 460-75.

[22] R. Moran, D.A. Pinotsis, and K. Friston, Neural masses and fields in dynamic causal modeling. Front Comput Neurosci 7 (2013) 57.

[23] H.R. Brown, and K.J. Friston, The functional anatomy of attention: a DCM study. Front Hum Neurosci 7 (2013) 784.

[24] J. Daunizeau, L. Lemieux, A.E. Vaudano, K.J. Friston, and K.E. Stephan, An electrophysiological validation of stochastic DCM for fMRI. Front Comput Neurosci 6 (2012) 103.

[25] K.J. Friston, J. Kahan, B. Biswal, and A. Razi, A DCM for resting state fMRI. Neuroimage 94 (2014) 396-407.

[26] M.I. Garrido, K.J. Friston, S.J. Kiebel, K.E. Stephan, T. Baldeweg, and J.M. Kilner, The functional anatomy of the MMN: a DCM study of the roving paradigm. Neuroimage 42 (2008) 936-44.

[27] D.A. Pinotsis, N. Brunet, A. Bastos, C.A. Bosman, V. Litvak, P. Fries, and K.J. Friston, Contrast gain control and horizontal interactions in V1: a DCM study. Neuroimage 92 (2014) 143-55.

[28] B. Li, J. Daunizeau, K.E. Stephan, W. Penny, D. Hu, and K. Friston, Generalised filtering and stochastic DCM for fMRI. Neuroimage 58 (2011) 442-57.

[29] J. Daunizeau, K.E. Stephan, and K.J. Friston, Stochastic dynamic causal modelling of fMRI data: should we care about neural noise? Neuroimage 62 (2012) 464-81.

[30] M.G. Sharaev, V.V. Zavyalova, V.L. Ushakov, S.I. Kartashov, and B.M. Velichkovsky, Effective Connectivity within the Default Mode Network: Dynamic Causal Modeling of Resting-State fMRI Data. Front Hum Neurosci 10 (2016) 14.

[31] K. Friston, C. Chu, J. Mourao-Miranda, O. Hulme, G. Rees, W. Penny, and J. Ashburner, Bayesian decoding of brain images. Neuroimage 39 (2008) 181-205.

[32] K. Friston, J. Mattout, N. Trujillo-Barreto, J. Ashburner, and W. Penny, Variational free energy and the Laplace approximation. Neuroimage 34 (2007) 220-34.

[33] C.F. Beckmann, and S.M. Smith, Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imaging 23 (2004) 137-52.

[34] S.M. Smith, K.L. Miller, G. Salimi-Khorshidi, M. Webster, C.F. Beckmann, T.E. Nichols, J.D. Ramsey, and M.W. Woolrich, Network modelling methods for FMRI. Neuroimage 54 (2011) 875-91.

[35] K.J. Friston, V. Litvak, A. Oswal, A. Razi, K.E. Stephan, B.C. van Wijk, G. Ziegler, and P. Zeidman, Bayesian model reduction and empirical Bayes for group (DCM) studies. Neuroimage 128 (2016) 413-31.

[36] K. Friston, P. Zeidman, and V. Litvak, Empirical Bayes for DCM: A Group Inversion Scheme. Front Syst Neurosci 9 (2015) 164.

# 5. Deep learning classification and support vector machines (SVM) to classify animals according to anaesthesia regimes on the basis of stationary and dynamic functional connectivity data

**Abstract**

Single subject machine learning classification using fMRI data is challenging due to the high dimensionality and low information content in the fMRI data. This also holds when classifying subjects according to the dose of a drug received, as group differences are in general small and given the inter-subject variability intrinsically difficult to detect. In resting state fMRI, information derived from stationary functional connectivity analysis has been used as features in machine learning algorithms. Recently, methods for studying dynamic functional connectivity have been introduced, i.e. the connectivity information on shorter time scales assuming that brain networks and their interaction varies over time. In this work we fed features extracted from stationary as well as dynamic functional connectivity analysis derived from mice exposed to the anaesthetic isoflurane at different doses to machine learning algorithms for both support vector machines (SVM) and deep belief networks (DBN). The results show that we were able to successfully classify between anaesthetic doses using features extracted from static and dynamic functional connectivity analysis. The features extracted from dynamic functional connectivity analysis were found to be more discriminative to different anaesthetic doses. This shows the potential for the use of features based on dynamic functional connectivity analysis. A major limitation regarding the use of machine learning in the context of our study was small sample size (N=12 per group), which led to an accuracy of less than 70% for most comparisons.

## 5.1 Introduction

The use of machine learning methods for classification of subjects in to specific groups, for example corresponding to a specific disease state, has gained rapid popularity in the recent years [1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11]. Machine learning is a statistical approach for identifying differentiating features from 'training' data and using these features for assigning subjects, whose data were not contained in the training data set, to classes/groups. Ideally, machine learning algorithms expect a training data set with a high number of samples and a low number of features, in order to effectively learn the data patterns. Features that contain more information and are more distinctive among the classes are considered better.

fMRI offers high spatial resolution to study human and animal brain function non-invasively. Typically, fMRI analysis uses linear regression models with several assumptions in a so-called mass univariate approach. This approach assumes that each voxel in the brain is independent of the other voxels, and therefore neglects the basic principles of brain functions that neurons interact with each other through synaptic connections and the activity in brain voxels is dependent on many other voxels through a complex underlying function. Machine learning as a multivariate approach addresses the shortcomings of mass univariate approaches by disregarding the independence assumption over voxels, and takes in to account the interdependence of voxels. While mass univariate models can be considered as encoding models to predict brain activity from experimental context, multivariate models are often termed as decoding models to predict experimental context from brain activity. In typical machine learning classification analysis, the fMRI data is the observed surrogate of brain activity. Comparing training data with known assignment to two or several classes, machine learning approaches search for patterns (features) that enable maximal discrimination between theses classes. Then an entirely new dataset consisting of the same or similar classes is used for model validation. Given the statistical nature of machine learning algorithms a large sample size with data characterized by a low number of features (low dimensional feature vector) favor proper feature identification. However, fMRI data is high dimensional in nature and data samples are typically small in number (<200 samples or subjects). Therefore, application of machine learning algorithms to the fMRI data is challenging and bears the risk of data overfitting. This may be largely avoided by applying dimensionality reduction algorithms [12; 13; 14; 15; 16]

Appropriate feature selection for neuroimaging data has been discussed in [2]. Many of the machine learning algorithms have to be considered as "black boxes"; it is not exactly clear what the algorithms have learned. This makes them prone to finding spurious associations. A proper training comprising of neurobiologically relevant features covering all the possible data variations and large sample size, and validation is required in order to avoid such pitfalls. It has also been argued that only the features that are meaningful with regards to neurobiological aspects should be considered relevant for enhancing our understanding of the brain function[17; 18].

Anesthesia has been shown to alter the functional connectivity both in humans and animals [19; 20; 21; 22; 23; 24; 25; 26; 27; 28; 29; 30]. Dose dependent effects of isoflurane in rats have been thoroughly investigated in [31] and suggests consistent changes in functional connectivity with increasing dose of anesthetics. Similarly, we have found alterations in functional networks of the mouse brain that depend on the dose of isoflurane administered, in particular loss of interhemispheric homotopic correlation and modular segregation (Chapter 3 of the thesis). In this work we evaluated, whether these changes in resting state patterns were of sufficient discriminative power to enable assignment of individual animals to specific isoflurane dose groups using machine learning algorithms, in particular Deep Belief Network (DBN) and Support Vector Machines (SVM). For feature extraction both stationary and dynamic functional connectivity analysis has been used.

### 5.1.1 Support Vector Machines (SVM)

SVM is considered one of the famous machine learning algorithm that identifies a separating hyperplane that maximizes the margin between the classes. Given a set of data points from two classes, the first step is identifying the support vectors, that are subset of training samples closest to the data from other group, and then maximizing the margin between the support vectors in a linearly separable plane. The problem of finding the optimal hyper plane is an optimization problem is solved by iteratively moving the hyperplane to maximize the distance between the groups. The hyperplane that represents the largest separation between the classes or that maximizes the distance from it to the nearest data point on each side, is called maximum-margin hyperplane. Figure 1 below shows data points from two groups that are linearly separated by a maximum margin hyperplane.
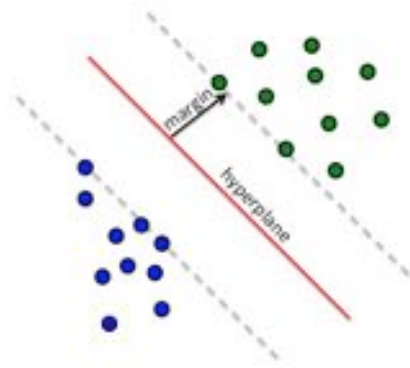
Figure 1: SVM separating the hyperplane with maximum margin between the two groups.

SVM is a linear classifier; this means SVM is capable of finding a hyperplane if the feature space is linearly separable. However, most of the real world data is non-linear in nature. In order to address this issue, an extension of SVM has been proposed in the literature with the use of kernels e.g. radial basis function to find the separating hyperplane [32]. The idea of the kernels is to apply a nonlinear transformation to the nonlinear input space, the result of which is a linearized feature space, in which SVM aims at identifying a hyperplane that linearly separates the classes in the transformed space. This is shown in Figure 2.



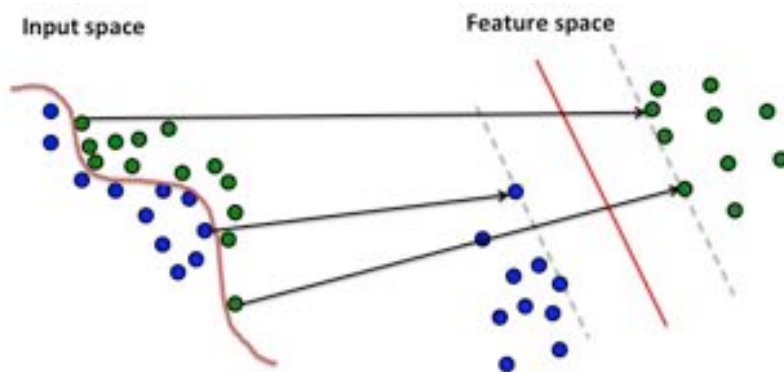Figure 2: Non linear SVM transformed the input space in to feature space using a non linear kernel. In the feature space, the two classes are linearly separable.

### 5.1.2 Deep learning

A Deep Belief Network (DBN) is a deep learning algorithm that consists of many layers of Restricted Boltzmann Machines (RBMs). RBMs are two-layer neural nets with the first layer of the RBM being the visible, or input, layer, and the second one

123

being the hidden layer. While each 'neuron' in each layer is connected to each neuron of the other layer, there is no connection between neurons within the same layer (no intra-layer communication, therefore 'restricted'). Each node (neuron) is a locus of computation that processes input, and begins by making stochastic decisions about whether to transmit that input or not. A DBN network can be decomposed into a series of RBMs, in which in a subsequent step the hidden layer is considered the input layer of the next RBM and so forth. A DBN is a generative graphical model neural network that can learn a probability distribution over its set of inputs. Fig. 5 shows a neural network with two layers. The lower layer $x$ is connected to the hidden layer $h$ through weights $W$. If the learning error is high, the system adapts, altering the weights in order to improve subsequent results. This procedure is expected to capture the high level representation of the original data at the output of the last layer [33; 34]. RBM uses learning method based on Gibbs sampling, that allows obtaining a sequence of observations without direct sampling, and adjusts its weights at each layer to minimize the reconstruction error.
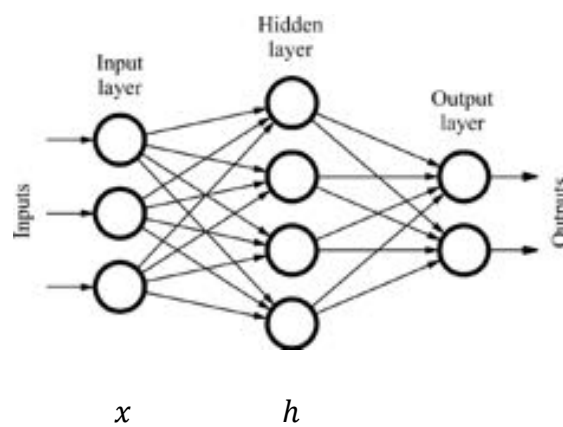


Figure 3: A simple feed forward neural network is shown. The first layer is the input layer, also called the visible layer, followed by the middle layer also called the hidden layer. Inputs are transformed to the hidden using a weighting function W. Neural network output is generated at the output layer. From Quiza and Davim [35]. Reproduced with permission.

If $x$ defined as the visible layer and $h$ is defined as the hidden layer as shown in the Fig. 3, then probability distribution of the model as indicated in [34] is given by

$$P(x, h) = \frac{e^{-E(x,h)}}{Z} \tag{1}$$

where $x$ and $h$ are stochastic binary variables. The energy function $E(x, h)$ and the partition function $Z$ are defined as

$$E(x_i, h_j) = -b_i \cdot x_i - c_j \cdot h_j - h_j \cdot W_{ij} \cdot x_i \tag{2}$$

$$Z(x, h) = \sum_{i,j} e^{-E(x_i, h_j)} \tag{3}$$

where $b$ and $c$ are the biases of the visible layer and the hidden layer respectively. The sum over $i, j$ represents all possible states of the model.

The condition probability of one layer given the other one is:

$$P(h|x) = \prod_i P(h_i \,|x) \tag{4}$$

Notice that if one layer is given, the distribution of the other layer is factorial. Since the neurons are binary the probability of a single neuron being on is given by

$$P(h_i = 1 \,|\, x) = sigm(c_i + W_i\, x) \tag{5}$$

Similarly the conditional probability for the visible layer can be found

$$P(x_i = 1 \,|\, h) = sigm(b_i + W_i\, h) \tag{6}$$

RBM uses a probabilistic version of the normal sigmoid neuron activation function with the goal to maximize the log likelihood of the training data i.e. make the model generate data like the training data.

The gradient of the negative log probability of the visible layer with respect to the model parameters $\theta$ is:

$$\frac{\partial}{\partial \theta}\left(-log\, P(x)\right) = \frac{\partial}{\partial \theta}\left(-\log \sum_h P(x, h)\right) \tag{7}$$

The derivation has been shown in [34] and it is reduced to

$$\frac{\partial}{\partial b}\left(-log\, P(x)\right) = -\mu_1[-x|x] - \mu_1[-x\,] \tag{8}$$

$$\frac{\partial}{\partial c}\left(-log\,P(x)\right) = -\mu_1[-h|x] - \mu_1[-h] \tag{9}$$

where $\mu_1$ is a function returning the first moment or expectation value.

Model parameters are optimized in a recursive manner,

$$x^{(n)} = P(x|h^{(n-1)}) \tag{10}$$
$$\downarrow$$
$$h^{(n)} = P(x^{(n)}|h) \tag{11}$$
$$\downarrow$$
$$x^{(n+1)} = P(x|h^{(n)}) \tag{12}$$

The complete derivation has been shown in [34]. The superscripts $1 \leq n \leq N$ denote the number of the iteration. At each iteration the entire layer is updated. The model is initialized at some arbitrary state, $N$ iterations are carried out, $N$ being a large number. The termination criterion used was, either error was smaller than a predefined threshold value or the set number of iterations are performed as defined by epochs. In order to make this efficient we initialize the model at a training sample, iterate one step, and use this as our sample. This is the contrastive divergence algorithm as introduced by Hinton et al. [33] with one step (CD-1). The logic is that, as the model distribution approaches the training data distribution, initializing the model with a training sample may be considered as approximation for model convergence. Finally, for computational efficiency, we will use stochastic gradient descent instead of the recurrent batch update rule described above. The final algorithm has been shown in [34].

The learning procedure consists of several steps of Gibbs sampling and adjusting the weights to minimize reconstruction error. All this procedure is in an unsupervised learning fashion (the algorithm is kept unknown about which data point belongs to which group), where the training labels are unknown to the algorithm. The output of the generative model is then used in a neural network classifier in a supervised learning fashion (training labels are known to the algorithm; i.e. the algorithm is

126

informed of which data point belongs to which group) in a procedure called stacking. This ensures that the algorithm learns from the labels/outputs and modifies the learned layer-wise parameters accordingly.

In deep belief network we train the RBM as described above and then train another RBM using the first RBM's hidden layer as the second RBMs visible layer and so on.

### 5.1.3 Dynamic functional connectivity analysis

Typically when extracting information on FC on the basis of rs-fMRI data, the whole time series is used to compute correlation values, assuming that the functional networks are stationary in time. That essentially means that information on events/interactions that occur at a shorter time scale is averaged out. Yet, neural processing occurs at a much shorter time scale and EEG studies led to the identification of microstates that involve network rearrangements in the millisecond time domain [36; 37]. Dynamic functional connectivity (dFC) analysis [38] is a concept that aims at estimating the FC changes over relatively short time intervals constituting a small fraction of the full time series recorded. Limiting factors when using fMRI data are the low-pass temporal filtering imposed by the hemodynamic response function and the sequential nature of MRI data acquisition, which limits the time required to capture an image volume to one or a few seconds. Despite these limitations, dFC has been suggested as a more accurate representation of functional brain networks since brain networks are dynamic. There have been many algorithms proposed in the literature to estimate dFC from resting state fMRI data. Sliding window based correlation analysis is a simple method to define a window of duration $\tau$, for which a conventional FC analysis is carried out. The window must comprise a sufficient number of sample points to allow for a meaningful correlation analysis, yet should not be too long in order to still capture dynamic aspects. The window is then shifted by an increment $\Delta t$ (typically one sampling point), while maintaining its length $\tau$ and the FC analysis is carried out again. Repeating this procedure several times allows monitoring the changes in FC networks over time. These resulting correlation matrices are then z-transformed and are called as dFC matrices $\mathbf{Z}_s(t_i; \tau)$. This procedure is illustrated in Figure 4.
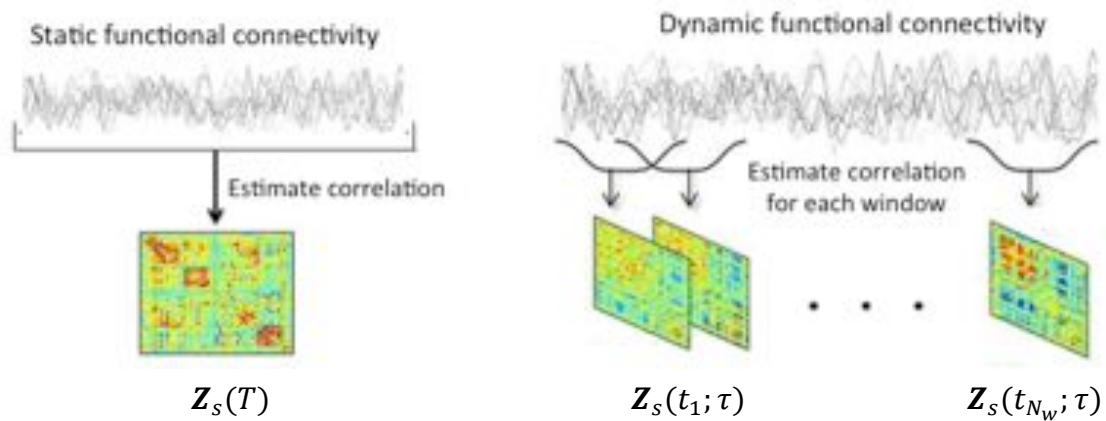
Figure 4: Static FC matrix $\mathbf{Z}_s(\quad)$ is obtained for each subject $s$ by computing the coefficients of correlation between the whole time series of duration $T$ for all pairs of regions (voxels) $i$ and $j$. dFC is estimated by carrying out the analogous analysis for time windows of duration $\quad$ separately, the starting point of which are incremented by $\Delta t$, yielding correlation matrices $\mathbf{Z}_s(t_i; \tau)$ with $1 \le i \le N_w$, $N_w = (T - \tau)/\Delta t$ being the number of windows. Adapted from [39].

Typically dFC had been used for analysing resting state fMRI data, however in 2010 Sakoglu et. al. applied dFC technique to the task related fMRI data [40]. They developed an approach that uses spatial ICA to estimate correlations between windowed time-courses of different brain networks (components). After estimating the spatial ICAs, they used the method developed by Calhoun and colleagues (http://mialab.mrn.org/software) based on the maximal lagged correlation approach [40] for estimating both stationary and dFC. The dFC was estimated by applying the window size of 64 time points.

**Algorithm**

We used the algorithm developed by Leonardi et al for estimating the dFC states. Feasibility of this algorithm has been demonstrated in few recent studies [41; 42; 43].

Sliding-window correlation between the time series x and y using the formula

$$r_{xy}(t) = corr(x[t, t + \tau], y[t, t + \tau]), \tag{13}$$

where $\tau$ was the window length in TRs and the window was shifted by $\Delta t$ for each estimation. The coefficients of correlation were computed according to Pearson correlation as given in eq. (14)

$$r_{xy} = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}, \tag{14}$$

where $r$ is the coefficient of correlation between $x$ and $y$.

The correlation coefficient computed were then z-transformed according to

$$z_{xy}(t) = \text{atanh}(r_{xy}(t)) \tag{15}$$

The dFC analysis is carried out for each subject individually. The resulting correlation matrices for each window were vectorized yielding a $K \ x \ N_w$ matrix $\boldsymbol{C}_s$ for each subject and condition $s(1 \leq s \leq N_s)$, with $K$ being the number of pair-wise correlations and $N_w = (T - \tau)/\Delta t$ the number of windows. The matrix $\boldsymbol{C}_s$ was row-wise de-meaned to solely address the fluctuations of connectivity over time regardless of their mean value [41]. This has been discussed further in Grandjean et al. (2017) with a detailed comparison of with and without this de-meaning step. The matrix $\boldsymbol{C}_s - \overline{\boldsymbol{C}}_s$ after demeaning then represents the increase/decrease in correlation strengths with respect to the mean (stationary FC).

The $N_s$ subject and condition-specific matrices $\boldsymbol{C}_s - \overline{\boldsymbol{C}}_s$ were then concatenated into a data matrix $\boldsymbol{X'}$ of dimension $K$ times the product $N_w \cdot N_s$

$$\boldsymbol{X'}^{(K \times (N_w \cdot N_s))} = \left( \boldsymbol{C}_1 - \overline{\boldsymbol{C}}_1 \big| ... \big| \boldsymbol{C}_{N_s} - \overline{\boldsymbol{C}}_{N_s} \right) = \left( x'_1 \big| ... \big| x'_{N_w \cdot N_s} \right) \tag{16}$$

with $x_i$ being vectors of length $K$.

## 1. Estimation of eigenconnectivities

The matrix $\boldsymbol{X'}$ describing the dynamic functional connectivity for a group of $N_s$ subject (or subjects times conditions) can be reduced, e.g. by principle component analysis (PCA), with a few PCs accounting for most of the variation across in the data matrix $\boldsymbol{X'}$. Leonardo et al. (2013) derived such components by estimating the eigenvectors and eigenvalues of $\boldsymbol{X'}$ according to

$$X' \cdot \widetilde{X}' = U \cdot \Lambda \cdot \widetilde{U} \tag{17}$$

with $U$ being a unitary matrix containing the orthonormal eigenvectors as columns and $\Lambda$ being a diagonal matrix $\Lambda_{ik} = \delta_{ik} \cdot \lambda_i$ comprising the eigenvalues $\lambda_i$ (Fig. 5). The eigenvectors, i.e. the columns of $U$, were termed eigenconnectivities.
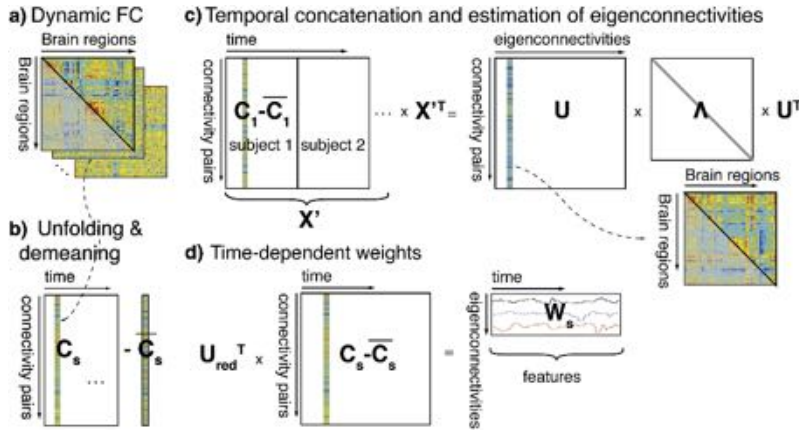


Figure 5: Schematics illustrating the dFC method. (from Leonardi et al 2013, reproduced with permission). (a) Dynamic FC between $n$ brain regions was computed as sliding window correlations between the activities of all regions for each subject. (b) The upper triangular part of each correlation matrix $_s(t_i; \tau)$ of dimension $n \times n$ was unfolded and concatenated across time to form a dynamic FC matrix $C_s$ of dimension $K \times N_w$, $K$ being the number of pairwise correlations and $N_w$ the number of time windows The mean correlation value across time was subtracted from all connectivity pairs (i.e., row-wise centering) yielding the matrix $C_s - \overline{C}_s$. (c) The dynamic FC matrices were concatenated across subjects to form a matrix $'$, for which eigenvalues $\lambda_i$ and eigenvectors (columns of unitary matrix $U$) were computed, so-called eigenconnectivities. Eigenconnectivities can be visualized after rearranging the eigenvectors of dimension $K \times 1$ them into a matrix of dimension $n \times n$ and symmetrizing, representing a indicating the interaction matrix corresponding to a specific eigenvalue in the basis of the brain regions. (d) The weight matrix $W_s$ containing the time dependent weights of each eigenconnectivity and subjectwas calculated by projecting the demeaned dynamic subject-specific FC matrix $C_s - \overline{C}_s$ onto a few eigenconnectivities, accounting for the major part of variation in the data.

Other algorithms may be used to estimate eigenconnectivities. For example, Grandjean et al. (2017) estimated eigenconnectivities based on dictionary learning algorithm. Dictionary learning algorithm aims to find the sparse representation of the input data and has been shown to produce good results in the fields of image classification and processing [44; 45]. In this work, we used dictionary-learning algorithm to estimate eigenconnectivities. Figure 6 shows that functional correlation matrices can be transformed in to dFC states by applying dictionary learning algorithm over it
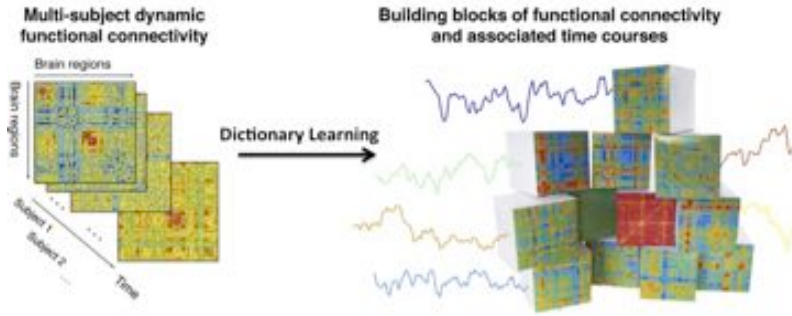


Figure 6: Functional correlation matrices estimates for each window and different subjects are concatenated in order to apply dictionary-learning algorithm for the estimation of dynamic functional states, or atoms; also referred to as the basic building blocks of dFC. Adopted from Leonardi et al 2013.

Dictionary learning algorithm involves the generation of a dictionary with

$$\boldsymbol{D}^{(K \times M)} = (d_1 | \dots d_M) \tag{18}$$

with $M$ $K$-dimensional column vectors as simple building blocks, the atoms, capturing whole-brain connectivity to a large extent, and a coefficient matrix of the concatenated subjects.

$$\boldsymbol{W}^{M \times (N_w \cdot N_s)} = \left(w_1 | \dots w_{N_w \cdot N_s}\right). \tag{19}$$

The columns of the concatenated coefficient matrix describe the approximation of the set of signals in $\boldsymbol{C}_s - \overline{\boldsymbol{C}}_s$ to combine these atoms. Like the number of IC components in ICA analysis, $M$ can be chosen arbitrarily, but following other papers we generated M=20 atoms.

Cost function minimization was implemented using a sparsity-enforcing algorithm

giving $W$ and $D$ as outputs.

$$f(W, D) = \frac{1}{N_w \cdot N_s} \sum_{i=1}^{N_w \cdot N_s} \|w_i\|_1 \, s.t. \, \|c_i - D \cdot w_i\|_2^2 \leq r \tag{20}$$

In the above equation cost function tries to minimize the squared distance between $c_i$ and the product of $D$ and $w_i$ and upper bounding it with a regularization parameter $r$

In order to retrieve easily interpretable atoms, we required them to be energy-bounded and positive by imposing the set of constraints

$$C \triangleq \{ D \in \mathbb{R}^{K \times M} \, s.t. \, \forall j = 1, \dots, M, \forall i = 1, \dots, K, \; \tilde{d}_j \cdot d_i \leq 1, D_{ij} \geq 0 \} \tag{21}$$

The first part of the equation shows $D \in \mathbb{R}^{K \times M}$ that the dictionary atoms belongs to the real numbers with the matrix dimension of K (number of pairwise correlation) times M (number of atoms estimated). Furthermore, eq. (21) states that the atoms are energy-bounded, $d_i \leq 1$, and positive, $D_{ij} \geq 0$.

Dictionary learning was performed 100 times (or folds; 400 iterations for the first fold, 200 for the subsequent ones). The obtained atoms were matched to the first fold using the Hungarian algorithm (Kuhn, 2010), with spatial correlation as the similarity metric. In order to get robust results, the fold instances exhibiting similarity above the median value across all folds were averaged for each atom

Animal-specific time-dependent contributions of atoms were obtained by back-projection of the dictionary onto the original dFC dataset using least-square fitting.

**5.2 Method**

*5.2.1 Animals, preparation, and anesthesia*

The experiments were performed in compliance with Swiss laws on animal protection and approved by the Veterinary Office of the Canton of Zurich. Female C57BL/6 mice (Janvier, Le Genest-St Isle, France) between 10 and 15 weeks old were studied. All mice were initially anesthetized with isoflurane in a 20% $O_2$ / 80% air mixture: 3.5% for induction, 2% for endotracheal intubation and during set-up on the animal bed. Throughout the duration of the experiment, animals were mechanically ventilated using a small animal ventilator (CWE, Ardmore, USA) with a 20% $O_2$ / 80% air mixture at a rate of 80 breaths/min, a respiration cycle of 25% inhalation, 75% exhalation, and an inspiration volume of 1.8 ml/min. The head was placed with the animal's incisors secured over a bite bar and fixated by ear bars, ophthalmic ointment was applied to the eyes, and a rectal temperature probe was inserted to keep the animal at $36.5 \pm 0.5$ °C by means of a warm-water circuit integrated into the animal holder (Bruker Biospin GmbH, Ettlingen, Germany). The tail vein was cannulated for intravenous (i.v.) administration of anesthetics and the neuromuscular blocking agent pancuronium bromide (Sigma-Aldrich, Steinheim, Germany).

Two independent set of studies were performed to evaluate the effects of isoflurane in a dose dependent manner.

Group 1 (Dose-escalation in individual mice): Twelve animals were used in the experiment. Isoflurane (Abbott, Cham, Switzerland) was sequentially increased from 1.1% to 1.3%, 1.5% and 2.0% in a 20% $O_2$ / 80% air mixture for each individual mouse. After each incremental increase of isoflurane concentration, there was a 10 min interval for equilibration before the fMRI data acquisition was started. Mice remained in the scanner throughout the duration of the experiment.

Group 2 (Single dose for each mouse): Isoflurane was administered in a 20% $O_2$ / 80% air mixture at a single defined dose per mouse in order to avoid any accumulation effects. Doses used were 1.1% (N=10 mice) and 1.5% (N=18).

Each animal received an i.v. bolus injection of 0.5 mg/kg pancuronium bromide dissolved in saline (0.5 mg/3 ml) followed by a continuous infusion of 0.5 mg/kg/h of pancuronium bromide corresponding to an infusion rate of the solution of 3 ml/kg/h.

Animal preparation, anesthesia protocols, and the conditions during resting-state measurements were identical for the fMRI experiments and for the assessment of systemic physiological parameters. Approximately 20 min were used for animal preparation, and a further 20 min for preparatory MRI scans. Subsequently, rs-fMRI

data sets of 6 min duration each were acquired. After the experiments, time for recovery from anesthesia and pancuronium bromide administration was provided for all the animals.

### 5.2.2 fMRI

MRI/fMRI experiments were carried out using a Bruker Biospec 94/30 small animal MR system (Bruker BioSpin MRI, Ettlingen, Germany) operating at 400 MHz (9.4 T). A four-element receive-only cryogenic phased array coil (Bruker BioSpin AG, Fällanden, Switzerland) was used in combination with a linearly polarized room temperature volume resonator for transmission (Bruker BioSpin MRI, Ettlingen, Germany). Anatomical images acquired in the sagittal and horizontal direction allowed exact positioning of 12 adjacent coronal slices of 0.5 mm slice thickness, which were used for the rs-fMRI scans. A gradient-echo echo-planar imaging (GE-EPI) sequence has been used for rs-fMRI data acquisition with field of view=16x7 mm2, matrix dimensions=80x35, TR=1 s, TE=12 ms, flip angle=60 degrees. The time series acquired was of 360 s length.

### 5.2.3 Data Processing

All the preprocessing was performed using tools from FMRIB's Software Library (FSL version 5). FSL's recommended pre processing pipeline was used. Motion correction, removal of non-brain structures, high pass temporal filtering with sigma = 75.0s; pre-whitening and global spatial smoothing of 0.2 mm was applied as part of the pre-processing.

After the pre-processing, the functional scans were aligned to the high-resolution anatomical QBI (Queensland Brain Institute) template using linear affine and nonlinear diffeomorphic transformation registration as implemented in ANTs (ANTs. v 1.9; http://picsl.upenn.edu/ANTS/).

ICA analysis allows determining spatial maps from fMRI data by concatenating all the subjects from all groups. [46; 47] describes this method using probabilistic ICA. The spatial IC maps consists of the representative anatomical regions of the fMRI data and can be used for further analysis. We used FSL's MELODIC for probabilistic independent component analysis [47]. The multi-session temporal ICA concatenated approach, as recommended for resting state data analysis, allowed to input all subjects

134

from all the groups in a temporally concatenated fashion for the ICA analysis. ConcatICA yielded different activations and artifact components without the need of specifying any explicit time series model.

A total of 50 independent components (IC maps) were extracted and the mixture model approach was applied on these estimated maps to perform for inference analysis. An alternative hypothesis test based on fitting a Gaussian/gamma mixture model to the distribution of voxel intensities within spatial maps [46; 48] was used to threshold the IC maps. A threshold of 0.5 ($p < 0.5$) was selected for the alternative hypothesis in order to assign equal 'cost' to false-positives and false-negatives. Out of the 50 independent components (IC maps), only around 20 numbers of components were selected, while the components that overlapped with vascular structures and ventricles were excluded from further analysis. Similarly regions at the brain surface, which are prone to be affected by the motion artifacts, were excluded.

### 5.2.4 Dynamic Functional Connectivity analysis

**5.2.4.1 Description of the parameters and hyperparameters**

Window length was set at 40 seconds. Since TR = 1 sec for our data, that means window length = 40 TRs. The step size was set at 1 TR. We set the number of components to keep at K=20. Number of dictionary learning folds were set at 30. Maximum number of restarts allowed for getting convergence was set at 5. We used least square projection method of least fitting for the dictionary learning backfitting.

**5.2.4.2 Significance analysis for group differences**

Statistical analysis was performed using inhouse MATLAB codes. We used the absolute sum of time-dependent contributions of atoms relative to the number of sliding-window frames as a response variable. Isoflurane dose (1.1%, 1.3% 1.5% and 2.0%) was modelled as a fixed effect, and the individual animal intercepts were modelled as random effects. A contrast was designed to compare isoflurane dose. False discovery rate (FDR) was used to correct for multiple comparisons performed across atoms. The initial anesthetic dose (isoflurane 1.1%) was compared against all other anesthetic doses (isoflurane 1.3%, 1.5% and 2.0%). Figure 3 shows the atoms where significance was achieved in all the 3 comparisons in the same atoms (1.1 vs 1.3; 1.1 vs 1.5; 1.1 vs 2.0%). Similarly we also compared the two extreme doses

(isoflurane 1.1% vs 2.0%) and significant atoms are shown in figure 4 for this comparison.

### 5.2.5 SVM

We used the MATLAB implementation of SVM with Sequential Minimal Optimization solver. We tested different non-linear [49] kernels of SVM, however since the best preliminary results on this data were achieved using the linear SVM, we used linear SVM in this project.

### 5.2.6 DBN

We used the matlab based DBN implementation by Palm [34]. We used the sigmoid activation function. We used DBN stacked on a neural network. We used a 66-45-35 hidden unit DBN that means that we used the following configuration. The first RBM has a visible layer of 66 units and 45 hidden units. The second RBM has a visible layer of 45 units and 25 hidden units. We used the weights of the DBN to initialise a neural network. Number of epochs was set to 500. Batch size was 4 and momentum was put as 0.7

### 5.2.7 Pre-processing of the features

The two feature generation steps used in this work are discussed below

### 5.2.7.1 z-correlation

Each IC map obtained from MELODIC was averaged across the time dimension to obtain an averaged time series per IC. Each of these time series was then used to calculate the Pearson correlation against all other time series obtained from other IC maps. The correlation coefficients were then converted in to z transform. These z-transformed correlation coefficients were fed in to SVM and DBN as feature vectors for classification.

### 5.2.7.2 dFC vector

We used the same IC maps as for calculating z correlations. Dictionary learning was used [41; 42; 43] for evaluating dynamic aspects in functional connectivity. The atoms (dFS) were estimated by concatenating the results obtained from all groups and then used as regressors for analysing rs-fMRI data sets for each individual anaesthesia dose, similar to the procedure applied in ICA analysis and dual regression. Twenty atoms

136

explaining approximately 50% of the variability have been considered for the analysis. The weights of each atom were concatenated to form the feature vectors.

### 5.2.8 Validation

We performed randomised permutations with leave one out cross validation (LOO) on 'n' subjects; that means during each permutation loop one of the subject was randomly taken out and LOO algorithm was applied to the remaining 'n-1' subjects. The left out subject was then mixed in to the group and in the next loop another randomly selected subject was taken out and LOO was performed on the remaining 'n-1' subjects. This procedure was repeated number of times to generate confidence interval for the prediction accuracy.

### 5.3 Results

Twenty-five IC maps were obtained from the MELODIC and attributed to the modular structures based on modules defined in previous work [50; 51] (Fig. 7).
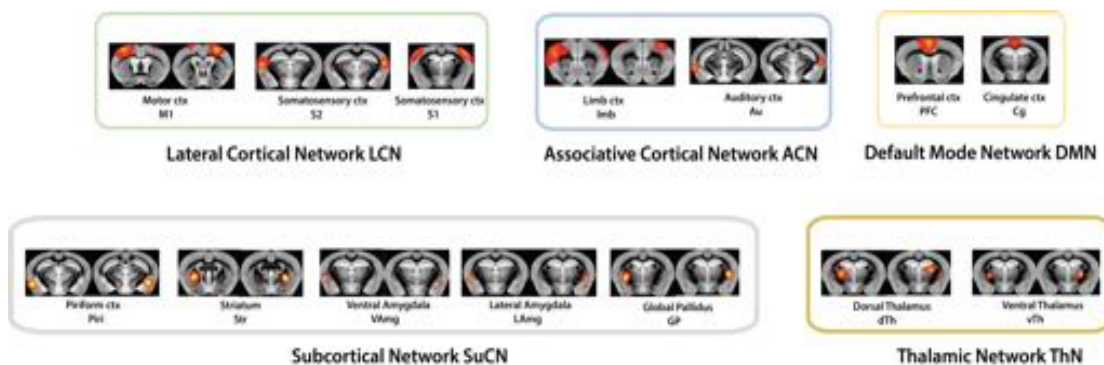


**Figure 7: Allocation of ICs to functional modules.** The 25 ICs that have been identified; have been attributed to the modules defined in [50; 51], i.e. lateral cortical network (LCN), associated cortical network (ACN), default mode network (DMN), subcortical network (SuCN) and thalamic network (ThN).

We used binary classification across all possible pairwise combinations of groups of mice exposed to a specific dose of isoflurane. Validation was done using 500 randomised permutations with LOO cross validation.

### 5.3.1 Classification results

DBN and SVM classification results using z-correlation coefficients as well as using dFC components as features are summarised in Table 1. Not surprisingly, comparison

137

to the group of mice receiving the highest isoflurane dose yielded the best classification results. Interestingly, SVM outperformed DBN in all cases analysed.

| Isoflurane dose (%) | z correlation | | dFC | |
|---|---|---|---|---|
| **dose(1) vs dose(2)** | **DBN** | **SVM** | **DBN** | **SVM** |
| **1.1 vs 1.3** | 59±6 | 66±4 | 60±5 | 70±5 |
| **1.1 vs 1.5** | 61±3 | 71±5 | 61±6 | 80±7 |
| **1.1 vs 2** | 67±4 | 78±5 | 69±5 | 78±6 |
| **1.3 vs 1.5** | 50±6 | 67±8 | 61±3 | 59±7 |
| **1.3 vs 2** | 57±7 | 76±9 | 71±5 | 90±8 |
| **1.5 vs 2** | 61±4 | 78±7 | 59±4 | 79±5 |

Table 1: Accuracy (in %) of classification of mice receiving isoflurane at different doses. Accuracy for pairwise comparison of treatment groups with corresponding values for standard deviation using DBN and SVM classification algorithms and either z correlation coefficients and dynamic functional states (dFS) as the features Validation was performed using randomised permutated LOO.

In a second analysis, the training set comprised data from mice of group 1 receiving 1.1% and 1.5% isoflurane; the trained network was the used to classify an independent dataset of group 2 comprising data from mice exposed to 1.1% and 1.5% dose of isoflurane (Table 2). The classification results were largely comparable to those of the first study.

| z correlation | | dFC | |
|---|---|---|---|
| DBN | SVM | DBN | SVM |
| 62 | 65 | 67 | 75 |

Table 2: Accuracy (in %) of classification of mice receiving isoflurane at doses 1.1 and 1.5%. Rd-fMRI data of mice of group 1 have been used as training data sets. While the analysis was carried out for animals from group 2 receiving the same isoflurane exposure. Accuracy of classification of test dataset (group 2) is given for

DBN and SVM classification algorithms using features based on both static (z correlation) and dynamic (dFC) functional connectivity analysis.

## 5.4 Discussion

Anaesthesia has been known to affect functional connectivity in rats and mice [19; 20; 21; 23; 24; 26; 30; 31; 52; 53; 54; 55; 56; 57; 58; 59]. Functional connectivity differences between different doses of anaesthetics have also been shown using seed based analysis as well as network modelling approaches.

Prediction at the individual subject level using neuroimaging data is a challenging problem due to the high dimensionality of fMRI data, low information content in the data (high level of undesired information in the signal) and low sample size (small number of subjects available for training and testing). With subtle changes in the dose of isoflurane and low number of sample (subjects) for training, it becomes even more challenging for the machine learning algorithms to identify and learn patterns that may reliably allocate a subject to a specific group.

In this work we used SVM and DBN classification algorithms to derive the dose of isoflurane administered to a mouse according to features derived from stationary and dynamic functional connectivity analysis of resting-state fMRI data. As expected, graded changes in anaesthesia level imposed minimal changes in connectivity patterns rendering differentiation according to dose level difficult. Nevertheless, we found that features based on dynamic functional connectivity can successfully classify mice (greater than chance level) according to the dose of isoflurane received, in particular when for comparisons between low dose (1.1 to 1.5%) and high dose of isoflurane (2%). Dynamic functional connectivity analysis outperformed static functional connectivity analysis for most of the classification analysis in this work. This shows the potential of the use of features based on dynamic functional connectivity for classification problems in neuroimaging. However, though the accuracy at present is statistically significant (greater than chance level after randomised permutations), for practical applications (e.g. in diagnosis or treatment prediction) it must be improved.

We also found that DBN is outperformed by the linear SVM in our small dataset. DBNs are typically trained for large number of training samples in order to learn complex underlying functions of the feature space. However in this study, the number

of subjects was found to be too small for optimal training of the DBN algorithm. This also shows that for classification problems with small number of samples, linear SVM may be a better choice over DBN. For an optimal DBN training, large multi centre datasets should be acquired and shared across the research centres.

There are several other limitations of this approach. The classification accuracy could have also been further improved by providing more informative features having meaningful neurobiological basis. Furthermore, the present method only gives classification accuracies as the output, however a more sophisticated approach could have given information about the under lying neural mechanisms that were used for classification. Generative embedding [17; 18; 60] provides an excellent alternative in this case by providing more informative features as well as providing information about the underlying neuronal network.

## 5.5 References

[1] V. Aguiar-Pulido, J.A. Seoane, J.R. Rabunal, J. Dorado, A. Pazos, and C.R. Munteanu, Machine learning techniques for single nucleotide polymorphism--disease classification models in schizophrenia. Molecules 15 (2010) 4875-89.

[2] Q. Bukhari, D. Borsook, M. Rudin, and L. Becerra, Random Forest Segregation of Drug Responses May Define Regions of Biological Significance. Front Comput Neurosci 10 (2016) 21.

[3] A. Cerasa, Machine learning on Parkinson's disease? Let's translate into clinical practice. J Neurosci Methods 266 (2016) 161-2.

[4] A.M. de la Villehuchet, M. Brack, G. Dreyfus, Y. Oussar, D. Bonnefont-Rousselot, M.J. Chapman, and A. Kontush, A machine-learning approach to the prediction of oxidative stress in chronic inflammatory disease. Redox Rep 14 (2009) 23-33.

[5] M. Dyrba, F. Barkhof, A. Fellgiebel, M. Filippi, L. Hausner, K. Hauenstein, T. Kirste, S.J. Teipel, and E.s. group, Predicting Prodromal Alzheimer's Disease in Subjects with Mild Cognitive Impairment Using Machine Learning Classification of Multimodal Multicenter Diffusion-Tensor and Magnetic Resonance Imaging Data. J Neuroimaging 25 (2015) 738-47.

[6] S. Li, F. Shi, F. Pu, X. Li, T. Jiang, S. Xie, and Y. Wang, Hippocampal shape analysis of Alzheimer disease based on machine learning methods. AJNR Am J Neuroradiol 28 (2007) 1339-45.

[7] A. Rizk-Jackson, D. Stoffers, S. Sheldon, J. Kuperman, A. Dale, J. Goldstein, J. Corey-Bloom, R.A. Poldrack, and A.R. Aron, Evaluating imaging biomarkers for neurodegeneration in pre-symptomatic Huntington's disease using machine learning techniques. Neuroimage 56 (2011) 788-96.

[8] R.R. Shamir, T. Dolber, A.M. Noecker, B.L. Walter, and C.C. McIntyre, Machine Learning Approach to Optimizing Combined Stimulation and Medication Therapies for Parkinson's Disease. Brain Stimul 8 (2015) 1025-32.

[9] L.R. Trambaiolli, A.C. Lorena, F.J. Fraga, P.A. Kanda, R. Anghinah, and R. Nitrini, Improving Alzheimer's disease diagnosis with machine learning techniques. Clin EEG Neurosci 42 (2011) 160-5.

[10] V. Vigneron, A. Kodewitz, A.M. Tome, S. Lelandais, and E. Lang, Alzheimer's Disease Brain Areas: The Machine Learning Support for Blind Localization. Curr Alzheimer Res 13 (2016) 498-508.

[11] J. Wu, Y.B. Diao, M.L. Li, Y.P. Fang, and D.C. Ma, A semi-supervised learning based method: Laplacian support vector machine used in diabetes disease diagnosis. Interdiscip Sci 1 (2009) 151-5.

[12] M. Harandi, M. Salzmann, and R. Hartley, Dimensionality Reduction on SPD Manifolds: The Emergence of Geometry-Aware Methods. IEEE Trans Pattern Anal Mach Intell (2017).

[13] G. Lee, D.E. Romo Bucheli, and A. Madabhushi, Adaptive Dimensionality Reduction with Semi-Supervision (AdDReSS): Classifying Multi-Attribute Biomedical Data. PLoS One 11 (2016) e0159088.

[14] R. Pang, B.J. Lansdell, and A.L. Fairhall, Dimensionality reduction in neuroscience. Curr Biol 26 (2016) R656-60.

[15] S.E. Viswanath, P. Tiwari, G. Lee, A. Madabhushi, and I. Alzheimer's Disease Neuroimaging, Dimensionality reduction-based fusion approaches for imaging and non-imaging biomedical data: concepts, workflow, and use-cases. BMC Med Imaging 17 (2017) 2.

[16] W. Yu, S. Lee, and T. Park, A unified model based multifactor dimensionality reduction framework for detecting gene-gene interactions. Bioinformatics 32 (2016) i605-i610.

[17] K.H. Brodersen, L. Deserno, F. Schlagenhauf, Z. Lin, W.D. Penny, J.M. Buhmann, and K.E. Stephan, Dissecting psychiatric spectrum disorders by generative embedding. Neuroimage Clin 4 (2014) 98-111.

[18] K.H. Brodersen, T.M. Schofield, A.P. Leff, C.S. Ong, E.I. Lomakina, J.M. Buhmann, and K.E. Stephan, Generative embedding for model-based classification of fMRI data. PLoS Comput Biol 7 (2011) e1002079.

[19] V.C. Austin, A.M. Blamire, K.A. Allers, T. Sharp, P. Styles, P.M. Matthews, and N.R. Sibson, Confounding effects of anesthesia on functional activation in rodent brain: a study of halothane and alpha-chloralose anesthesia. Neuroimage 24 (2005) 92-100.

[20] A.G. Hudetz, General anesthesia and human brain connectivity. Brain Connect 2 (2012) 291-302.

[21] K.K. Kaisti, L. Metsahonkala, M. Teras, V. Oikonen, S. Aalto, S. Jaaskelainen, S. Hinkka, and H. Scheinin, Effects of surgical levels of propofol and sevoflurane anesthesia on cerebral blood flow in healthy subjects studied with positron emission tomography. Anesthesiology 96 (2002) 1358-70.

[22] A. Schroeter, F. Schlegel, A. Seuwen, J. Grandjean, and M. Rudin, Specificity of stimulus-evoked fMRI responses in the mouse: the influence of systemic physiological changes associated with innocuous stimulation under four different anesthetics. Neuroimage 94 (2014) 372-84.

[23] P. Boveroux, A. Vanhaudenhuyse, M.A. Bruno, Q. Noirhomme, S. Lauwick, A. Luxen, C. Degueldre, A. Plenevaux, C. Schnakers, C. Phillips, J.F. Brichant, V. Bonhomme, P. Maquet, M.D. Greicius, S. Laureys, and M. Boly, Breakdown of within- and between-network resting state functional magnetic resonance imaging connectivity during propofol-induced loss of consciousness. Anesthesiology 113 (2010) 1038-53.

[24] J. Grandjean, A. Schroeter, I. Batata, and M. Rudin, Optimization of anesthesia protocol for resting-state fMRI in mice based on differential effects of anesthetics on functional connectivity patterns. Neuroimage 102 Pt 2 (2014) 838-47.

[25] S.P. Kim, E. Hwang, J.H. Kang, S. Kim, and J.H. Choi, Changes in the thalamocortical connectivity during anesthesia-induced transitions in consciousness. Neuroreport 23 (2012) 294-8.

[26] Z. Liang, J. King, and N. Zhang, Intrinsic organization of the anesthetized brain. J Neurosci 32 (2012) 10183-91.

[27] G.A. Mashour, and M.T. Alkire, Consciousness, anesthesia, and the thalamocortical system. Anesthesiology 118 (2013) 13-5.

[28] R.N. Mhuircheartaigh, D. Rosenorn-Lanng, R. Wise, S. Jbabdi, R. Rogers, and I. Tracey, Cortical and subcortical connectivity changes during decreasing levels of consciousness in humans: a functional magnetic resonance imaging study using propofol. J Neurosci 30 (2010) 9095-102.

[29] N.S. White, and M.T. Alkire, Impaired thalamocortical connectivity in humans during general-anesthetic-induced unconsciousness. Neuroimage 19 (2003) 402-11.

[30] F. Zhao, T. Zhao, L. Zhou, Q. Wu, and X. Hu, BOLD study of stimulation-induced neural activity and resting-state connectivity in medetomidine-sedated rat. Neuroimage 39 (2008) 248-60.

[31] K. Masamoto, M. Fukuda, A. Vazquez, and S.G. Kim, Dose-dependent effect of isoflurane on neurovascular coupling in rat cerebral cortex. Eur J Neurosci 30 (2009) 242-50.

[32] J. Shawe-Taylor, and N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, 2004.

[33] G.E. Hinton, S. Osindero, and Y.-W. Teh, A fast learning algorithm for deep belief nets. Neural Comput. 18 (2006) 1527-1554.

[34] R.B.Palm, Prediction as a candidate for learning deep hierarchical models of data. Master Thesis Technical University of Denmark (2012).

[35] R. Quiza, and J.P. Davim, Computational Methods and Optimization. in: J.P. Davim, (Ed.), Machining of Hard Materials, Springer London, London, 2011, pp. 177-208.

[36] T. Koenig, L. Prichep, D. Lehmann, P.V. Sosa, E. Braeker, H. Kleinlogel, R. Isenhart, and E.R. John, Millisecond by millisecond, year by year: normative EEG microstates and developmental stages. Neuroimage 16 (2002) 41-8.

[37] D. Lehmann, P.L. Faber, S. Galderisi, W.M. Herrmann, T. Kinoshita, M. Koukkou, A. Mucci, R.D. Pascual-Marqui, N. Saito, J. Wackermann, G. Winterer, and T. Koenig, EEG microstate duration and syntax in acute, medication-naive, first-episode schizophrenia: a multi-center study. Psychiatry Res 138 (2005) 141-56.

[38] A.A. Ioannides, Dynamic functional connectivity. Curr Opin Neurobiol 17 (2007) 161-70.

[39] E.A. Allen, E. Damaraju, S.M. Plis, E.B. Erhardt, T. Eichele, and V.D. Calhoun, Tracking whole-brain connectivity dynamics in the resting state. Cereb Cortex 24 (2014) 663-76.

[40] U. Sakoglu, G.D. Pearlson, K.A. Kiehl, Y.M. Wang, A.M. Michael, and V.D. Calhoun, A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. MAGMA 23 (2010) 351-66.

[41] N. Leonardi, J. Richiardi, M. Gschwind, S. Simioni, J.M. Annoni, M. Schluep, P. Vuilleumier, and D. Van De Ville, Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest. Neuroimage 83 (2013) 937-50.

[42] N. Leonardi, W.R. Shirer, M.D. Greicius, and D. Van De Ville, Disentangling dynamic networks: Separated and joint expressions of functional connectivity patterns in time. Hum Brain Mapp 35 (2014) 5984-95.

[43] N. Leonardi, and D. Van De Ville, On spurious and real fluctuations of dynamic functional connectivity during rest. Neuroimage 104 (2015) 430-6.

[44] J. Mantilla, J. Paredes, J.J. Bellanger, E. Donal, C. Leclercq, R. Medina, and M. Garreau, Classification of LV wall motion in cardiac MRI using kernel Dictionary Learning with a parametric approach. Conf Proc IEEE Eng Med Biol Soc 2015 (2015) 7292-5.

[45] E. Varol, and C. Davatzikos, Supervised block sparse dictionary learning for simultaneous clustering and classification in computational anatomy. Med Image Comput Comput Assist Interv 17 (2014) 446-53.

[46] C.F. Beckmann, M. DeLuca, J.T. Devlin, and S.M. Smith, Investigations into resting-state connectivity using independent component analysis. Philos Trans R Soc Lond B Biol Sci 360 (2005) 1001-13.

[47] C.F. Beckmann, and S.M. Smith, Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imaging 23 (2004) 137-52.

[48] C.F. Beckmann, and S.M. Smith, Tensorial extensions of independent component analysis for multisubject FMRI analysis. Neuroimage 25 (2005) 294-311.

[49] R.-E. Fan, P.-H. Chen, and C.-J. Lin, Working Set Selection Using Second Order Information for Training Support Vector Machines. J. Mach. Learn. Res. 6 (2005) 1889-1918.

[50] V. Zerbi, J. Grandjean, M. Rudin, and N. Wenderoth, Mapping the mouse brain with rs-fMRI: An optimized pipeline for functional network identification. Neuroimage 123 (2015) 11-21.

[51] A. Liska, A. Galbusera, A.J. Schwarz, and A. Gozzi, Functional connectivity hubs of the mouse brain. Neuroimage 115 (2015) 281-91.

[52] D.V. D'Souza, E. Jonckers, A. Bruns, B. Kunnecke, M. von Kienlin, A. Van der Linden, T. Mueggler, and M. Verhoye, Preserved modular network organization in the sedated rat brain. PLoS One 9 (2014) e106156.

[53] G. Deshpande, C. Kerssens, P.S. Sebel, and X. Hu, Altered local coherence in the default mode network due to sevoflurane anesthesia. Brain Res 1318 (2010) 110-21.

[54] F. Ferrarelli, M. Massimini, S. Sarasso, A. Casali, B.A. Riedner, G. Angelini, G. Tononi, and R.A. Pearce, Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. Proc Natl Acad Sci U S A 107 (2010) 2681-6.

[55] R.M. Hutchison, M. Hutchison, K.Y. Manning, R.S. Menon, and S. Everling, Isoflurane induces dose-dependent alterations in the cortical connectivity

profiles and dynamic properties of the brain's functional architecture. Hum Brain Mapp 35 (2014) 5754-75.

[56] Z. Liang, X. Liu, and N. Zhang, Dynamic resting state functional connectivity in awake and anesthetized rodents. Neuroimage 104 (2015) 89-99.

[57] K. Masamoto, and I. Kanno, Anesthesia and the quantitative evaluation of neurovascular coupling. J Cereb Blood Flow Metab 32 (2012) 1233-47.

[58] S.J. Peltier, C. Kerssens, S.B. Hamann, P.S. Sebel, M. Byas-Smith, and X. Hu, Functional connectivity changes with concentration of sevoflurane anesthesia. Neuroreport 16 (2005) 285-8.

[59] T.L. Wu, A. Mishra, F. Wang, P.F. Yang, J.C. Gore, and L.M. Chen, Effects of isoflurane anesthesia on resting-state fMRI signals and functional connectivity within primary somatosensory cortex of monkeys. Brain Behav 6 (2016) e00591.

[60] S. Raman, L. Deserno, F. Schlagenhauf, and K.E. Stephan, A hierarchical model for integrating unsupervised generative embedding and empirical Bayes. J Neurosci Methods 269 (2016) 6-20.

**6. Random Forest Segregation of Drug Responses May define Regions of Biological Significance**

**Bukhari Q, Borsook D, Rudin M, and Becerra L**

**Abstract**

The ability to assess brain responses in unsupervised manner based on fMRI measure has remained a challenge. Here we have applied the Random Forest (RF) method to detect differences in the pharmacological MRI (phMRI) response in rats to treatment with an analgesic drug (buprenorphine) as compared to control (saline). Three groups of animals were studied: two groups treated with different doses of the opioid buprenorphine, low (LD) and high dose (HD), and one receiving saline. PhMRI responses were evaluated in 45 brain regions and RF analysis was applied to allocate rats to the individual treatment groups. RF analysis was able to identify drug effects based on differential phMRI responses in the hippocampus, amygdala, nucleus accumbens, superior colliculus and the lateral and posterior thalamus for drug vs. saline. These structures have high levels of mu opioid receptors. In addition these regions are involved in aversive signaling, which is inhibited by mu opioids. The results demonstrate that buprenorphine mediated phMRI responses comprise characteristic features that allow a supervised differentiation from placebo treated rats as well as the proper allocation to the respective drug dose group using the RF method, a method that has been successfully applied in clinical studies.

**6.1 Introduction**

Optimal dosing is an important process in the evaluation or development of pharmaceutical agents. For CNS drugs, parameters evaluated comprise pharmacokinetic readouts such as drug penetration through the blood-brain barrier [1], receptor binding, or analysis of drug concentration in cerebro-spinal fluid (CSF) [2]. Some of them rely on invasive procedures and are therefore of limited clinical use. In addition, they do not provide information on pharmacodynamic efficacy. Alternatively, drug dosing may be based on assessing pharmacodynamic responses, which for neuroactive drugs may include the analysis of effects on brain circuits using objective readouts such as functional magnetic resonance imaging (fMRI).

fMRI responses constitute an objective measure that can be used in disease diagnosis, prognosis, and evaluation of treatment effects [3]; [4]; [5]. Yet, fMRI response patterns are complex and often difficult to analyze. In recent years, machine learning and pattern recognition have entered the field of neuroimaging [6] based on their ability of detecting subtle, non-strictly localized effects, that commonly would escape univariate statistical analyses. Machine learning tools enable pattern recognition algorithms to uncover a functional relationship among the brain response patterns, in particular by identifying features that allow classification into different groups for diagnostic purposes, prognosis, or for the analysis of therapy responses [7].

Most analytical approaches concentrate on obtaining general, population-based results. While such analyses are important, methods that allow proper allocation of individual patients to the respective groups are of critical importance for diagnostic purposes. Several approaches are available to solve this problem; using support vector machines (e.g. generative embedding [8] or other machine learning techniques [9]. These techniques, however, do not discriminate the feature vectors based on their importance of classification.

Here we use Random Forest (RF) as a means of identifying brain regions that display differential responses under different pharmacological conditions (high and low doses of buprenorphine and saline), which should be suited for diagnosis at the level of the individual patient providing classification probabilities. RF is based on combining two independent ideas of random selection of features and bagging to construct decision trees with controlled variance. It has been used increasingly in medicine [10]; [11]; [12]. A big advantage of using RF is that it adds a confidence label to the classification due to its probabilistic nature. This is not the case for many other

classification algorithms including SVM: even though SVM has been repeatedly used as classification tool in neuroimaging studies, it does not provide a probabilistic classification. As a consequence SVM may add labels to a sample even if it is unable to properly classify it. In contrast, upon using RF such samples might be identified based on their probability values of 0.5 (50%), and thus be labelled correspondingly (e.g. as 'unclassified')

RF allows for estimating the importance of feature vectors that are used for its classification, thereby providing information regarding the biological basis of the classification results. As RF also generates probabilistic results, it yields a measure of confidence in the classification results obtained [13]. The methodology described in that paper is unsupervised and thus suited for analysis of experimental data, for which modeling the experimental paradigm into the analysis is difficult or not possible. This is also the case for the study presented here given the temporal fMRI response to the administration of the drug is not known. However, the method is not restricted to conditions lacking a model description since differences in brain regions are calculated from the data and depend on the power contained within the group to differentiate them.

The goal of this study was to develop a methodological formulation based on RF to identify differences in fMRI responses in a region-specific manner in groups treated with the drug buprenorphine at different doses and a control (saline) group. We have previously reported the pharmacological effects of buprenorphine vs. saline using fMRI evaluation of the drug in rats [14]. Buprenorphine is a semisynthetic opioid compound with m and k receptor affinity that has been used in the treatment of opioid-dependent patients [15] as well as for treatment of pain patients [16]. The drug was shown to have similar analgesic effects in rodents and hence, is therefore well suited for evaluating a classification scheme that would differentiate drug treatment from controls, as well as potentially discriminate different drug doses.

## 6.2 Methods

### 6.2.1 Imaging

The study was approved by the Massachusetts General Hospital's Animal Care and Use Committee. Male Sprague-Dawley rats (~300g) were used for these experiments with 12 animals injected with 0.04mg/kg (low dose; LD), 12 with 0.1mg/kg (high dose; HD) buprenorphine and 13 (controls) with saline. Solutions were prepared to

149

have a 1ml/kg concentration, for saline a 1ml/kg infusion was administered. For imaging, anesthesia was induced with 3% isoflurane for 15 minutes and the rats were positioned in the MRI cradle. A tail vain was placed for drug infusion. The infusion scan lasted 25 minutes; after 5 minutes of baseline scanning, the drug/saline was infused over a period of 2 minutes. fMRI data were acquired using a 4.7T Biospec scanner (Bruker Biospin Ltd, Billerica) with a surface coli for transmit/receive. An EPI sequence with TR/TE=2.5s/11ms was used, with 12 slices (1.5mm thick, FOV= 3.0cm, matrix=64x64) recorded. 600 volumes were acquired resulting in an acquisition time of 25min. A short TE was used to reduce susceptibility artifacts while maintaining sufficient contrast. For detailed information the reader is referred to [14].

**6.2.2 Analysis**

We have proposed a novel approach in this work to find the brain regions that differentiate between two different drug states. The pipeline we have proposed is shown in **Figure 1a** and **1b** and is discussed below. We used Random Forest to differentiate between the drug groups and further identify the most important feature components that gives us the brain regions that differ between the groups. The steps below describe each processing step that we applied over the data.

Figure 1. Flow chart of the proposed processing pipeline. (A) compares PCA, t-SNE, and isomaps to find the best suited dimensionality reduction for our experiment. The decision was taken by testing for classification and validated using LOO validation (B) goes on to apply Random Forest as the classification algorithm followed by the LOO validation.

## 6.2.2.1 *Pre-processing*

Pre-processing was carried out utilizing FSL tools [17] adapted for rat brain anatomy and included motion correction and spatial smoothing (0.7mm). No high pass filtering was applied. Brain extraction was performed using in-house software. Functional data were registered to an in-house atlas for group analysis. A Gaussian smoothing kernel for functional volume and reference volume was used.

Region of Interest (ROI) time series extraction: ROIs across the whole brain were extracted based on our internal MRI atlas developed from a histological one [18]. In total, forty-five structures were used to extract time series.

## 6.2.2.2 *Dimensionality Reduction*

Time series from predefined ROIs were extracted from the pre-processed data. In order to feed these time series into a machine-learning algorithm, it was necessary to reduce the dimension of data in an efficient way. We used isomaps as dimensionality reduction technique after having carried out a comparison analysis with Locally-Linear Embedding (LLE), t-Distributed Stochastic Neighbor Embedding (t-SNE) and PCA using DELFT implementations. The reduced time data set for a ROI was then correlated with that of other ROIs to form a z-correlation matrix. Random forest algorithm was applied to the z-correlation maps of the unlabeled data set to identify whether there was enough power contained in the dataset to classify the groups correctly. Isomaps generated the lowest false classification probabilities when used as the dimensionality reduction technique (**Figure 2**). It is important to realize that the result of dimensionality reduction is dependent on the classification result and vice versa.
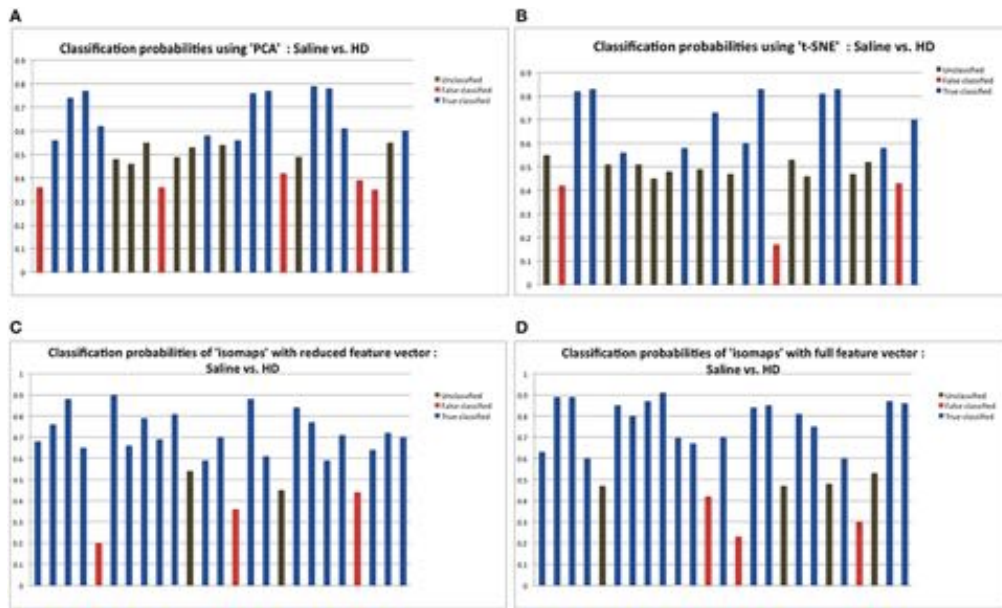
Figure 2. The probabilities of classification results. Data below 0.45 probabilities presents a false result, while anything greater than 0.55 probability presents the correct result. The classification probabilities between 0.45 and 0.55 were considered "unclassified" because of the uncertainty in classification results. Dimensionality reduction using isomaps clearly presents better results that PCA and t-SNE method as shown by the experiments. Probabilities shown here are the result of applying random forest for classification using the specific dimensionality reduction technique over the individual ROIs. (A) Probabilities showing 11 correct classification using PCA as the dimensionality reduction method for the individual ROIs. (B) Probabilities showing 10 correct classification using t-SNE. (C) Probabilities showing 20 correct classification using isomaps. (D) Comparison between full and reduction feature set for prediction: Classification probabilities and the predictions using the reduced feature set, that is only from the most important 6 features as obtained from the Random Forest variable importance graph.

*6.2.2.3 Preprocessing of the feature vectors*

After the dimensionality reduction, we applied a high pass filter to smooth the data. Since the data had a lot of undesired small fluctuations, high pass filtering gave us a better model of the time series data contained in every region. We also normalized each feature (column) to have mean 0 and standard deviation 1.

6.2.2.4 *Pearson Correlations across all regions and normalization of values.*

Pearson correlation between ROIs was carried out using the dimensionally reduced vectors to determine a full connectivity map between all the regions. Connection strengths between two ROIs are expressed by:

$$\rho X, Y = corr(X, Y)$$
$$= \frac{E[(X - \mu X)(Y - \mu Y)]}{\sigma X \sigma Y}$$

Where X=X(t) and Y=Y(t) represent the two time series corresponding to the two ROIs. The correlation values were then fitted to a Gaussian curve adopting the Fisher z-transformation with the following formula:

$$z_{X,Y} = 0.5 \cdot \log\left[(1 + \rho_{X,Y})/(1 - \rho_{X,Y})\right]$$

*6.2.2.5 Conversion of a matrix in to a vector*

The z scores from the Pearson correlation were passed through another pre-processing step. We sequentially converted the complete matrix in to a 1D vector, keeping the record of its dimensions so at any time the backtracking can result in the actual brain regions and voxel that are of interest.

*6.2.2.6 Classification of groups by Random Forest using the generated feature vectors*

Classification of the groups was done using Random Forest. RF is based on the principle of aggregating several binary decision trees built on several bootstrap samples drawn uniformly from the learning set. The aggregate of all the tree classifiers constitute the final prediction of the Random Forest. Since each tree predicts a class, a confidence interval is generated that described the percentage of votes for either class.

Feature vectors comprising the connectivity values across all brain regions were used as input for the classification. In order to evaluate the performance of the classification tool, we used the 'leave-one-out' (LOO) cross validation technique [19] [20] , thereby each and every sample was tested with regard to the classification in an unbiased fashion. LOO requires the classification algorithm to run N times, leaving one of the subjects each time out, treating it as a test subject and training the classifier on N-1 other subjects. This assures that each and every of the subject is tested and the results neither contain any bias nor any chance sampling. Prediction error and variable importance was estimated from the 'out-of-bag' sample of observations.

*6.2.2.7 Calculation of importance of feature vectors*

We used the Random Forest library from R [21] to calculate the importance of feature vectors in order to find the most important feature vectors for successful classification

of groups. This provides a variable importance index for feature vectors using RF permutation index as the indicator.

Random Forest calculates variable importance by estimating out of bag (OOB) error, which is the proportion of misclassified data. For each OOB sample we permute at random the i-th variable values of the data. The variable importance of the i-th variable is the mean increase of the error of a tree. The higher the value, the more important is the variable. The Supplementary Figure 1 shows the importance of variables graph in a decreasing order. It is clear that first few features in the graph carry more information than the rest. To keep the uniformity among the experiments, we selected 10 most important features for further processing. The other features carry similar and smaller information contents as shown in the figure. As an additional robustness check, we ran the algorithm with features from 11-990, however as can be seen from supplementary Figure 2, the prediction accuracy is decreased. This indicates that the most important information was contained in the 10 most important features as calculated by variable importance.

The most important feature vectors for LD vs. HD group were selected based on the analysis of the other two groups that are Saline vs. LD and Saline vs. HD. The uncommon correlation pairs between these two groups were selected as the features of interest.

*6.2.2.8 Re-evaluation of classification to verify the power contained in the feature vectors.*

In order to validate the feature vectors obtained through variable importance index, we classified the groups again but this time with a reduced set of features that were the top important features as selected by the R library using the variable importance function. We restricted ourselves to select 10 most important features out of 990 total features in each case. Dimensionality reduction, important feature selection and final classification were all done inside the LOO framework.

## 6.3 Results

**Figure 2** depicts the classification results as a function of the dimensionality reduction approach used: PCA, t-SNE, and Isomaps. Shown are the probabilities that the individual animals are correctly attributed to its group be it saline (control) or high-dose (HD) group of buprenorphine treatment. Using Isomaps as dimensionality reducing technique generated the highest correct classification probabilities (N=20) as compared to PCA (N=12) and t-SNE (N=11). Hence, Isomaps was used as the method of dimensionality reduction for the whole study.

Though the classification was successful, we still needed to find the most important features (regional connectivities) that made this classification possible. This is illustrated in **Figure 2(d)** depicting the prediction results of a set of selected brain regions as indicated from RF variable importance for the comparison control versus LD. **Figure 2(c)** should be compared with **Figure 2(d)**, which shows the analogous analysis for ROIs across the whole brain. The results indicate that using specific but more informative regions preserves the classification result, and thus proves the concept that these regions contain most of the useful information for the classification between the two groups. Classification accuracy was evaluated using the LOO method (**Table 1**)

|  | Average correctly predicted in k times leave one out validation | Average false predicted in k times leave one out validation | Average unclassified in k times leave one out validation | Total number of subjects "k" |
|---|---|---|---|---|
| Saline vs. LD | 18 | 4 | 3 | 25 |
| Saline vs. HD | 18 | 3 | 4 | 25 |
| LD vs. HD | 6 | 9 | 9 | 24 |

Table 1. Classification accuracy based on leave one out cross validation with all 45 regions (990 features) considered for the classification.

|  | Average correctly predicted in k times leave one out validation | Average false predicted in k times leave one out validation | Average unclassified in k times leave one out validation | Total number of subjects "k" |
|---|---|---|---|---|
| Saline vs. LD | 20 | 2 | 3 | 25 |
| Saline vs. HD | 20 | 3 | 2 | 25 |
| LD vs. HD | 16 | 6 | 2 | 24 |

Table 2. Classification accuracy based on leave one out cross validation after selecting the top 10 features from the variable importance as indicated by Random Forest.

Similar analyses have been carried out for the HD group. Classification was first applied with the complete feature set (990 features), followed by the calculation of

important features. These important features were then used for re-classification. The accuracy of the classification procedure was evaluated using the LOO method. Reducing the number of feature vectors to include the 10 most important ones preserves the classification accuracy, proving that the most important information lies in the selected feature vectors (**Table 2**). When comparing LD versus HD, the initial classification using all 990 features with leave one out validation generated only chance probability. Thus, the lack of significant result also prohibited us from further continuing the analysis to find the most important features for classification. To solve this problem we used the mutually exclusive method from sets, i.e. we selected the anatomical regions which were found among the most important features of Saline vs. LD and Saline vs. HD comparisons, however selected only those anatomical regions present in one of the two comparisons only. The rationale behind was if it exists in only one of the comparisons, it is more likely to be the effect of the dose rather than the saline or other mutual effects in the comparison. Once these uncommon correlation pairs between these two groups were selected as features of interest, we applied the classification algorithm over the reduced feature set as selected from this method, and applied LOO cross-validation to obtain classification accuracy of 66.6%. While this work-around yielded some reasonable classification results, the results need to be handled with care.

**Table 3** indicates the brain structures that anchor the classification using the reduced set of features. Common structures that discriminate fMRI response of the three treatment groups included thalamus, hypothalamus, hippocampus, caudate putamen and colliculus. Only the 10 most important features in the classification are listed, while few extra regions are also listed with their rank among importance of feature vectors, to provide better comparison between Saline vs. LD and Saline vs. HD analysis.

| Cortical structures | Regions | Sides | Saline vs. LD | Saline vs. HD | LD vs. HD |
|---|---|---|---|---|---|
| Sensorimotor cortex |  | Right | 2.76 | 2.26 | |
| | | Left | <2 | <2 | |
| Anterior cingulate cortex |  | | 2.21 | 2.13 | * |
| Entorhinal cortex |  | Right | 2.03 | 2.21 | |
| | | Left | <2 | <2 | |
| Insula |  | Right | 2.09 | <2 | |
| | | Left | 2.05 | <2 | * |
| Hippocampus |  | Right | 2.21 | <2 | |
| | | Left | <2 | 2.91 | |

| Subcortical structures | Regions | Sides | Saline vs. LD | Saline vs. HD | LD vs. HD |
|---|---|---|---|---|---|
| Thalamus ventral |  | Right | 5.3 | 3.87 | |
| | | Left | 4.6 | 3.23 | |
| Thalamus posterior |  | Right | <2 | 3.57 | |
| | | Left | <2 | 2.09 | * |
| Hypothalamus |  | Right | 2.9 | 3.87 | |
| | | Left | 1.95 | 2.41 | |
| Cudate Putamen |  | Right | 2.09 | 2.96 | |
| | | Left | <2 | 2.13 | * |
| Amygdala basal laterai |  | Right | 2.87 | 2.34 | |
| | | Left | 2.86 | 2.23 | |
| Amygdala anterior |  | Right | 2.07 | 2.27 | |
| | | Left | <2 | <2 | |
| Nucleus accumbens |  | Right | <2 | 2.26 | * |
| | | Left | 2.86 | 3.57 | |
| Superior colliculus |  | Right | <2 | 2.53 | * |
| | | Left | <2 | 2.53 | * |
| Inferior colliculus |  | Right | <2 | 2.97 | * |
| | | Left | <2 | 2.97 | * |

The table below shows the cortical and subcortical structures with the individual importance value for each classification.
For Saline vs. LD and Saline vs. HD, only those values are shown that are greater than 2. For LD vs. HD, the regions that were found important for the classification are marked with asterisk (*).

Table 3. Anatomical structures found important for the classification.

**6.4 Discussion**

While classification using machine learning approaches have been used for pain states on the basis of fMRI data, the approach has been hardly applied for evaluating drug efficacy [22]. Here, we have used RF for identifying brain regions displaying differential responses in response to treatment with the opioid drug buprenorphine and saline in a supervised manner. A critical step in preprocessing data for the machine learning tool is dimensionality reduction based on the experimental data. We have found that isomaps yields better classification accuracy compared to PCA or t-SNE based reduction methods. A feature of Isomaps is that it considers voxels in their context by intrinsic construction of a neighborhood graph based on the geodesic distance. Such neighborhood reletionships are important factors when analyzing brain functional data, as local networks of connected voxels make up a functional region. This might explain why isomaps outperformed other classifiers in our case.

For the classification, we choose the RF classifier, since based on theoretical arguments it should yield optimal classification performance in the sample limit, which has been successfully demonstrated to be the case for many biomedical classification problems [23]. The explanation of the good performance of RF is related to the good quality of each regression tree [24]. One of the big advantages that RF offers is that it automatically saves the features that are most critical for the classification purpose: these features are related to a relatively small number of anatomical regions that apparently play an important role in capturing treatment response. It is to our knowledge the first time that such method has been applied for analysis of fMRI response to pharmacological stimulation.

Brain region displaying a differential response in buprenorphine treated rats at either dose compared to saline treated controls are listed in **Table 2**. Regions identified for both drug doses upon comparison with saline (HD vs. Saline; LD vs. Saline) have been reported to process pain [25]. Additional regions were identified in both groups such as the caudate putamen, superior and inferior colliculus. Interestingly, all regions identified for the LD buprenorphine group were also found in the HD group, though not necessarily in the same rank order. Regions that show differential responses depending on the buprenorphine dose were amygdala, hypothalamus, nucleus accumbens, posterior thalamus, and sensorimotor, insular and entorhinal cortices. Many of these areas appear as classifiers either in HD vs. saline or LD vs. Saline. The relevance of the classification results is also supported by the notion that essentially all

the structures important for classification display high levels of mu opioid receptors. In addition they are involved in aversive processing, which might be inhibited by opioids. In [14], cingulate cortex, insula, cerebellum and thalamus had been shown as important anatomical regions that differentiate between LD vs. HD group using model based analysis (GLM) and our data largely agree with their findings. However in our results, insula appears in LD vs. Saline and LD vs. HD comparisons but not in HD vs. Saline comparison. Amygdala was not reported in the GLM paper, however it appears as an important region in our results. Model based approaches are limited by the intrinsic nature of the model: i.e. features not comprised in the model cannot be extracted from experimental data. However model-free approach such as the RF classification is not based on the correspondence of the basic assumptions with the actual experimental data. Being data driven it just searches for differences in the responses among two (or more groups) irrespective of their actual shape, and thus might identify regions that are not detected with GLM or related methods.

Brain regions of high discriminative power for all three comparisons (LD vs. saline, HD, versus saline, LD vs HD) were hippocampus and thalamus, while insula, posterior thalamus, amygdala differentiated the LD from the HD group. What may be important in the functionality of these regions that contribute to the differentiation process? There may be a number of interrelated processes including (1) opioid receptor numbers, (2) function of specific regions in endogenous pain control, and (3) connectivity between these regions contributing to whole brain 'interrogation' of the three conditions.

Buprenorphine is known to bind to the m and k opioid receptors [14], the nociceptin/orphanin receptor [26], and the opioid receptor like (ORL-1) receptor [27]. The m-opiod receptors occur with high concentration in cerebral cortex, thalamus, striatum (striosomes), amygdala, periaqueductal gray [28], while the k-opioid receptors are found in hypothalamus and also periaqueductal gray. In general the distribution profiles of the primary opioid receptors – m, d, k - are slightly different but also share significant overlap in structures including the amygdala and hippocampus. High levels of nocicpetin receptor occur in cortex, hippocampus (dentate gyrus), amygdala, hypothalamus and septal nuclei [29], while ORL-1 receptor is found predominantly in cortical areas, olfactory regions, limbic structures, and thalamus [30]. The regions identified in the RF signature discriminating drug effects from vehicles are essentially those outlined by this receptor distribution (Table 2).

While it is unclear how the specific regions contribute as a result of their own primary function (e.g., anterior insula and awareness) or interactions as a result of buprenorphine-receptor activation of neurons with efferent projections, the drug must produce alterations in brain circuits that differentiate the three conditions. The amygdala is involved in analgesia, emotion and also decision-making [31]. Of the various structures identified, amygdala has probably has the highest binding of all three receptor targets of buprenorphine and it is therefore not surprising that it shows a strong response to the drug as compared to saline. The hippocampus known to be involved in memory formation, spatial orientation and pain modulation, also participates in the stress response [32]. This latter response can be diminished by opioids, presumably due to its m receptor effects [33], which may explain its prominent role as classifier in discriminating the response to drug treatment as compared to controls. Nevertheless, it should be remembered that buprenorphine interacts with several opioid and opioid-like receptors systems, the effects of which might be even counteracting (e.g. ORL-1 activity versus m receptor effects). Hence interpreting the occurrence of specific brain areas in the discriminative feature vectors in terms of cognitive and emotional effects remains speculative, in particular when dealing with anesthetized animals as anesthesia may further modulate the fMRI responses to drug administration. On the other hand it is reassuring that regions associated with nociceptive processing clearly show up, indicative of the analgesic activity of the drug.

While RF yielded reasonable classification results regarding the nature of structures identified, we observed that several of the important regions appear unilaterally only. Given the nature of the condition, i.e. a pharmacological stimulus with a systemically administered drug, and the more or less symmetrical distribution of its molecular targets across the brain, on would have expected that feature vectors display bilaterally symmetry. The question then arises whether this left/right asymmetry is of biological origin or whether it is an artifact of the analysis. Finding the most relevant features involves inherently a ranking and when maintaining of fixed number of 'most important feature vectors' also a thresholding. As a consequence, it is well conceivable that a structure within one hemisphere may not pass this threshold, leading to an apparent laterality. Obviously, this may be accounted for by relaxing the threshold criterion, i.e. by maintaining bilaterality if the difference between the two hemispheres is within 'the noise range'. This also became obvious, when analyzing the dose dependence of the buprenorphine response. When limiting the number of feature

vectors in the analysis to 10, there was the counter-intuitive result that features discriminating drug from saline appeared in the LD but not in the HD group. However, all LD features were contained in the 17 most important features of the HD group indicating that when analyzing the data it is important not to just be restricted to a fixed number of features.

The reasons underlying the failure of typical classification between LD vs. HD can be explained in several ways. LD and HD doses were selected based on publications indicating minimal and significant analgesic effects on behavioral measures in rats. Discrepancies between behavioral outcome and imaging findings are not unusual, as the former depend on the specific test paradigm applied and may be confounded by processes such as learned behavior or reflexive responses. In contrast imaging responses depend on the physiological baseline state, which is affected by the use of anesthesia. Other factors that might explain the lack of a difference between the LD and HD group might arise from inter-individual differences in the bioavailability of the drug, which would result in larger variability reducing the statistical power in discriminating the two states. . Finally there might be a ceiling effect regarding the fMRI response. Future studies using expanded dose ranges and larger cohorts should clarify this aspect.


**6.5 Conclusion**

We have used RF to classify rats based on the fMRI signature in response to systemic administration of buprenorphine at to different doses or saline to rats. The regions that turned out to be most important for the proper classification of animals were those displaying high levels of opioid and opioid-like receptors known to be the buprenorphine target, in particular structures associated with nociceptive processing, but also the limbic system. Using the LOO approach, the classification accuracy was 80% for the comparison of drug versus placebo, while there were 66 % of correct assignments for the comparison LD versus HD. RF appears an attractive machine learning tool suited for classification of individuals based on their response to a neuroactive compound.

**6.6 Supplementary material**
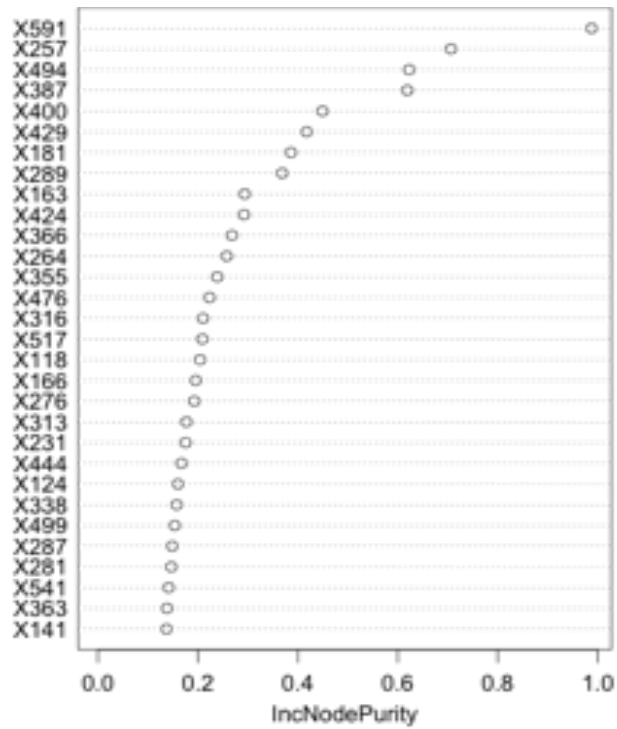


**Supplementary Figure 1.** The figure shows the variable importance map with the top features representing more information than the rest.
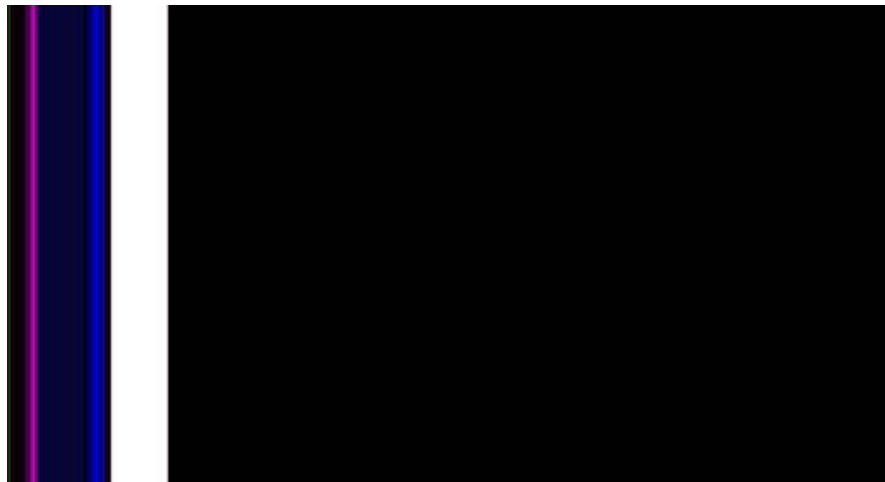


**Supplementary Figure 2.** The figure shows the classification results for Saline vs. HD with features 11 to 990 in the decreasing order of their importance values. The classification results with LOO show that the prediction accuracy with top 10 important features is greater than prediction accuracy with features 11-990

## 6.7 References

[1] M.S. Alavijeh, M. Chishty, M.Z. Qaiser, and A.M. Palmer, Drug metabolism and pharmacokinetics, the blood-brain barrier, and central nervous system drug discovery. NeuroRx 2 (2005) 554-71.

[2] M. Friden, H. Ljungqvist, B. Middleton, U. Bredberg, and M. Hammarlund-Udenaes, Improved measurement of drug exposure in the brain using drug-specific correction for residual blood. J Cereb Blood Flow Metab 30 (2010) 150-61.

[3] D. Borsook, J. Upadhyay, M. Klimas, A.J. Schwarz, A. Coimbra, R. Baumgartner, E. George, W.Z. Potter, T. Large, D. Bleakman, J. Evelhoch, S. Iyengar, L. Becerra, and R.J. Hargreaves, Decision-making using fMRI in clinical drug development: revisiting NK-1 receptor antagonists for pain. Drug Discov Today 17 (2012) 964-73.

[4] D. Borsook, R. Hargreaves, and L. Becerra, Can Functional Magnetic Resonance Imaging Improve Success Rates in CNS Drug Discovery? Expert Opin Drug Discov 6 (2011) 597-617.

[5] D. Borsook, L. Becerra, and R. Hargreaves, Biomarkers for chronic pain and analgesia. Part 2: how, where, and what to look for using functional imaging. Discov Med 11 (2011) 209-19.

[6] E. Formisano, F. De Martino, and G. Valente, Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning. Magn Reson Imaging 26 (2008) 921-34.

[7] K. Strimbu, and J.A. Tavel, What are biomarkers? Curr Opin HIV AIDS 5 (2010) 463-6.

[8] K.H. Brodersen, L. Deserno, F. Schlagenhauf, Z. Lin, W.D. Penny, J.M. Buhmann, and K.E. Stephan, Dissecting psychiatric spectrum disorders by generative embedding. Neuroimage Clin 4 (2014) 98-111.

[9] J. Schrouff, M.J. Rosa, J.M. Rondina, A.F. Marquand, C. Chu, J. Ashburner, C. Phillips, J. Richiardi, and J. Mourao-Miranda, PRoNTo: pattern recognition for neuroimaging toolbox. Neuroinformatics 11 (2013) 319-37.

[10] S. Ghose, A. Oliver, J. Mitra, R. Marti, X. Llado, J. Freixenet, D. Sidibe, J.C. Vilanova, J. Comet, and F. Meriaudeau, A supervised learning framework of statistical shape and probability priors for automatic prostate segmentation in ultrasound images. Med Image Anal 17 (2013) 587-600.

[11] R. Casanova, S. Saldana, E.Y. Chew, R.P. Danis, C.M. Greven, and W.T. Ambrosius, Application of random forests methods to diabetic retinopathy classification analyses. PLoS One 9 (2014) e98587.

[12] A.H. Simonsen, J. Mattila, A.M. Hejl, E. Garde, M. van Gils, C. Thomsen, J. Lotjonen, H. Soininen, and G. Waldemar, Application of the PredictAD decision support tool to a Danish cohort of patients with Alzheimer's disease and other dementias. Dement Geriatr Cogn Disord 37 (2014) 207-13.

[13] F. Disanto, and T. Wiehe, Exact enumeration of cherries and pitchforks in ranked trees under the coalescent model. Math Biosci 242 (2013) 195-200.

[14] L. Becerra, J. Upadhyay, P.C. Chang, J. Bishop, J. Anderson, R. Baumgartner, A.J. Schwarz, A. Coimbra, D. Wallin, L. Nutile, E. George, G. Maier, S. Sunkaraneni, S. Iyengar, J.L. Evelhoch, D. Bleakman, R. Hargreaves, and D. Borsook, Parallel buprenorphine phMRI responses in conscious rodents and healthy human subjects. The Journal of pharmacology and experimental therapeutics 345 (2013) 41-51.

[15] R.P. Mattick, C. Breen, J. Kimber, and M. Davoli, Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. The Cochrane database of systematic reviews 2 (2014) CD002207.

[16] J. Cote, and L. Montgomery, Sublingual buprenorphine as an analgesic in chronic pain: a systematic review. Pain Med 15 (2014) 1171-8.

[17] M. Jenkinson, C.F. Beckmann, T.E. Behrens, M.W. Woolrich, and S.M. Smith, Fsl. Neuroimage 62 (2012) 782-90.

[18] Paxinos, The Rat Brain in Stereotaxic Coordinates, Academic Press 2007.

[19] P.A.L.a.M.R. Mickey, Estimation of error rates in discriminant analysis. Technometrics 10 (1968) 1-12.

[20] A.L.a.V. Brailovsky, On estimation of characters obtained in statistical procedure of recognition (in Russian). Techicheskaya Kibernetica 3 (1969).

[21] A.L.a.M. Wiener, Classification and Regression by Random Forest. R News 2 (2002) 18-22.

[22] R. Salat, and K. Salat, The application of support vector regression for prediction of the antiallodynic effect of drug combinations in the mouse model of streptozocin-induced diabetic neuropathy. Comput Methods Programs Biomed 111 (2013) 330-7.

[23] A.S.a.C.F. Aliferis, Are Random Forests Better than Support Vector Machines for Microarray-Based Cancer Classification? AMIA Annu Symp Proc. 2007 (2007) 686-690.

[24] Breiman, Random Forests. Machine Learning 45 (2001) 5-31.

[25] I. Tracey, Imaging Pain. British journal of anaesthesia 101 (2008) 32-9.

[26] T.V. Khroyan, J. Wu, W.E. Polgar, G. Cami-Kobeci, N. Fotaki, S.M. Husbands, and L. Toll, BU08073 a buprenorphine analogue with partial agonist activity at mu-receptors in vitro but long-lasting opioid antagonist activity in vivo in mice. Br J Pharmacol 172 (2015) 668-80.

[27] K. Lutfy, and A. Cowan, Buprenorphine: a unique drug with complex pharmacology. Curr Neuropharmacol 2 (2004) 395-402.

[28] A. Mansour, H. Khachaturian, M.E. Lewis, H. Akil, and S.J. Watson, Autoradiographic differentiation of mu, delta, and kappa opioid receptors in the rat forebrain and midbrain. J Neurosci 7 (1987) 2445-64.

[29] T. Houtani, M. Nishi, H. Takeshima, K. Sato, S. Sakuma, S. Kakimoto, T. Ueyama, T. Noda, and T. Sugimoto, Distribution of nociceptin/orphanin FQ precursor protein and receptor in brain and spinal cord: a study using in situ hybridization and X-gal histochemistry in receptor-deficient mice. J Comp Neurol 424 (2000) 489-508.

[30] C. Mollereau, and L. Mouledous, Tissue distribution of the opioid receptor-like (ORL1) receptor. Peptides 21 (2000) 907-17.

[31] J. LeDoux, The amygdala. Curr Biol 17 (2007) R868-74.

[32] J.D. Gray, T.G. Rubin, R.G. Hunter, and B.S. McEwen, Hippocampal gene expression changes underlying stress sensitization and recovery. Mol Psychiatry 19 (2014) 1171-8.

[33] R. Okutani, K. Kono, O. Kinoshita, H. Nakamura, H. Ishida, and D.M. Philbin, Variations in hemodynamic and stress hormonal responses in open heart surgery with buprenorphine/diazepam anesthesia. J Cardiothorac Anesth 3 (1989) 401-6.

## 7. Discussion

Resting state fMRI provides a method for studying whole brain functional connectivity and large-scale brain networks without the need of any explicit task condition [1; 2; 3]. This is of great value in small animal imaging, where task based paradigms or stimulus-evoked paradigms have been challenging due to their unspecific response reported in mice, and limited in their application because of the small choice of possible tasks/stimuli (e.g. electrical stimulation, heat stimulation etc.) available in anesthetised rodents. It is therefore not surprising that the number of rs-fMRI studies in rodents has rapidly increased over the last years.

fMRI analysis tools are typically developed for human. Due to inter-species differences, increased susceptibility to magnetic inhomogeneity at large magnetic field strengths leading to image distortion and intra-voxel dephasing (signal voids), as well as differences in basic physiology (respiration frequency, heart rate, blood pressure), the application of these analysis tools to analyse small animal fMRI data is not straightforward and requires adaptations. Similarly, machine-learning tools have been largely restricted to application in human fMRI data.

A major difference between human and animal (rodent) fMRI is the use of anaesthesia, which is common in small animal fMRI studies despite their known intrinsic effects on brain activity, but rarely used in human studies, unless the purpose of the study is to investigate the effects of an anaesthetic drug on brain function/networks. In animal studies, anaesthesia is essential for immobilizing the animal during data acquisition. Anaesthesia will inevitably affect the results of functional brain studies; it is therefore important to assess and understand the alterations in brain activity patterns/network induced by the anaesthetic drug. Several groups have studied effects of anaesthetics on rodent fMRI data and not surprisingly reported differential results based on the agent used [4; 5; 6; 7; 8; 9; 10; 11; 12]. Optimized anaesthesia regimes for rodent fMRI studies have been suggested [13], yet this likely depends on the question to be addressed. A major issue when analysing the effects of anaesthesia is the lack of a valid reference state. Comparison to data obtained in the awake state is feasible in humans, but difficult in rats and mice as even in trained animals confounds due to stress and arousal responses may not be excluded due to the immobilization stress and scanning noise [14]. Even in mice under light anaesthesia, stimulus evoked fMRI responses were found severely confounded by

arousal related systemic hemodynamic responses [15]. The questions remains, how to assess the effect of an anaesthetic on brain activity/brain networks. One approach might be to translate results obtained in awake and anaesthetized humans to other mammals, for example the use of isoflurane at a specific dose predominantly affects brain area X or the functional connection X-Y. Yet, this would not account for differences across species; in addition, many of the anaesthetics used in rodents are not used in a clinical setting. An alternative strategy might be to relate functional changes as derived from MRI to clinical, behavioural measures of anaesthesia depth such as loss of sensation, analgesia, muscle relaxation, and loss of consciousness, as a function of anaesthesia depth (dose of the anaesthetic drug).

A prerequisite for such studies are reliable tools for analysing fMRI, and in particular resting-state fMRI data, that yield semi-quantitative information on brain networks and changes in interactions strength when types or dose of the anaesthetic agent is altered. We therefore evaluated several tools originally developed for the analysis of human fMRI data: dual regression and network analysis for identifying major brain network modules and studying within and between network interactions [16], dynamic functional connectivity analysis [17; 18] for probing for hidden information on network interactions not apparent from conventional pseudo-stationary resting-state fMRI analysis, and finally the use of machine learning tools in classifying resting-sate fMRI data obtained for different anaesthesia regimes.


## 7.1 Effects on anaesthetics on mouse functional networks und their interaction

In a first study we analysed resting-state fMRI patterns for different anaesthetic regimens including isoflurane, medetomidine and the combination of the two, and studied their effects on brain networks using multivariate methods and network modelling. Clearly the tools were able to identify differential effects of the respective agents on the brain activity patterns. Mice under low dose isoflurane anaesthesia displayed predominantly intra- and inter-cortical interactions with only minor interactions involving subcortical structures. In particular cortico-thalamic connectivity was found attenuated. In contrast, medetomidine-anesthetized mice displayed significant subcortical functional connectivity including interactions between cortical and thalamic ICA components, while intracortical networks appeared less prominent. Combining the two anaesthetics at low dose resulted in network

interactions that constituted the superposition of the interactions observed for each anaesthetic alone, though we acknowledge that some interactions were better preserved under isoflurane or medetomidine alone rather than under the combination anesthesia. As an attractive feature of the isoflurane-meditomidine combination anaesthesia showed minimum effects on systemic hemodynamic parameters and appears therefore suitable for longitudinal studies. There are mixed reports in literature regarding the effects of anaesthetics on cortico-thalamic interactions [19; 20; 21]. Our data indicate that cortico-thalamic interaction appears to be modulated by the type and depth of anesthesia, since we found loss of cortico-thalamic activity in isoflurane anesthetized mice but an anticorrelation pattern in mice under the effect of medetomidine and iso-med combination anesthetic regimens.

In a second study we evaluated the dependence of mouse resting-state fMRI patterns on the dose of isoflurane using pseudo-stationary and dynamic functional connectivity analysis and found that at higher isoflurane levels spatial segregation among the brain regions is lost. Also the functional connectivity between homotopic regions was found to be lost at higher dose of isoflurane.

The optimal choice of anesthesia, the specific regimen as well as the dose, depends on the specific brain regions and the specific problem to be studied, in order to minimize interference with the processes to be studied. Furthermore, the insight on the effects of anesthetics on large scale brain networks may help in optimization of anesthesia protocols.


## 7.2 Application of resting-state fMRI analysis tools to mouse data

Analysis tools for neuroimaging data are typically developed for use in humans. In this work we tested the applicability of several state of the arts analysis tools for rodent fMRI studies. We also developed machine-learning tools to identify brain regions affected by different doses of drugs.


When analysing (mouse) whole brain fMRI data for changes induced by neuroactive agents, methods free of operator bias are in general preferred. In this sense, we considered ICA superior to a seed based approach. We also adapted a dual regression approach for the analysis of rodent resting-state fMRI data. The combination of dual regression with ICA analysis provides an alternative to typical resting analysis techniques, minimizing operator's bias. We used partial correlation estimates for the

network analysis in order to eliminate spurious edge effects due to an indirect third region in-between the two regions. We also performed dFC analysis to determine the dose dependent effects of isoflurane. dFC analysis revealed significant interactions among functional networks that were not apparent from the conventional stationary FC analysis. We also applied machine learning algorithms to analyse rodent fMRI data. Our random forest pipeline showed that machine learning algorithms can be used in rodent fMRI studies for identifying the brain regions affected by the different doses of drugs. Preliminary results using support vector machine (SVM) analysis and deep learning analysis indicated huge potential for future use in rodent fMRI studies. In the following sections some of these aspects are discussed in more detail.

## 7.3 Use of dual regression, network modelling and dynamic functional connectivity in small animal fMRI

Dual regression and network modelling approaches have been mostly applied for analysing human data so far, with the exception of a few reports of applying network analysis methods in rats [22; 23]. Dual regression provides an alternative to the GLM analysis based on seed based method for analysing resting state fMRI data. In the absence of an experimental paradigm, dual regression provides an unbiased and more sophisticated approach to find activations in the brain regions associated with the effects of interest. Despite that fact that the mouse brain and in particular its cortical organization is simpler than human brain and is subject to less inter-individual variability, the applicability and the value of these tools to/for small animal imaging remains questionable given the biological differences across species and the intrinsic differences in raw data quality (SNR values constitute a major challenge in rodent fMRI). SNR values reported in mice are below 100 [24] while in humans it may go up to more than 250 [25]. Nevertheless the reported SNR values in mice are without cryogenic coils and we have observed an increase in SNR values up to 150 with the use of cryogenic coils.

In this work, we applied Dual Regression and Network Modeling for identifying brain network differences between different anaesthetics regimens and doses. Dual Regression identified the loss of connectivity between homotopic brain regions for higher doses of anaesthetics. Furthermore different anaesthetic regimens were found to induce different modulatory effects on functional connectivity, which was best studied

by determining the fMRI responses at the level of brain networks. We also applied dynamic functional connectivity to determine the effects that might have been averaged out in the stationary FC analysis. dFC revealed interesting connectivity maps showing loss of modular structure of functional connectivity at higher dose of anaesthesia. Dual regression, network modelling and dynamic functional connectivity revealed reasonable, neurobiologically plausible results for mouse rs-fMRI data. Our reported results are statistically significant and in line with our previous findings [13; 26]. The results were also found to be consistent with other reports in human studies.

A critical aspect is the intrinsic quality of the raw data, which should display both an intrinsic high image SNR and minimize contributions to physiological noise due to physiological instability. Combining cryogenic MRI detection [27] with careful mouse preparation [13; 15; 28] involving intubation, mechanical ventilation in combination with muscle relaxant, constituted critical success factors in this regard.

### 7.1.3 Machine learning in small animal fMRI

Machine learning in neuroimaging studies is gaining wide spread recognition. However, up to now such analyses are largely limited to their application in human fMRI. We successfully applied machine-learning algorithms to mouse fMRI data. In human fMRI, prediction of group assignment (labels) can be useful as it can be used to diagnose certain neurological disorders; brain areas that contributed the most to the classification are of relevance as it extends our biological understanding of which areas have been affected by the certain disorder or drug. On the other hand, in a typical animal experiment, we know the history of the animal, i.e. we know to what group it belongs. However, animal studies should be carried in a blinded fashion, i.e. the person carrying out the analysis should not be aware of the treatment. In this sense, the use of automated classification tools might constitute an unbiased approach for analysing treatment effects.

To explore the use of machine learning in classifying animals to specific treatment groups and in identifying features critical for the differentiation, we used Random Forest algorithm to highlight brain regions that played the most important role for classifying rats treated with vehicle or the analgesic buprenorphine at two different doses. Classification on the basis of the 10 most important features yielded a 20% error rate when classifying vehicle versus drug-treated rats, while the error rate

increased to 40% when comparing the low and high dose buprenorphine group. Reduction of the number of feature vectors to the most important once increased the quality of the classification. Nevertheless, there remained a substantial error rate, which eventually might be linked to the quality of raw data available. The identification of the brain regions through feature importance maps contributed to the understanding of buprenorphine effects on the brain.

In a second study, we have used SVM and DBN, a deep learning algorithm, to allocate mice anaesthetized with isoflurane to the appropriate dose group. The study was designed for evaluating the potential of deep learning algorithms in neuroimaging studies, and the results are promising to be used in future for both human and animal imaging studies. The results show that we were able to successfully classify between anaesthetic doses using features extracted from static and dynamic functional connectivity analysis. The features extracted from dynamic functional connectivity analysis were found to be more discriminative to different anaesthetic doses. This shows features based on dynamic functional connectivity analysis are capable of identifying subtle changes in mice brains using rs fMRI data. A major limitation regarding the use of machine learning in the context of our study was small sample size (N=12 per group), which led to an accuracy of less than 70% for most comparisons.

## 7.4 Limitations of animal imaging

Animals imaging has gained popularity over the years due to their low inter population differences, ability to conduct tests on genetically modified mice, as models of human diseases and for evaluating novel therapies. Animal studies also provide valuable insights into basic mechanisms because of the high structural similarity across mammalian brains, similar cell types, same neurotransmitters etc. With regard to functional neuroimaging study the simple topology of animal brain is attractive in this context, and results obtained might eventually be translated to the more complex human architecture.

While technical limitations (SNR etc) have already been discussed and will not be repeated here, there are also conceptual issues regarding fMRI studies in rodents.

1) While simplicity of the brain organization is an advantage in some cases, it is certainly a disadvantage (complexity is not for nothing). There may be human networks not found in the mouse (example is discussion on DMN)

2) Despite high level of similarity across mammalian brains there are species differences at every level (molecules, cells, structures) which may render translation nevertheless difficult.

3) The use of anaesthesia limits our options to study brain functions related to the cognitive tasks and emotional tasks etc.

4) Use of anaesthetics may alter basic physiology, which may be an issue as we are measuring a physiological readout

5) Use of anaesthesia is found to decrease the pulse distention and blood pressure. Since fMRI is an indirect measure of cerebral activation, which depends on the cerebrovascular baseline state, changes in blood pressure will affect the amplitude of the BOLD signal fluctuations and thereby potentially also the correlation analysis

## 7.5 Outlook

We will discuss two aspects in the outlook section: 1. How could we derive markers of a state of unconsciousness based on neuroimaging information, and 2. How can we use the methods described, in particular the classification techniques, in personalized patient care, for example in predicting therapy outcome in patients suffering from psychiatric disorders.

### 7.5.1 EEG-fMRI to identify neuromarkers of unconsciousness in the action of anaesthetics

Anaesthesia is an integral part of essentially all mouse fMRI studies. Though there are a few reports of awake rodent studies, their applicability in mice is certainly limited due to the difficulty in training mice. Furthermore, awake rodent studies have also been criticised for inducing the bias of stress in the animals due to immobilization (head fixation) and the high noise level in the scanner. Therefore fMRI studies in mice typically use anaesthetics in the experiments, despite the fact that anaesthetics are known to alter the functional connectivity. A major observation in our studies was loss

of spatial segregation of brain functional modules when increasing the dose of the isoflurane. However these findings are based on hemodynamic readouts and alterations in cardiac output or efficiency of neurovascular coupling might thus influence the results. Hence, confirmation using a technique that measures neural activity more intimately is desirable. Effects of anaesthesia have been widely studied using EEG because of its high temporal resolution. However EEG has an intrinsic low spatial resolution that limits its usability in studying whole brain connectivity. Hybrid EEG-fMRI setups combine the advantages of both systems and we believe that anaesthesia can be well studied with such a system. There have been a few attempts in this direction [30], however so far none of the studies using such system aimed to identify the tipping point of unconsciousness. Furthermore these findings should also be replicated in awake rodent studies. We have previously discussed that different anaesthetic agents affect different brain areas/networks. On the other hand there must be some similarity as most of them lead to the same clinical/physiological result: analgesia, muscle relaxation, loss of consciousness. Hence, by comparing the effect of different anaesthetic drugs on brain function at different dose levels, it might be possible to tease out domains / features that are ultimately associated with the physiological/ clinical readouts of anaesthesia. These features might serve as marker regions to optimize anaesthesia regimen. Also comparing activity patterns of agents known to be merely sedative with agents known to lead to a complete loss of consciousness may help identifying responsible brain areas

### 7.5.2 Optogenetics to control consciousness/unconsciousness

We have identified connectivity between several important brain regions that play a vital role in unconsciousness e.g. the cortico-thalamic interactions. However in order to identify their particular role in unconsciousness, these brain connectivities may be artificially altered to determine if it changes their conscious condition. Optogenetics is a breakthrough technique developed recently to artificially control genetically modified light sensitive neurons in the brain by using light. Neurons that are present in the anaesthesia-affected regions according to our results may be genetically modified to become light sensitive and then their effects on the consciousness can be determined. If consciousness can be controlled this way, such a study would become a major step in understanding the effects of anaesthesia and altered states of consciousness.

### 7.5.3 Psychiatric treatment outcome predictions

Machine learning in neuroimaging is expected to play an important role in the coming years. Recently researchers were able to reconstruct the content of the dream by analysing the brain activity during sleep and matching it to the already known response to the visual stimuli [31]. In clinical psychology, researchers have used machine learning to predict the intrusive moments in PTSD patients [32]. There have been several other reports of using machine learning for identifying noise related components from the ICs and then regressing those components to improve the signal to noise ratio [33; 34; 35].

However use of machine learning in clinical neuroimaging has largely been confined to the prediction of diagnosis at the single subject level [36; 37; 38; 39; 40]. Psychiatrists on the other hand, trust their experience and diagnosis manuals for the diagnosis. This led to limited to no use of fMRI data analysis techniques for identification of psychiatric disorders in clinical practice. However an area where the psychiatrists require assistance is in selection of alternative treatment options. The psychiatrist usually selects one treatment from many alternate treatment choices available, and there is little available evidence supporting the selection of one treatment over another. This is important because many patients fail to respond to any particular treatment.

By analysing the treatment response using neuroimaging data, machine learning algorithms can be trained to predict the treatment response in new subjects. This would allow us to suggest personalised medicine for psychiatric patients based on their neuroimaging data. Personalized medicine has the potential to optimize patient care with fewer treatment trials; by predicting treatment response based on the neuroimaging data and thus selecting the optimal treatment. Patients with similar symptom profiles may have differing neurobiology that may impact treatment effectiveness, and therefore analyzing the patient-specific neuroimaging data would be vital in selecting the optimal treatment. This way the best-suited therapy can be given to the patients, saving both money and time.

Prediction of treatment response based on the baseline neuromarkers is an emerging research area. There are only a few studies that have attempted to predict the response of the treatment in brain disorder patients using neuroimaging data [10; 11; 12; 13],

however those studies lack substantial evidence due to very small sample size (less than 10 subjects) and unexplained brain basis of treatment efficacy. However in future, experiments may be performed with large sample size to predict the alternative treatment response using multi modal neuroimaging data.

Furthermore such prediction algorithms at the level of single subject would require definition of highly informative and low level features. fMRI raw data is typically high dimensional. Though we have used feature reduction techniques in our work, we acknowledge that further progress can be made to identify low-level highly informative features. Brodersen et al. [36; 41] suggested the use of weights of DCM networks as features for classification in a technique called 'generative embedding'. Future machine learning applications may use the similar feature selection strategy. Generative embedding with deep learning algorithms for classification may help achieving high level of prediction accuracy at the single subject level.

## 7.5.4 Validation studies in fMRI

fMRI has been widely used in neuroimaging research over the decade. Researchers have reported interesting insights in to the brain function under normal conditions as well as in different pathological conditions. Similarly cross species functional networks have also been identified regularly. However in the absence of the proper ground truth, many of these reported results need separate validation studies. Recently Eklund et al. [42] reported the incorrect use of statistical tools to analyse fMRI data, possibly impacting thousands of publications. This speaks for validating the results in independent studies across different research centres. Unfortunately the scientific community hails the one who presents the novel findings and the reward for those validating the findings is minimum. However the scientific community needs to adapt to give proper credit to those validating the previously known scientific results. Since fMRI data analysis aims to determine small changes in the blood flow of the brain caused by the pathological conditions or due to a specific task, fMRI results need to be taken with caution before independent validation studies are performed. Providing the data as open access to researchers around the world for validating the results may be the first step in this direction. The computational power has increased greatly in the last few years and with automatic analysis paradigms, e.g. using nipype, the whole analysis paradigms can be scripted to run automatically to validate results from previously published work.

## 7.6 References

[1] B. Biswal, F.Z. Yetkin, V.M. Haughton, and J.S. Hyde, Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med 34 (1995) 537-41.

[2] M.E. Raichle, A.M. MacLeod, A.Z. Snyder, W.J. Powers, D.A. Gusnard, and G.L. Shulman, A default mode of brain function. Proc Natl Acad Sci U S A 98 (2001) 676-82.

[3] J.L. Vincent, G.H. Patel, M.D. Fox, A.Z. Snyder, J.T. Baker, D.C. Van Essen, J.M. Zempel, L.H. Snyder, M. Corbetta, and M.E. Raichle, Intrinsic functional architecture in the anaesthetized monkey brain. Nature 447 (2007) 83-6.

[4] D.V. D'Souza, E. Jonckers, A. Bruns, B. Kunnecke, M. von Kienlin, A. Van der Linden, T. Mueggler, and M. Verhoye, Preserved modular network organization in the sedated rat brain. PLoS One 9 (2014) e106156.

[5] R.M. Hutchison, S.M. Mirsattari, C.K. Jones, J.S. Gati, and L.S. Leung, Functional networks in the anesthetized rat brain revealed by independent component analysis of resting-state FMRI. J Neurophysiol 103 (2010) 3398-406.

[6] Z. Liang, X. Liu, and N. Zhang, Dynamic resting state functional connectivity in awake and anesthetized rodents. Neuroimage 104 (2015) 89-99.

[7] X. Liu, X.H. Zhu, Y. Zhang, and W. Chen, The change of functional connectivity specificity in rats under various anesthesia levels and its neural origin. Brain Topogr 26 (2013) 363-77.

[8] K. Masamoto, M. Fukuda, A. Vazquez, and S.G. Kim, Dose-dependent effect of isoflurane on neurovascular coupling in rat cerebral cortex. Eur J Neurosci 30 (2009) 242-50.

[9] A. Silva, H. Cardoso-Cruz, F. Silva, V. Galhardo, and L. Antunes, Comparison of anesthetic depth indexes based on thalamocortical local field potentials in rats. Anesthesiology 112 (2010) 355-63.

[10] K.A. Williams, M. Magnuson, W. Majeed, S.M. LaConte, S.J. Peltier, X. Hu, and S.D. Keilholz, Comparison of alpha-chloralose, medetomidine and isoflurane anesthesia for functional connectivity mapping in the rat. Magn Reson Imaging 28 (2010) 995-1003.

[11] F. Zhao, T. Zhao, L. Zhou, Q. Wu, and X. Hu, BOLD study of stimulation-induced neural activity and resting-state connectivity in medetomidine-sedated rat. Neuroimage 39 (2008) 248-60.

[12] O. Akeju, M.L. Loggia, C. Catana, K.J. Pavone, R. Vazquez, J. Rhee, V. Contreras Ramirez, D.B. Chonde, D. Izquierdo-Garcia, G. Arabasz, S. Hsu, K. Habeeb, J.M. Hooker, V. Napadow, E.N. Brown, and P.L. Purdon, Disruption of thalamic functional connectivity is a neural correlate of dexmedetomidine-induced unconsciousness. Elife 3 (2014) e04499.

[13] J. Grandjean, A. Schroeter, I. Batata, and M. Rudin, Optimization of anesthesia protocol for resting-state fMRI in mice based on differential effects of anesthetics on functional connectivity patterns. Neuroimage 102 Pt 2 (2014) 838-47.

[14] J.A. King, T.S. Garelick, M.E. Brevard, W. Chen, T.L. Messenger, T.Q. Duong, and C.F. Ferris, Procedure for minimizing stress for fMRI studies in conscious rats. J Neurosci Methods 148 (2005) 154-60.

[15] A. Schroeter, F. Schlegel, A. Seuwen, J. Grandjean, and M. Rudin, Specificity of stimulus-evoked fMRI responses in the mouse: the influence of systemic physiological changes associated with innocuous stimulation under four different anesthetics. Neuroimage 94 (2014) 372-84.

[16] S.M. Smith, K.L. Miller, G. Salimi-Khorshidi, M. Webster, C.F. Beckmann, T.E. Nichols, J.D. Ramsey, and M.W. Woolrich, Network modelling methods for FMRI. Neuroimage 54 (2011) 875-91.

[17] N. Leonardi, W.R. Shirer, M.D. Greicius, and D. Van De Ville, Disentangling dynamic networks: Separated and joint expressions of functional connectivity patterns in time. Hum Brain Mapp 35 (2014) 5984-95.

[18] N. Leonardi, and D. Van De Ville, On spurious and real fluctuations of dynamic functional connectivity during rest. Neuroimage 104 (2015) 430-6.

[19] S.P. Kim, E. Hwang, J.H. Kang, S. Kim, and J.H. Choi, Changes in the thalamocortical connectivity during anesthesia-induced transitions in consciousness. Neuroreport 23 (2012) 294-8.

[20] X. Liu, K.K. Lauer, B.D. Ward, S.J. Li, and A.G. Hudetz, Differential effects of deep sedation with propofol on the specific and nonspecific thalamocortical systems: a functional magnetic resonance imaging study. Anesthesiology 118 (2013) 59-69.

[21] N.S. White, and M.T. Alkire, Impaired thalamocortical connectivity in humans during general-anesthetic-induced unconsciousness. Neuroimage 19 (2003) 402-11.

[22] A.E. Mechling, N.S. Hubner, H.L. Lee, J. Hennig, D. von Elverfeldt, and L.A. Harsan, Fine-grained mapping of mouse brain functional connectivity with resting-state fMRI. Neuroimage 96 (2014) 203-15.

[23] A. Liska, A. Galbusera, A.J. Schwarz, and A. Gozzi, Functional connectivity hubs of the mouse brain. Neuroimage 115 (2015) 281-91.

[24] E. Jonckers, J. Van Audekerke, G. De Visscher, A. Van der Linden, and M. Verhoye, Functional connectivity fMRI of the rodent brain: comparison of functional connectivity networks in rat and mouse. PLoS One 6 (2011) e18876.

[25] M. Welvaert, and Y. Rosseel, On the definition of signal-to-noise ratio and contrast-to-noise ratio for FMRI data. PLoS One 8 (2013) e77089.

[26] J. Grandjean, R. Derungs, L. Kulic, T. Welt, M. Henkelman, R.M. Nitsch, and M. Rudin, Complex interplay between brain function and structure during cerebral amyloidosis in APP transgenic mouse strains revealed by multi-parametric MRI comparison. Neuroimage (2016).

[27] C. Baltes, N. Radzwill, S. Bosshard, D. Marek, and M. Rudin, Micro MRI of the mouse brain using a novel 400 MHz cryogenic quadrature RF probe. NMR Biomed 22 (2009) 834-42.

[28] J. Grandjean, and M. Rudin, What can functional connectivity in mice tell us about Alzheimer\U+2019\s disease? An investigation with resting-state fMRI at 9.4T, ETH-Zürich, Zürich, 2014, pp. 1 Band.

[29] R. Wang, T. Foniok, J.I. Wamsteeker, M. Qiao, B. Tomanek, R.A. Vivanco, and U.I. Tuor, Transient blood pressure changes affect the functional magnetic resonance imaging detection of cerebral activation. Neuroimage 31 (2006) 1-11.

[30] P.L. Purdon, E.T. Pierce, G. Bonmassar, J. Walsh, P.G. Harrell, J. Kwo, D. Deschler, M. Barlow, R.C. Merhar, C. Lamus, C.M. Mullaly, M. Sullivan, S. Maginnis, D. Skoniecki, H.A. Higgins, and E.N. Brown, Simultaneous electroencephalography and functional magnetic resonance imaging of general anesthesia. Ann N Y Acad Sci 1157 (2009) 61-70.

[31] T. Horikawa, M. Tamaki, Y. Miyawaki, and Y. Kamitani, Neural decoding of visual imagery during sleep. Science 340 (2013) 639-42.

[32] I.A. Clark, K.E. Niehaus, E.P. Duff, M.C. Di Simplicio, G.D. Clifford, S.M. Smith, C.E. Mackay, M.W. Woolrich, and E.A. Holmes, First steps in using machine learning on fMRI data to predict intrusive memories of traumatic film footage. Behav Res Ther 62 (2014) 37-46.

[33] L. Griffanti, G. Salimi-Khorshidi, C.F. Beckmann, E.J. Auerbach, G. Douaud, C.E. Sexton, E. Zsoldos, K.P. Ebmeier, N. Filippini, C.E. Mackay, S. Moeller, J. Xu, E. Yacoub, G. Baselli, K. Ugurbil, K.L. Miller, and S.M. Smith, ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. Neuroimage 95 (2014) 232-47.

[34] G. Salimi-Khorshidi, G. Douaud, C.F. Beckmann, M.F. Glasser, L. Griffanti, and S.M. Smith, Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90 (2014) 449-68.

[35] V. Zerbi, J. Grandjean, M. Rudin, and N. Wenderoth, Mapping the mouse brain with rs-fMRI: An optimized pipeline for functional network identification. Neuroimage 123 (2015) 11-21.

[36] K.H. Brodersen, L. Deserno, F. Schlagenhauf, Z. Lin, W.D. Penny, J.M. Buhmann, and K.E. Stephan, Dissecting psychiatric spectrum disorders by generative embedding. Neuroimage Clin 4 (2014) 98-111.

[37] M. Dyrba, F. Barkhof, A. Fellgiebel, M. Filippi, L. Hausner, K. Hauenstein, T. Kirste, S.J. Teipel, and E.s. group, Predicting Prodromal Alzheimer's Disease in Subjects with Mild Cognitive Impairment Using Machine Learning Classification of Multimodal Multicenter Diffusion-Tensor and Magnetic Resonance Imaging Data. J Neuroimaging 25 (2015) 738-47.

[38] A. Khazaee, A. Ebrahimzadeh, and A. Babajani-Feremi, Application of advanced machine learning methods on resting-state fMRI network for identification of mild cognitive impairment and Alzheimer's disease. Brain Imaging Behav 10 (2016) 799-817.

[39] L.R. Trambaiolli, A.C. Lorena, F.J. Fraga, P.A. Kanda, R. Anghinah, and R. Nitrini, Improving Alzheimer's disease diagnosis with machine learning techniques. Clin EEG Neurosci 42 (2011) 160-5.

[40] P. McGuire, J.R. Sato, A. Mechelli, A. Jackowski, R.A. Bressan, and A. Zugman, Can neuroimaging be used to predict the onset of psychosis? Lancet Psychiatry 2 (2015) 1117-22.

[41] K.H. Brodersen, T.M. Schofield, A.P. Leff, C.S. Ong, E.I. Lomakina, J.M. Buhmann, and K.E. Stephan, Generative embedding for model-based classification of fMRI data. PLoS Comput Biol 7 (2011) e1002079.

[42] A. Eklund, T.E. Nichols, and H. Knutsson, Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci U S A 113 (2016) 7900-5.

## 8. Curriculum Vitae

I was born on 3rd October 1985 in Karachi, Pakistan and did my bachelors in Electrical Engineering from NED University of Engineering and Technology in Karachi. My major in my bachelor studies was Electrical Power System. However, I was very much interested in the signal processing area, and was selected for the prestigious Erasmus Mundus Masters program (CIMET) with full scholarship in August 2009. During my Masters, I spent my semesters in Finland, Spain and Norway and graduated with three degrees, Masters in Computer Science from University of Eastern Finland, Finland; Masters in Physics from University of Granada, Spain and Masters in Media Technology from Gjovik University College (now part of NTNU), Norway. I also had the opportunity to do 2 months internship at Heidelberg Collaboratory for Image Processing (HCI) under the supervision of Prof. Bernd Jähne and Dr. Markus Jehle. In 2011, I started my Master Thesis at Light Microscopy Center (LMC) ETH Zurich, under the supervision of Prof. Wolf Hardt and Dr. Peter Horvath. My goal was to track salmonella flowing in 3D using 2D microscopic images. We modelled point-spread function of the microscope through image processing algorithms to obtain depth information. That was my first exposure of applying engineering tools to solve biological problems and I quickly developed interest in this area. In 2012, I started my PhD at Institute of Biomedical Engineering at ETH Zurich under the supervision of Prof. Markus Rudin. Since then I am interested in understanding the brain functions by applying multivariate and pattern recognition tools to the MRI data. During my PhD, I also had the opportunity to spend a few months at PAIN lab, Children's Hospital Boston under the supervision of Prof. David Borsook and Dr. Becerra.