Reproducibility of Functional Connectivity Estimates in Motion Corrected Fetal fMRI

Daniel Sobotka[1]([✉]), Roxane Licandro[1,3], Michael Ebner[4], Ernst Schwartz[1], Tom Vercauteren[4], Sebastien Ourselin[4], Gregor Kasprian[2], Daniela Prayer[2], and Georg Langs[1]

[1] Computational Imaging Research Lab, Department of Biomedical Imaging and Image-Guided Therapy, Medical University of Vienna, Vienna, Austria
daniel.sobotka@meduniwien.ac.at
[2] Division of Neuroradiology and Musculoskeletal Radiology, Department of Biomedical Imaging and Image-Guided Therapy, Medical University of Vienna, Vienna, Austria
[3] Computer Vision Lab, Institute of Visual Computing and Human-Centered Technology, TU Wien, Vienna, Austria
[4] School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

Abstract. Preprocessing and motion correction are essential steps in resting state functional Magnetic Resonance Imaging (rs-fMRI) of the fetal brain. They aim to address the difficult task of removing artefacts caused by fetal movement or maternal breathing, and aim to suppress erroneous signal correlations caused by motion. While preprocessing standards have been established in the adult brain, motion correction of fetal rs-fMRI and subsequent interpretation of processed data is still challenging. Here, we evaluate the effect of different preprocessing methods and motion correction on rs-fMRI sequences by assessing reproducibility of functional connectivity estimates. For slice-based motion correction of 4D fetal rs-fMRI, we extend a high-resolution reconstruction approach presented for structural fetal MRI. Correlation, standard deviation and structural similarity index are evaluated on the whole cortex, on specific regions and at different gestational ages. Results show improved reproducibility and signal interpretability after preprocessing with motion correction enabling the quantification of long-range correlation patterns of the developing default mode network in the fetal brain.

1 Introduction

The analysis of spatial and temporal brain signal correlations forms a key component to understand the maturation processes of brain activity, their interaction and their link to cognition in the developing brain [4]. Preprocessing methods used in functional Magnetic Resonance Imaging (fMRI) have been developed for adult or infant brains and have recently been also applied on

fetal rs-fMRI [4,13]. Here, motion correction is particularly important and necessary, due to continuous movement of the fetus itself or causes such as maternal breathing. Subsequent analysis of Functional Connectivity (FC) relies on the assumption that measurements have neural origin, while signal disruption and motion artifacts can artificially increase the correlation between brain voxels even after re-alignment of image data and thus distort study results [10]. Existing fetal studies [3,14,15,17] used different processing combinations as normalization, smoothing, motion censoring, motion regression or motion correction. The specific effects of these methods on the reliability of the resulting fetal rs-fMRI signals and corresponding FC are poorly understood.

Contribution. Here, we assess the effect of state-of-the-art preprocessing techniques on the *reproducibility* of rs-fMRI signals and the computation of short- and long-range functional connectivity in the fetal brain, providing an evaluation scheme and corresponding metrics. Jakab et al. [5] used correlation, standard deviation and the structural similarity index as metrics for evaluating within-subject reproducibility in diffusion tensor imaging. Inspired by this approach, we applied these metrics on resting state fetal fMRI for the proposed preprocessing pipelines. To correct for motion, we extend the 3D Motion Correction (MC) and High-Resolution-Reconstruction (HRR) approach for fetal MRI proposed in [1] for 4D fetal rs-fMRI. Quality assessment of the signal is a necessary step, since there exists no standardized pipeline for fetal fMRI preprocessing. We present different quality assessment schemata to evaluate the signal before and after different preprocessing approaches on the cortex, on specific regions and age related dependencies. The proposed reproducibility evaluation scheme is introduced in Sect. 2. The evaluation results are presented in Sect. 3 and in Sect. 4 this work concludes with a discussion of optimal preprocessing of fetal rs-fMRI and discussion of possible future directions.

2   Methodology

In this section, the proposed evaluation framework and slice-based motion correction of 4D fetal rs-fMRI is summarized. Subsequently, the proposed signal quality assessment strategy is presented. The study population and imaging protocol used for evaluation is introduced in Sect. 3.

Structural Preprocessing: Fetal MRI preprocessing included atlas-based alignment, brain segmentation, generating of cortex meshes [11] and manual registration with functional data.

Preprocessing Pipelines: We incorporated 7 different fMRI preprocessing pipelines (cf. Table 1 for more detail) into the reproducibility test framework proposed. Inspired by [9] we used combinations of bias field correction [19], slice timing correction [6], high resolution 4D motion correction (see Sect. 2.1 for detailed information) and motion regression [9].

2.1 4D High Resolution Motion Correction (HRMC)

In this work, two different HRMC strategies are proposed: (1) Volume-to-Volume (V2V) and (2) Slice-to-Volume (S2V) HRMC for fetal brain rs-fMRI. *Volume-to-Volume HRMC* is performed by rigidly registering each stack (time point) individually to a target fMRI stack using symmetric block-matching based on normalized cross correlation [8]. For individual *Slice-to-Volume HRMC*, a higher-resolution reference volume is estimated by using the first 15 time points to create a 1 mm isotropic volume with the super-resolution reconstruction framework [1], whereby three two-step motion-correction/reconstruction cycles are performed. Subsequently, all slice stacks each acquired at the same time point are rigidly registered to this higher-resolution reference using normalized cross correlation as similarity measure. The final volumes are reconstructed on the original grid by solving the slice acquisition model [1,2] in a least-squares formulation using first-order Tikhonov regularization, i.e.

$$\min_{\mathbf{x} \geq 0} \Big(\sum_{k=1}^{K} \frac{1}{2} \|\mathbf{y}_k - \mathbf{A}_k\mathbf{x}\|_{\ell^2}^2 + \frac{\alpha}{2} \|\boldsymbol{\nabla}\mathbf{x}\|_{\ell^2}^2 \Big), \tag{1}$$

for all individual slices $\mathbf{x}_k$, $k = 1, \ldots, K$ associated with a single time point. This takes into account either the obtained Volume-to-Volume or Slice-to-Volume motion estimates for the linear blurring and downsampling operator $\mathbf{A}_k$ [1].

Table 1. Functional preprocessing pipelines incorporated into the framework proposed. Each pipeline has different combinations of bias field correction (BFC), slice timing correction (STC), Slice-to-Volume motion correction (S2V), Volume-to-Volume motion correction (V2V) and motion regression (MR)

	BFC	STC	BFC+STC	STC+BFC	S2V	V2V	MR
Pipeline 1 (P1)	✓						
Pipeline 2 (P2)		✓					
Pipeline 3 (P3)			✓				
Pipeline 4 (P4)				✓			
Pipeline 5 (P5)			✓		✓		
Pipeline 6 (P6)			✓			✓	
Pipeline 7 (P7)			✓		✓		✓

2.2 Short-Range and Long-Range Connectivity Computation

The Pearson correlation coefficient is computed between the time course t ($t = 1, \ldots, M$; M is the number of time frames) of each brain node $x_i(t)$ and $x_j(t)$ (i, j $= 1, \ldots, N$, where N is the number of nodes observed) [7,12]:

$$\mathbf{CM_t} = \frac{\sum[(x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j)]}{\sqrt{\sum[(x_i(t) - \bar{x}_i)^2(x_j(t) - \bar{x}_j)^2]}} \tag{2}$$

As a result an $N \times N$ correlation matrix CM_t for every subject S was obtained, with $\bar{x}_i$, $\bar{x}_j$ the mean node intensity across all time points at position i and j. To define short- and long-range connectivity, we calculate the Euclidean distance (ED) between coordinates of nodes. For every cortical node, we count high correlating time courses (threshold ≥ 0.4), and assign them to short- and long-range splitting at a distance roughly equivalent to 15 mm in an adult brain [12]. This distance is changed from 4.4 mm (gestational age of 20 weeks) to 8.8 mm (gestational age of 40 weeks) in relation to the fetal brain size, since fetus' brains are resampled on a standard brain (fsaverage5)[1], which can introduce correlations from nearby brain nodes [7,12].

2.3   Assessment of Reproducibility

According to [10] signal disruption and motion artifacts increase the correlation between brain voxels and distort signals. We hypothesize that signals of two time ranges of a subject should be more similar after preprocessing, compared to the uncorrected signals, if artefacts are removed. Thus, we divided the rs-fMRI associated with each fetus in two time ranges u and v. We observed that there may be more fetal movement and maternal breathing at the beginning of the recording session, which led us to the following definition of the two time ranges: $u = [[1, \frac{M}{4}], [\frac{2M}{4}, \frac{3M}{4}]]$ and $v = [[\frac{M}{4}, \frac{2M}{4}], [\frac{3M}{4}, M]]$ where M is the number of time points in each dataset. For assessing the reproducibility of a subject's signal after preprocessing, the difference of correlations (ΔC) and standard deviations ($\Delta \sigma$) between a subject's S extracted time courses $x(u)$ and $x(v)$ are computed as well as the *SSIM* index [5].

Correlation Difference ΔC. In a first step for $x_S(u)$ and $x_S(v)$ correlation matrices CM_u^S and CM_v^S are computed following Eq. 2. Subsequently, the correlation difference is computed following Eq. 3

$$\Delta C_S = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |CM_u(i,j) - CM_v(i,j)| \tag{3}$$

Standard deviation Difference $\Delta \sigma$. The standard deviation σ of a time course t at node x of a subject is calculated using Eq. 4, where $\bar{x}$ is the mean of the time course $x(t)$ at node x:

$$\sigma_t = \sqrt{\frac{1}{M} \sum_{t=1}^{M} (x(t) - \bar{x})^2} \tag{4}$$

Subsequently, for every subject the standard deviation difference $\Delta \sigma$ is computed based on standard deviation estimates of time course u and v using Eq. 5.

$$\Delta \sigma = \frac{1}{N} \sum_{i=1}^{N} |\sigma_u - \sigma_v| \tag{5}$$

---

[1] https://surfer.nmr.mgh.harvard.edu/.

Structural Similarity (SSIM) Index. Is a quality assessment metric [5,16], which is calculated between $x(u)$ and $x(v)$ for all brain nodes of a subject.

$$SSIM(u,v) = [l(u,v)]^\alpha [c(u,v)]^\beta [s(u,v)]^\gamma \tag{6}$$

It consists of three terms, the luminance, contrast and structural term:

$$l(u,v) = \frac{2\mu_u\mu_v + c_1}{\mu_u^2 + \mu_v^2 + c_1} \tag{7}$$

$$c(u,v) = \frac{2\sigma_u\sigma_v + c_2}{\sigma_u^2 + \sigma_v^2 + c_2} \tag{8}$$

$$s(u,v) = \frac{\sigma_{uv} + c_3}{\sigma_u\sigma_v + c_3} \tag{9}$$

where μ_u, μ_v, σ_u, σ_v and σ_{uv} are the means, standard deviations and cross covariance. α, β and γ are used to adjust relative importance of the three terms, where the constants c_1, c_2 and c_3 are included to avoid term instabilities [16].

3   Results

We analysed the reproducibility of a subject's signal after the application of 7 different preprocessing pipelines using the difference of correlations (ΔC), standard deviations ($\Delta\sigma$) and the *SSIM* index [5] as evaluation metrics (introduced in Sect. 2.3).

Data. The study includes a total of 21 fMRI sequences from fetuses between the 20th and 40th gestational week (GW, mean: 28.43, standard deviation: 5.43) with normal brain development. Functional magnetic resonance imaging was performed on a 1.5 T clinical scanner (Philips Medical Systems, Best, The Netherlands) using a sensitivity encoding (SENSE) cardiac coil with five elements (three posterior, two anterior) wrapped around the mother's abdomen, utilizing single-shot gradient-recalled echo-planar imaging (EPI) and no cardiac gating with the following setup: 50 ms echo time, 1000 ms repetition time, 3 mm slice thickness, 18 slices and 96 volumes. The pregnant women were examined in the supine or left decubitus position (feet first), and no contrast agents or sedatives were administered. In order to receive the optimal MR signal, the coil was readjusted depending on the position of the fetal head during the imaging procedure.

3.1   Reproducibility of Functional Connectivity on the Cortex

In Fig. 1 (upper left plot) a boxplot of the ΔC metric estimated over all subjects, for the uncorrected signal and for the signal after every 7 preprocessing approaches is visualised. The $\Delta\sigma$ and *SSIM* metric are visualised in the same way in the upper right and lower right part of Fig. 1. In case of correlation and standard deviation a low value refers to better reproducibility, while for the similarity
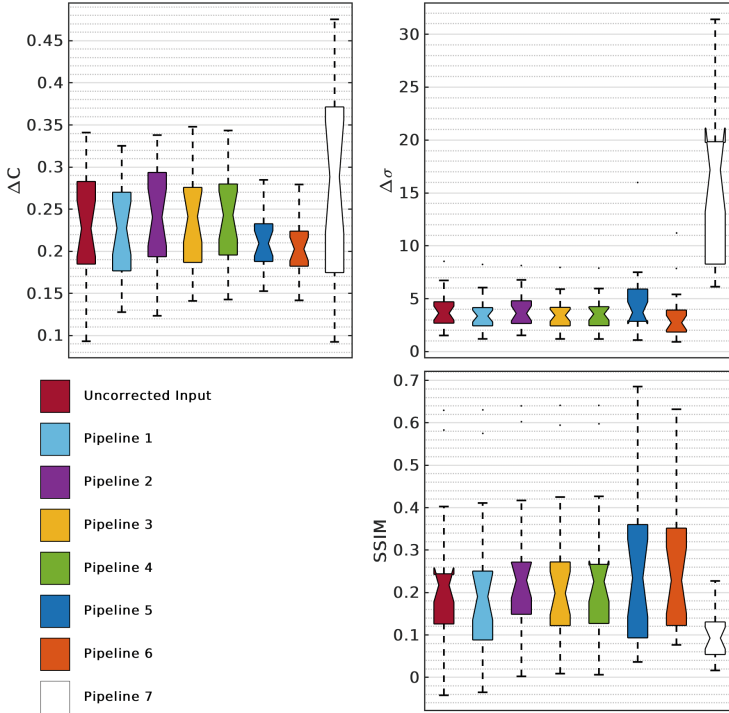
Fig. 1. Reproducibility metrics with correlation differences, standard deviation differences and structural similarity index comparison between the uncorrected input (UNC), bias field correction (BFC), slice timing correction (STC), Slice-to-Volume motion correction (S2V MC), volume to volume motion correction (V2V MC) and motion regression (MR).

index a higher value is interpreted as better reproducibility. First we evaluated if bias field correction and slice timing correction have a positive impact on the reproducibility. Therefore, the uncorrected signal (UNC) is preprocessed using Pipeline P1, P2, P3 and P4 introduced in Sect. 2.

Among P1–P4, P3 shows the best result, since the correlation differences (mean: 0.24, SD: 0.06) and standard deviation differences (mean: 3.53, SD: 1.65) are reduced and the SSIM score shows similar results (mean: 0.23, SD: 0.17) compared with the pipelines P1, P2 and P4. Thus, building on the P3, the Slice-to-Volume (S2V) and Volume-to-Volume (V2V) motion correction approaches are evaluated (P5 and P6) and visualised in Fig. 1. The correlation differences of S2V (P5, mean: 0.21) and V2V (P6, mean: 0.21) show similar results, while S2V leads to higher standard deviation differences (mean: 4.52), but a higher SSIM value (mean: 0.25, SD: 0.18, Q3: 0.36) compared to V2V (mean: 0.25, SD: 0.16, Q3: 0.35). Therefore, we chose P5 as the best preprocessing pipeline. An increase of the average SSIM mean value from 0.2 (UNC) to 0.25 is observable after motion correction, which can be refered to a positive effect from the motion

correction technique. Motion regression (P7) relies on the precise estimate of motion parameters during the alignment, errors in the estimates can cause the regression to introduce or amplify artifacts in the data leading to comparably worse reproducibility (mean SSIM: 0.1, SD SSIM: 0.06, ΔC: 0.27, $\Delta\sigma$: 15.69). In that light, using other proxy measures of motion induced signal might be a better strategy. The three evaluation measures assess the reproducibility of signal correlation analysis, and the overall loss of structure in the data. The value of reproducibility as a quality measure relies on the assumption that motion is different across the entire scan.

3.2 Reproducibility of Functional Connectivity in 7 Yeo Networks

We used the Yeo parcellation [18] to subdivide the brain into seven networks (visual (Yeo 1), somatomotor (Yeo 2), dorsal attention (Yeo 3), ventral attention (Yeo 4), limbic (Yeo 5), frontoparietal (Yeo 6) and default mode network (Yeo 7)). Figure 2 shows boxplots of the correlation differences over all subjects for uncorrected (UNC, red) and Pipeline 5 (blue) for all Yeo networks. The results indicate that the signal after applying Pipeline 5 is more reproducible compared to the uncorrected input, since a reduced correlation difference and a higher SSIM values are observable. Furthermore it shows consistant differences across networks, with highest SSIM in ventral attention (mean: 0.27), limbic (mean: 0.36) and frontoparietal networks (mean: 0.3).
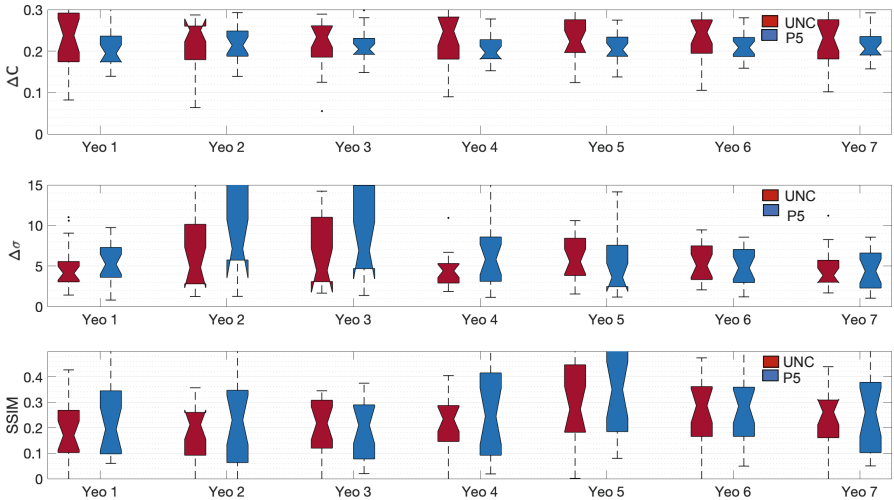

Fig. 2. Correlation differences (top), standard deviation differences (middle) and SSIM (bottom) between the uncorrected input (UNC) and after application of Pipeline 5 for each Yeo network. (Color figure online)

3.3   Age-Related Reproducibility

To test if age has an influence on reproducibility, we divided our dataset into two age ranges: GW 20–24 (6 subjects) and GW 25–40 (15 subjects), motivated by pronounced cortical folding process starting around the GW 24 [11]. In both ranges, motion correction improves reproducibility, and the resulting value ranges are largely comparable, but more data is needed to test for specific trends. Figure 3 shows values for the default mode network (Yeo 7).
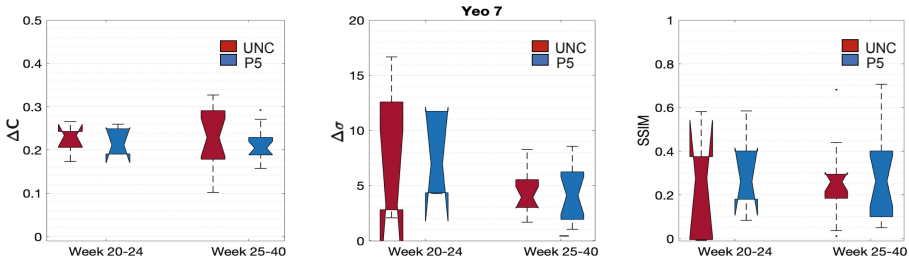

Fig. 3. Age related correlation differences (left), standard deviation differences (middle) and SSIM value (right) in the default mode network (Yeo 7) of the uncorrected input (UNC) and after application of Pipeline 5 (P5).

3.4   Connectivity Comparison

Finally, in the last experiment we compare the degree of short- and long-range connectivity before and after preprocessing on every cortical surface point. Figure 4 shows for the uncorrected input (top row) for each of the two parts of the time course the mean short- and long-range degree value visualized on the surface over all subjects. The bottom row shows the connectivity after the best reproducibility preprocessing pipeline P5 including bias field, slice timing and Slice-to-Volume motion correction. The short-range connectivity (left side) is less sensitive to motion compared to long-range connectivity (right side), and preprocessing shows a stronger effect. In particular long-range connectivity shows high values across the entire cortex, while after motion correction, a more nuanced image emerges. High long-range connection areas partly corresponding to the default mode network become visible, suggesting that these network develops already during gestation. The SSIM values between the two time windows on the cortex for the uncorrected input (short-range: 0.92, long-range: 0.36) and after preprocessing (short-range: 0.93, long-range: 0.26) indicate, that with preprocessing a higher short-range reproducibility is achieved. The limitations of the SSIM metric are visible in the long-range comparison, where the motion motivated uncorrected input obtained a higher SSIM value as after preprocessing.
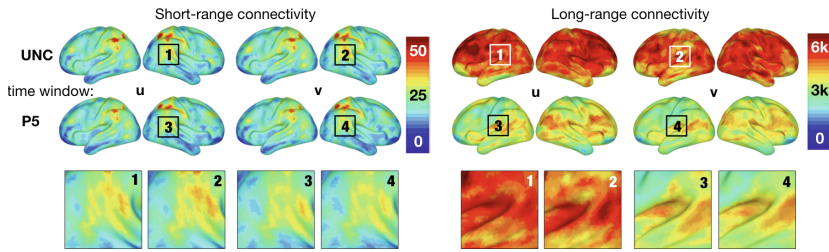
Fig. 4. Short- and long-range mean connectivity degree value visualized on the surface between the uncorrected input and after preprocessing with Pipeline 5. Long-range connections benefit substantially from preprocessing.

4   Conclusion

In this work, we introduced a *reproduciblity test framework*, for evaluating the effect of 7 different preprocessing and motion correction pipelines for fetal rs-fMRI sequences and corresponding functional connectivity estimates. The comparisons of the proposed pipelines were performed based on the reproduciblity of correlation, standard deviation and the structural similarity index for two parts of every time course from each subject. The combination of bias field, slice timing and slice-to-volume motion correction performed best. We showed that preprocessing with motion correction leads to better reproducibility results on the whole cortex and on the Yeo 7 networks. We show that preprocessing has a positive effect on reproducibility for in utero rs-fMRI acquisitions, and in particular that long-range connectivity is more sensitive to motion artefacts compared to short-range connectivity patterns. Reproducible long-range connectivity are located at the default mode network after applying preprocessing and motion correction. For future work, we will use a greater population to increase the generalisability and investigate how short-range an long-range patterns develop during gestation across the cortex. We did not study the link between motion and gestational age in this paper, but note that there might be a relationship. Another point of future work is to take motion estimates for assessing the impact of different levels of motion into account.

References

1. Ebner, M., et al.: An automated localization, segmentation and reconstruction framework for fetal brain MRI. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 313–320. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_36

2. Gholipour, A., Estroff, J.A., Warfield, S.K.: Robust super-resolution volume recon-struction from slice acquisitions: application to fetal brain MRI. IEEE Trans. Med. Imaging **29**(10), 1739–1758 (2010)

3. van den Heuvel, M.I., et al.: Hubs in the human fetal brain network. Dev. Cogn. Neurosci. **30**, 108–115 (2018)

4. Jakab, A., et al.: Fetal functional imaging portrays heterogeneous development of emerging human brain networks. Front. Hum. Neurosci. **8**, 852 (2014)

5. Jakab, A., Tuura, R., Kellenberger, C., Scheer, I.: In utero diffusion tensor imaging of the fetal brain: a reproducibility study. NeuroImage Clin. **15**, 601–612 (2017)

6. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: Fsl. Neuroimage **62**(2), 782–790 (2012)

7. Licandro, R., Nenning, K.H., Schwartz, E., Kollndorfer, K., Bartha-Doering, L., Langs, G.: Changing functional connectivity in the child's developing brain affected by ischaemic stroke. In: Proceedings MICCAI PIPPI (2016)

8. Modat, M., Cash, D.M., Daga, P., Winston, G.P., Duncan, J.S., Ourselin, S.: Global image registration using a symmetric block-matching approach. J. Med. Imaging **1**(2), 024003 (2014)

9. Poldrack, R.A., Mumford, J.A., Nichols, T.E.: Handbook of Functional MRI Data Analysis. Cambridge University Press, Cambridge (2011)

10. Power, J.D., Schlaggar, B.L., Petersen, S.E.: Recent progress and outstanding issues in motion correction in resting state fMRI. Neuroimage **105**, 536–551 (2015)

11. Schwartz, E., Kasprian, G., Jakab, A., Prayer, D., Schöpf, V., Langs, G.: Modeling fetal cortical expansion using graph-regularized Gompertz models. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9900, pp. 247–254. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46720-7_29

12. Sepulcre, J., Liu, H., Talukdar, T., Martincorena, I., Yeo, B.T., Buckner, R.L.: The organization of local and distant functional connectivity in the human brain. PLoS Comput. Biol. **6**(6), e1000808 (2010)

13. Thomason, M.E., et al.: Intrinsic functional brain architecture derived from graph theoretical analysis in the human fetus. PLoS ONE **9**(5), e94423 (2014)

14. Thomason, M.E., et al.: Prenatal lead exposure impacts cross-hemispheric and long-range connectivity in the human fetal brain. NeuroImage **191**, 186–192 (2019)

15. Thomason, M.E., et al.: Weak functional connectivity in the human fetal brain prior to preterm birth. Sci. Rep. **7**, 39286 (2017)

16. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., et al.: Image quality assess-ment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)

17. Wheelock, M., et al.: Sex differences in functional connectivity during fetal brain development. Dev. Cognitive Neurosci. **36**, 100632 (2019)

18. Yeo, B.T., et al.: The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. **106**(3), 1125–1165 (2011)

19. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging **20**(1), 45–57 (2001)