Functional connectivity analysis for thalassemia disease based on a graphical lasso model

Julie Coloigner, Ronald Phlypo, Adam Bush, Natasha Lepore, John Wood

HHS Public Access

Author manuscript

Proc IEEE Int Symp Biomed Imaging. Author manuscript; available in PMC 2018 October 17.

Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2016 April ; 2016: 1295–1298. doi:10.1109/ISBI.2016.7493504.

FUNCTIONAL CONNECTIVITY ANALYSIS FOR THALASSEMIA DISEASE BASED ON A GRAPHICAL LASSO MODEL

Julie Coloigner[1], **Ronald Phlypo**[2], **Adam Bush**[3], **Natasha Lepore**[1], and **John Wood**[3]

[1]CIBORG, Department of Radiology, Childrens Hospital Los Angeles, USA

[2]Research group ViBS (vision and brain signal processing), GIPSA-lab, Grenoble, France

[3]Department of Cardiology, Childrens Hospital Los Angeles, USA

Abstract

Thalassemia is a congenital disorder of hemoglobin synthesis which can lead to thromboembolic events and stroke in the brain. In this work we propose to use a functional connectivity model to discriminate between control and diseased subjects. Our connectivity measure is based on functional magnetic resonance imaging, and hence common variations of the blood oxygenation level in spatially distant areas. Analyzing this connectivity could highlight abnormal neuronal activation and provide us with a descriptor (bio-marker) of the disease. To estimate the connectivity, we propose a robust learning scheme based on the graphical lasso model, whose hyperparameter is validated within a cross-validation scheme. To analyze model fit, we transfer the mean connectivity from the control group to the thalassemic patient group. Our null hypothesis is that the model learned on control subjects is perfectly adequate (in the maximum likelihood sense) to describe the patients. The results of the permutation test suggest that the some patients with thalassemia do not have the same connectivity structure as the control.

Index Terms

fMRI; Resting state; Connectivity; Thalassemia disease; Graph theory; Graphical Lasso method

1. INTRODUCTION

β-thalassaemia is a congenital disorder of hemoglobin synthesis that is characterized by a decreased production of globin β-chains and the subsequent accumulation of unpaired α-globin subunits [1]. The presence of this globin precipitates also leads to increased red blood cell destruction (hemolysis), resulting in chronic anemia. The patients require regular blood transfusion and iron chelation therapy. Observational studies showed a high prevalence of thromboembolic events, which are mostly venous and their occurrence increases with age [1, 2]. Its effect on the brain has not been widely studied and the incidence of strokes has only recently been established in [2]. 37% of patients showed asymptomatic brain damage, which involved the subcortical white matter in all patients on magnetic resonance imaging (MRI) [3].

In addition to these investigations into structural MRI, analysis of functional connectivity remains promising as a neuroimaging marker of thalassemia. Indeed, functional MRI (fMRI)

studies have found that spontaneous low-frequency ($0.009 \leq f \leq 0.08$ Hz) blood oxygenation level-dependent contrast fluctuations measured during rest showed high temporal coherence between functionally related brain regions [4]. This result suggests that fMRI might also be appropriate for examining functional connectivity. Mathematically, graph theory is employed to evaluate fMRI brain network connectivity, which models the brain as a complex network composed of nodes associated with regions of interest, and edges representing functional connectivity between nodes. Various statistical methods have been adopted to infer the latent connectivity from neuroimaging data. One of the techniques is partial correlation analysis (normalized version of the inverse covariance matrix), which provides a much better modelization of brain connectivity than simple correlation analysis. The inverse covariance matrix can be considered a faithful representation of a Gaussian graphical model. Indeed a zero entry corresponds to conditional independence of a pair of nodes, given their complement in the network. This constraint is also valid because neurological findings have demonstrated that a brain region usually only interacts directly with a few other brain regions. We will here use graphical lasso to estimate a such sparse network [5]. Extension of these techniques estimate intra-subject brain connectivity model under the assumption that they have the same structure can be found in all individuals in the group [6, 7, 8]. It has proven sensitive to correctly detect existing connections in a graph structure under some mild assumptions. Different patterns of brain connectivity have been shown in Alzheimer's disease [9].

In this paper, starting from a correlation metric, a robust learning scheme is proposed based on a graphical Lasso model whose hyperparameter is fixed within a crossvalidation scheme, in order to estimate a descriptor of reference brain activation. Thalassemia patients could be affected by strokes, altering some connectivity patterns. Given this variability of the damage and the abnormal blood flow in thalassemia, we hypothesize that at least for some patient's abnormalities in connectivity patterns exist in patients as compared to control subjects. Based on the Gaussian likelihood model, a network cross-validated on control subjects is individually verified on thalassemia patient data. Our null hypothesis is that such a model learned on control subjects is perfectly adequate to describe patients. Permutation tests are used to assess the p-value of every patient. The statistical method is tested on a cohort of 20 control subjects and 11 thalassemia patients.

2. MATERIALS

2.1. Dataset

This study includes thalassemia patients and control subjects that were recruited from the Children's Hospital Los Angeles between January 2012 and September 2015. The institutional review board approved the protocol and written informed consent was obtained from all subjects. All participants underwent MR imaging by using a 3T Philips Achieva. The 3D T1-weighted images were acquired covering the whole brain (160 sagittal slices) with TR = 8.20 s, TE = 3.77 ms, flip angle = 8, in-plane resolution = 256 × 256, FOV = 256 mm × 224 mm and thickness/gap =1.0/0 mm). During resting-state fMRI scanning, subjects were instructed to close their eyes, keep still as much as possible, not to think of anything, and not to fall asleep. The functional images were acquired with the following parameters:

TR = 2000 ms, TE = 50 ms, flip angle = 90, in-plane resolution = 96 × 96, FOV = 220 mm × 220 mm, 26 axial slices, thickness/gap= 5/0 mm. A total of 240 volumes were collected in 8 minutes. At this stage of the experiment, the population sample consists of 11 patients with thalassemia and 20 control volunteers, but data acquisition is ongoing.

2.2. Preprocessing steps

Imaging data was first preprocessed with the FMRIB Software Library (FSL), using standard spatial preprocessing steps: images were (1) slice-time corrected, (2) realigned to remove physiological motion with FSL's MCFLIRT, (3) intensity normalized and (4) smoothed with a 6-mm Gaussian kernel. The fMRI images were co-registered with their anatomical counterpart using FSL and linearly transformed to the Montreal Neurological Institute (MNI) template space. After fMRI preprocessing, in order to remove residual motion, the parameters estimated by FSL's MCFLIRT, as well as theirs derivatives, were used as regressors, computed by backwards differences (12 regressors). Moreover, the physiological noise was also reduced using other nuisance regressors that were calculated by the CompCor [10] method as implemented in Nilearn (http://nilearn.github.io/). These 5 regressors correspond to the principal components from noisy regions-of-non-interest, such as white matter, cerebral spinal fluid and out-of brain.

2.3. Regions of interest selection

In order to analyze functional connectivity, the resting state data was parcellated into 39 regions using the multi-subject probabilistic atlas from [11]. It has been shown to provide good support to define regions of interest in fMRI studies, and it is known that the the choice of the regions of interest (ROI) that define the nodes of the graphs has a great influence on the assessment of the connectomes. Regional mean time-series are estimated by averaging the fMRI signals over all voxels within each atlas region. From these regional mean time-series, we can compute the empirical correlation matrix, denoted $\boldsymbol{S}^{(k)}$, for each participant. We also show an example of this matrix used for learning the model, as described in Figure 1. Note that it displays features with high connectivity in the upper left part, corresponding to the default mode network.

3. METHOD

In this section, we present a robust learning scheme, through which we estimate a descriptor of the reference functional connectivity structure. A sparse modeling framework is employed based on a graphical Lasso model associated with a cross-validation scheme. Then, based on the Gaussian likelihood model, a mean network estimated with control subjects is applied to the thalassemia patient data. Permutation tests are used to assess the sensitivity and specificity of our model.

3.1. Learning model

Starting from a training set of K control subjects, the graphical Lasso model is represented through a precision matrix, $\boldsymbol{\Theta}$, corresponding to the inverse correlation matrix, $\boldsymbol{\Theta} = \boldsymbol{S}^{-1}$. We assume that the variables have a multivariate Gaussian distribution with covariance $\boldsymbol{S}$. If the

edge linking nodes j and i is absent, then nodes j and i are conditionally independent given all others, and the corresponding entry of the inverse covariance matrix, $\boldsymbol{\Theta}$, is zero.

The estimation of $\boldsymbol{\Theta}$ is based on 3 steps: i) the set of precision matrices $\{\boldsymbol{\Theta}^{(1)}, \cdots, \boldsymbol{\Theta}^{(K-1)}\}$ conditional on a value λ are first estimated. ii) λ is fixed through a 10-fold crossvalidation (each subset of $K-1$ controls are validated on 1 control with $K=10$) iii) The model is tested on the K-th subject by calculating the likelihood of the patient data in the model with parameter $\widehat{\boldsymbol{\Theta}}$, the mean precision.

For step i), to estimate a sparse symmetric positive definite parameter, $\boldsymbol{\Theta}^{(K)}$, we maximize the penalized Gaussian log-likelihood, for all $K-1$ subjects [12]:

$$\boldsymbol{\Theta}^{(k)} = \underset{\boldsymbol{\Theta}^{(k)} > 0}{\arg\min} \text{Trace}(\boldsymbol{\Theta}^{(k)}\boldsymbol{S}^{(k)}) - \text{logdet}(\boldsymbol{\Theta}^{(k)}) + \lambda\|\boldsymbol{\Theta}^{(k)}\|_1 \quad (1)$$

over a grid of values for the penalization parameter λ and where $\boldsymbol{S}^{(k)}$ is the covariance matrix of the k-th subject. Note that log det and Trace correspond to the logarithm of the determinant and the trace of the matrix, respectively. The operator $\| \|_1$ is the l^1-norm matrix. The parameter λ is a penalty parameter, which controls the amount of sparsity. The block-coordinate descent optimization algorithm proposed by Friedman et al. [5] is used, which ensures the symmetric positive definiteness of the estimate $\boldsymbol{\Theta}^{(K)}$, during the optimization procedure.

3.2. Cross-validation

To train the model, a leave-one-out cross-validation scheme is used. A subject is extracted from the training set as a validation subject, and the model is trained (i.e. estimate the model parameter λ is estimated) with the remaining $K-1$ participants. This is repeated such that each control subject is used as a validation sample. We report the optimal parameter λ, that maximizes the mean log-likelihood over all K-folds.

Using this optimal λ for every subject k, the precision matrix, $\boldsymbol{\Theta}^{k}$ is estimated again. From $\{\boldsymbol{\Theta}^{1}, \cdots, \boldsymbol{\Theta}^{K-1}\}$, the Frechet mean is used to estimate the reference connectivity model, $\widehat{\boldsymbol{\Theta}}$ [6].

3.3. Permutation test

The model determined above to estimate healthy brain activation is applied to the Q data of thalassemia patients. Our null hypothesis (H0) is that the model learned on control subjects is perfectly adequate (in the maximum likelihood sense) to describe the patients. For each of the patients (with labels $q=1, \ldots, Q$), the log-likelihood in a Gaussian distribution with mean 0 and the covariance $\widehat{\boldsymbol{\Theta}}$ is calculated. We permute the data by shuffling the subjects' labels (control or patient), and then calculate the same log-likelihood statistics for every permutation. Finally, the statistical significance, expressed in terms of a p-value, is calculated as the fraction of likelihood that is at least as high as the original (non-permuted) statistic, which was derived from the correctly labeled data.

4. RESULTS

First, the healthy brain functional connectivity structure was estimated from the 12 control subjects. A grid value of λ from 0 to 1 was used and a leave-one-out scheme was repeated $K = 12$ times. Figure 2 displays the inverse correlation model, giving a descriptor of reference activation, obtained by the learning scheme. Compared to the covariance matrix of a subject, displayed in Figure 1, we observe that the precision matrix is more sparse, as expected. Indeed, only direct connections between two ROIs are represented in the inverse covariance matrix. In addition, graphical representation of the connection for the atlas-based ROIs are shown in Figure 3. It displays the highest scores related to the homologous interhemispheric connections; in other words, each region tends to be highly correlated with the corresponding region in the opposite hemisphere. The sagittal view of the Figure 3 shows the connections between the medial prefrontal cortex and the precuneus, implicated in well known networks such as the default mode network.

Then, the model estimated previously is applied to the $Q = 11$ thalassemia patient data. Our null hypothesis is that such a model learned on control subjects is perfectly adequate (in the maximum likelihood sense) to describe thalassemia patients. Three patients had a p-value inferior to 5%, leading to the conclusion that H0 is rejected. The results of the permutation test suggest that the some patients with thalassemia do not have the same connectivity structure as the controls. Higher p-values (p-value>0.1) were observed for 7 patients, which means that the likelihood of these data in the model do not allow to reject H0. Our classification results were compared with the one obtained by a correlation matrix. The p-values of all subjects are high. This model seems not to show differences between the patients. Significant correlation was found between p-values and global CBF estimations.

5. CONCLUSION

In this paper, we proposed a robust learning scheme based on a graphical Lasso model to estimate the reference brain functional connectivity structure. Based on a permutation test, we showed that the model learned on control subjects is not adequate to describe several thalassemia patients. The results of the permutation test suggest that three thalassemia patients do not have the same connectivity structure as compared to reference subjects. Indeed, the strokes and abnormal blood flow in thalassemia patients can affect brain activation patterns. This method may help to identify the patients at greatest risk for cerebral strokes and for neurocongitive dysfunctions. This study will provide insight into risk factors for cerebrovascular disease in SCD patients, which will facilitate improved neuroprotection and quality of life. Future works will consider the inclusion of individual’s clinical variables, such as stroke, hemoglobin levels and neurocognitive functioning and will explore their impact on the connectivity patterns. Finally, this is an ongoing study, and we plan to pursue it on a larger database.

References

1. Musallam KM, Taher AT, Karimi M, Rachmilewitz EA. Cerebral infarction in β-thalassemia intermedia: breaking the silence. Thrombosis research. 2012; 130(5):695–702. [PubMed: 22857801]

2. Musallam KM, Beydoun A, Hourani R, Nasreddine W, Raad R, Koussa S, Taher AT. Brain magnetic resonance angiography in splenectomized adults with β-thalassemia intermedia. European journal of haematology. 2011; 87(6):539–546. [PubMed: 21913989]

3. Taher AT, Musallam KM, Nasreddine W, Hourani R, Inati A, Beydoun A. Asymptomatic brain magnetic resonance imaging abnormalities in splenectomized adults with thalassemia intermedia. Journal of Thrombosis and Haemostasis. 2010; 8(1):54–59. [PubMed: 19817994]

4. Fox MD, Greicius M. Clinical applications of resting state functional connectivity. Front Syst Neuroscience. 2010:4–19.

5. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008; 9(3):432–441. [PubMed: 18079126]

6. Varoquaux G, Baronnet F, Kleinschmidt A, Fillard P, Thirion B. Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. MICCAI, vol 6361 of Lecture Notes in Computer Science. 2010:200–208.

7. Phlypo R, Thirion B, Varoquaux G. MICCAI. Springer; 2014. Deriving a multi-subject functional-connectivity atlas to inform connectome estimation; 185–192.

8. Ng B, Varoquaux G, Poline JB, Thirion B. Information Processing in Medical Imaging. Springer; 2013. A novel sparse group gaussian graphical model for functional connectivity estimation; 256–267.

9. Huang S, Li J, Sun L, Ye J, Fleisher A, Wu T, Chen K, Reiman E. Learning brain connectivity of alzheimer's disease by sparse inverse covariance estimation. NeuroImage. 2010; 50(3):935–949. [PubMed: 20079441]

10. Behzadi Y, Restom K, Liau J, Liu T. A component based noise correction method (compcor) for bold and perfusion based fmri. Neuroimage. 2007; 375(1):90–101.

11. Varoquaux G, Gramfort A, Pedregosa F, Michel V, Thirion B. Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. Information Processing in Medical Imaging Springer. 2011:562–573.

12. Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ. Functional connectivity: the principal-component analysis of large (pet) data sets. Journal of cerebral blood flow and metabolism. 1993; 13:5–5. [PubMed: 8417010]

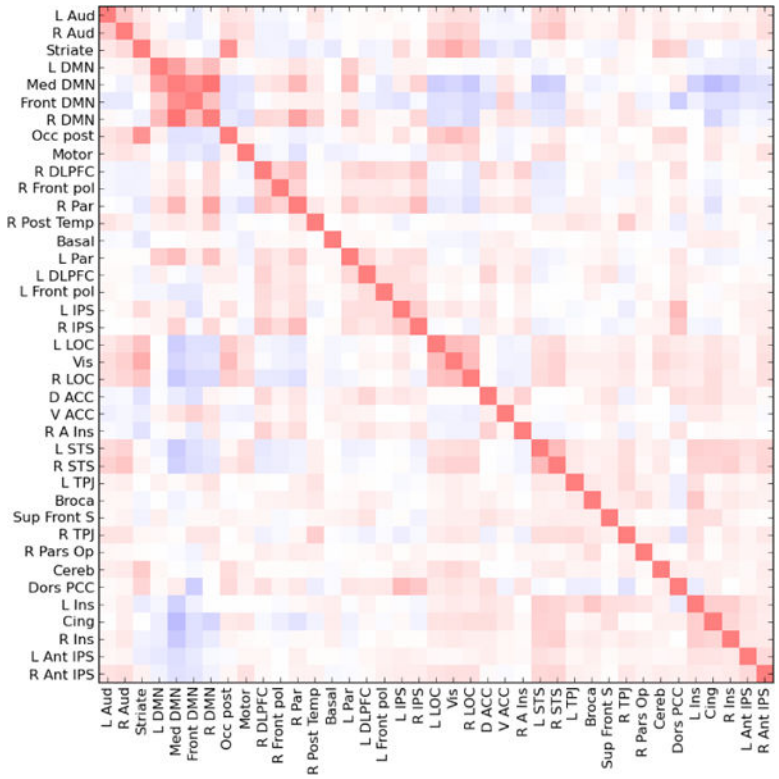

Fig. 1.

Connectivity matrix from the 39 regions: each coefficient in the matrix represents the correlation coefficient between two ROIs [7].

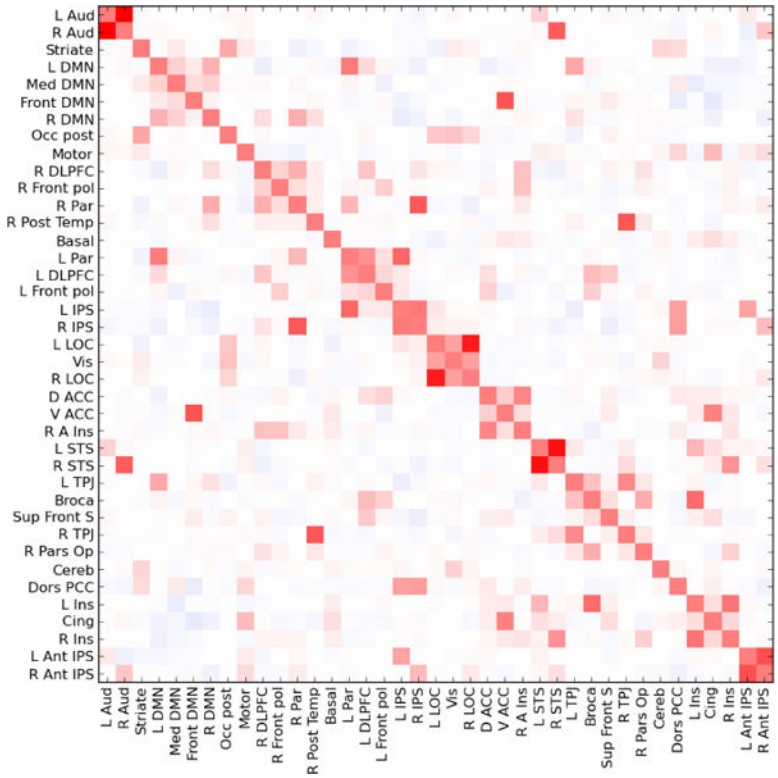

Fig. 2.

Inverse correlation model, giving a descriptor of reference brain activation, computed by the procedure described in sections 3.1 and 3.2.

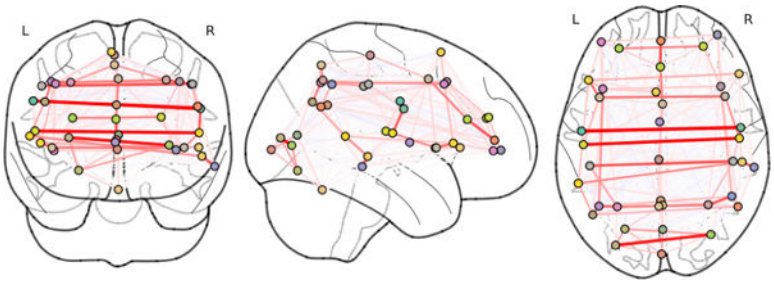

Fig. 3.
Graphical representation of the reference connectivity model.