VirtualBrainCloud

Personalized Recommendations for Neurodegenerative Disease

www.VirtualBrainCloud-2020.eu

Public deliverable report

D3.1: Initial version of full-scope containerized pipeline developed

Date	November 2020
Authors	Institute of Neuroscience and Medicine (Brain and Behaviour, INM-7) Forschungzentrum Jülich: Kyesam Jung, Shufei Zhang, Oleksandr V. Popovych, Michael Hanke, Simon B. Eickhoff
Dissemination level	public
Website	www.VirtualBrainCloud-2020.eu

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under grant agreement No 826421

Table of content

1. Introduction

Any analysis of empirical neuroimaging data is preceded by a rather complicated procedure of data processing, where the useful signal contained in raw data has first to be cleaned from numerous measurement and motion artifacts, confounds and noise, and then extracted for investigation (Salimi-Khorshidi et al. 2014; Ciric et al. 2017). However, the lack of ground truth and golden standards for the data processing caused intense debates in the literature as to which algorithms and parameters should provide the most plausible results (Salimi-Khorshidi et al. 2014; Ciric et al. 2017; Qi et al. 2015; Maier-Hein et al. 2017; Lindquist 2020). Hence, the modern neuroimaging data processing remains at the level of the best practice, which requires further investigations, especially, when the empirical data is used for derivation and validation of the whole-brain dynamical models (Honey et al. 2009).

The crucial component for the data-driven and model-based investigation of the human brain activity and connectivity is an appropriate extraction of the structural connectivity (SC) approximating the axonal bundles in the brain and used to build an underlying network of the structural connectome (Sporns, Tononi, and Kötter 2005; Honey et al. 2009). One of the widely used tools to provide a macro-level human structural connectome *in vivo* is a whole-brain tractography (WBT) based on diffusion-weighted imaging (DWI) (Sporns, Tononi, and Kötter 2005). The processing of the data from magnetic resonance imaging (MRI) corresponding to this analysis type involves many steps and approaches, and numerous software tools exist for the processing DWI, tractography and also functional MRI (Soares et al. 2013; Schirner et al. 2015; Esteban et al. 2019). Consequently, effects of different methods of neuroimaging data processing and brain parcellations on structural architecture have been reported (Bassett et al. 2011; Qi et al. 2015). Another study also demonstrated that processing the same neuroimaging data by different pipelines can affect the obtained results (Lindquist 2020). Therefore, a valid and reproducible data processing is one of the most important issues in human neuroimaging research (Maier-Hein et al. 2017; Parkes et al. 2018). However, solving this problem is a challenge, and addressing it is inevitably connected with evaluation of the impact of pipeline parameters on the extracted empirical data and, consequently, on the modeling results. The same applies to the processing of resting-state functional MRI (fMRI) data (Salimi-Khorshidi et al. 2014; Ciric et al. 2017).

In this report we provide a full-scope pipeline for the processing of DWI and resting-state fMRI human neuroimaging data that takes the raw data and delivers the whole-brain structural and functional connectomes. The pipeline has a modular structure, where every module can be executed independently of the others if the corresponding input data is available, which contributes to the flexibility of pipeline application. Furthermore, the pipeline is distributed in a containerized form, which supports its cross-platform usage on different hardware and software environments ranging from single-core desktops and local clusters to supercomputers (Jülich Supercomputing Centre 2018). The pipeline was optimized for parallel processing of several subjects on multi-thread computational nodes and can be applied to process large datasets, e.g. the Human Connectome Project (HCP) dataset.

The developed pipeline is an open-source collection of shell scripts based on several well-established and freely available software packages that are widely used in the neuroimaging community. The main goal of its development was, on the one hand, to simplify the procedure of the high-quality and state-of-the-art data processing for the end user, where the complexity of the appropriate selection and application of data processing steps is resolved inside of the provided scripts. On the other hand, enough freedom for selection of parameters and algorithms of the data cleaning and signal extraction should

nevertheless be preserved such that unrestricted investigation of their impact on the derived results remains possible, in particular, by means of model-based approaches (Popovych et al. 2019).

2. Partners involved

This deliverable was prepared by the Institute of Neuroscience and Medicine (Brain and Behaviour, INM-7) from the Forschungszentrum Jülich (FZJ). The computational resources were granted through JARA on the JURECA computational cluster at FZJ (Jülich Supercomputing Centre 2018). The containerized pipeline used for the neuroimage processing were optimized on the CPU partitions of JURECA in collaboration with the Jülich Supercomputing Centre (JSC) of FZJ.

3. Description of work performed

The developed pipeline consists of two parts. It was designed to (1) preprocess the raw T1-weighted (T1w) and diffusion-weighted MRI (dwMRI) data, extract the whole-brain tractography (WBT) and calculate the structural connectivity (SC); and (2) preprocess the raw T1w and fMRI data, normalize it to the standard MNI152 space and extract the blood-oxygenation level dependent (BOLD) signals and corresponding functional connectivity (FC). The two pipeline parts can be referred to as dwMRI and fMRI pipelines, respectively. In the pipeline scripts some functions of the following third-party software packages were used:

- FSL Version 6.0 (Smith et al. 2004); FSL software license (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Licence); Web: http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/.
- MRtrix Version 3.0 (RC3) (Tournier et al. 2019); Mozilla Public License (http://mozilla.org/MPL/2.0/); Web: http://www.mrtrix.org/.
- ANTs Version 3.0.0.0.dev64-geecaf (Tustison et al. 2010); ANTs Licence (https://github.com/stnava/ANTs/blob/master/COPYING.txt); Web: https://github.com/stnava/ANTs.
- FreeSurfer Version 6.0.0 (Dale, Fischl, and Sereno 1999); The open source FreeSurfer license (https://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferSoftwareLicense); Web: http://freesurfer.net/fswiki.
- AFNI Version Debian-18.0.05+git24-gb25b21054~dfsg.1 -1~nd90+1 (Cox 1996); AFNI license (https://afni.nimh.nih.gov/Legal_info); Web: https://afni.nimh.nih.gov.
- Connectome Workbench Version: 1.4.2 (Glasser et al. 2013); Connectome Workbench license (https://www.humanconnectome.org/software/connectome-workbench-license); Web: https://www.humanconnectome.org/software/connectome-workbench.

3.1. Processing of dwMRI data

The dwMRI pipeline consists of 4 modules: preprocessing, tractography, atlas transformation, and SC reconstruction as illustrated in Figure 1. The modules can be executed either serially and automatically one after another if a complete processing of the subject data is necessary, or independently of each other if some of the processing tasks were already completed, and the corresponding data is available as an input for a given module. For example, if WBT for a given subject is available, the extraction of SC

for another brain parcellation can be performed only by involving modules 3 and 4 which require much less computational resources than module 2.

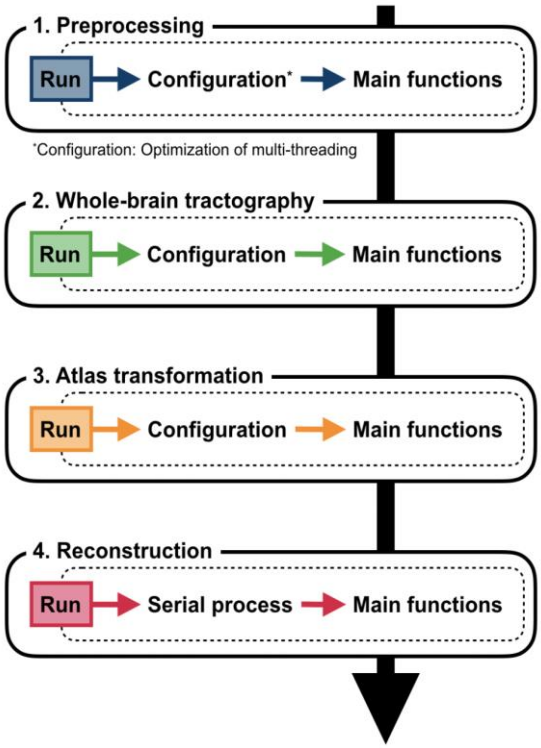

Figure 1: Workflow of the modular structure of the dwMRI pipeline consisting of four modules which can be executed either serially one after another or separately of each other.

In more detail, the mentioned modules of the dwMRI data processing are as follows.

1. The preprocessing module performs image processing for T1w and DWI such as intensity normalization, denoising, motion correction with calculation of diffusion tensor image (DTI), registration, and tissue segmentation.
2. The tractography module estimates fibre oriented distribution (FOD) based on spherical deconvolution derived from the processed DWI (Jeurissen et al. 2014). After the FOD estimation, streamlines are generated by a tracking algorithm with various parameters, for instance, total number of streamlines (Figure 3).
3. The atlas transformation module uses the registration results in the preprocessing module for atlas transformation from the MNI standard space (6th generation in FSL) to a native DWI space (Figure 4).
4. The reconstruction module extracts SC and an average path-length (PL) of WBT streamlines for each region-to-region connection based on the transformed atlas (Figure 5).

Additional details on the structure of the dwMRI pipeline are presented in Table 1, where the main steps of the data preprocessing are indicated in the order of their execution (left column) together with the used software packages mentioned above (middle column) and the corresponding functions (right column).

Table 1: Main steps of the data processing, utilized software packages and executed functions of the dwMRI pipeline.

Preprocessing module	Software	Function
Process T1w image	Freesurfer	recon-all
De-noise diffusion-weighted images (DWI)	MRtrix3	dwidenoise
Bias-field correction for DWI	MRtrix3 and ANTs	dwibiascorrect
Motion-correction and DTI calculation	Freesurfer	dt_recon
Co-registration and registration	FSL	flirt, fnirt, applywarp
Generate a cortical mask	FSL and Freesurfer	fslmaths, applywarp, mri_convert
Tractography module	**Software**	**Function**
Estimation of fibre oriented distributions	MRtrix3	dwi2response, dwi2fod
Tracking streamlines	MRtrix3	tckgen
Atlas transformation module	**Software**	**Function**
Atlas transformation	FSL and Freesurfer	fslmaths, applywarp, mri_convert
Reconstruction module	**Software**	**Function**
Reconstruction of connectivity matrices of SC and PL	MRtrix3	tck2connectome

For the end users several important parameters of the dwMRI pipeline are left open for modification in the input file of the pipeline including selection of the brain atlas that is supposed to be in the 1 mm iso-cubic MNI space (6th generation in FSL), the number of total WBT streamlines (tract), the maximum angle in degrees between successive steps (tckgen_angle), and the minimum and the maximum length of streamlines. The default values are indicated in the supplied example of input parameters. There are other available options for the image processing, tracking algorithms, and reconstructions that can be modified at the pipeline execution. This provides freedom for configuration of the data processing steps and parameters that can be used to investigate their impact on the processed data and extracted signals.

3.2. Processing of the resting-state fMRI data

The fMRI pipeline also contains 4 modules designed to process the raw structural images, perform a minimal and enhanced processing of raw fMRI data and signal extraction as illustrated in Figure 2, discussed in detail below and listed in Table 2.

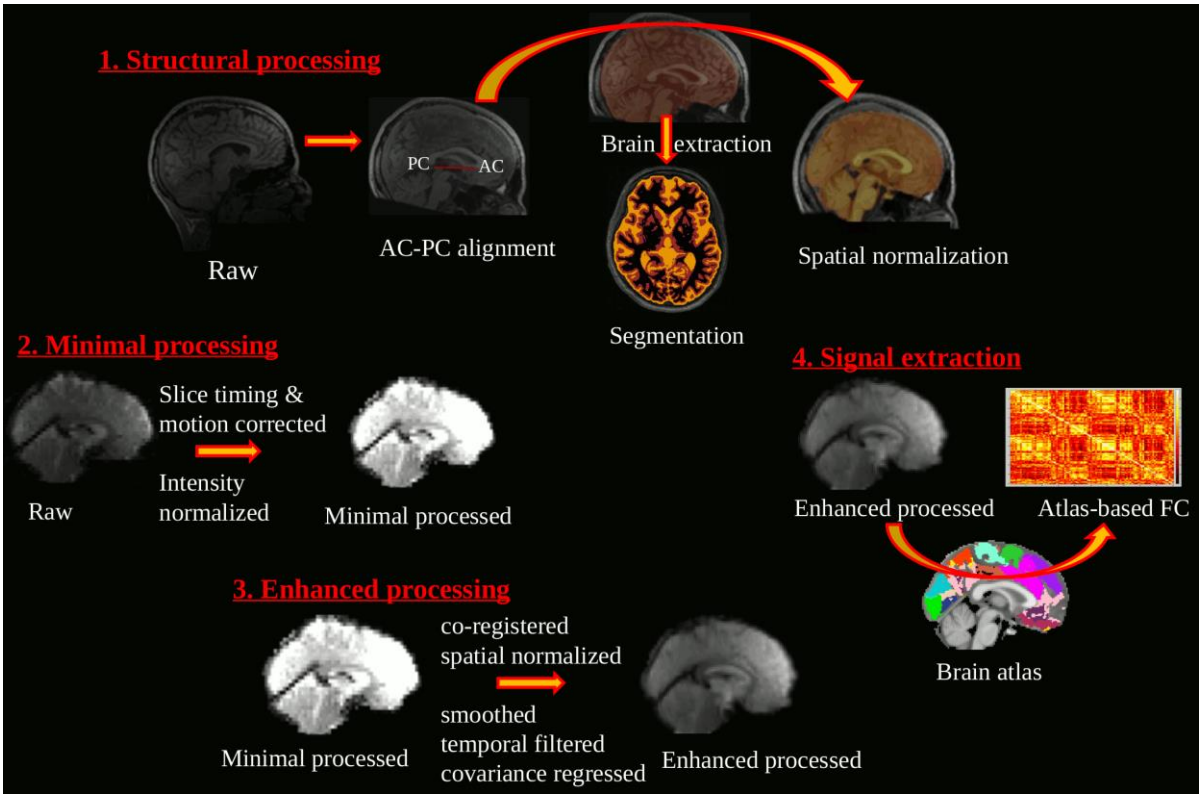

Figure 2: Workflow of the modular structure of the fMRI pipeline consisting of four modules which can be executed either serially one after another or separately from each other.

1. The structural processing module processes the raw structural T1w image (also T2w if available) by performing ACPC alignment, brain extraction, bias correction, segmentation and spatial normalization to the standard MNI152 template.
2. The minimal processing module processes the raw resting-state fMRI images by first excluding a few initial volumes from the processed data and then performing slice-timing and motion corrections and intensity normalization. In addition, 24-parameter motion regressors are calculated.
3. The enhanced processing module includes additional steps for more sophisticated fMRI processing such as co-registration with the T1w image and two-step spatial normalization to the MNI152 standard template, smoothing, temporal filtering and covariance regression by using the calculated 24 motion regressors and global whole-brain and/or tissue signals.
4. The signal extraction module calculates the BOLD signals and functional connectivity for any brain parcellation provided by the user in volumetric form in MNI space.

Additional details on the structure of the fMRI pipeline are presented in Table 2 in the same format as for the dwMRI pipeline (Table 1).

Table 2: Main steps of the data processing, utilized software packages and executed functions of the fMRI pipeline.

Structural processing module	Software	Function
Anatomical average	FSL	robustfov, midtrans,fslmaths

ACPC-alignment	FSL	flirt/aff2rigid
Brain-extraction	ANTs/FSL	fnirt-based brain extraction, antsBrainExtraction, bet
Bias-correction	FSL	fsl_anat, epi_reg (if T2w used)
Segmentation	FSL	fast
Spatial normalization	FSL/ANTs	flirt, fnirt, antsRegistration
Minimal processing module	**Software**	**Function**
Volume exclusion	FSL	fslroi
Slice-timing correction	FSL	slicetimer
Motion-correction	FSL	mcflirt
24-parameters head motion calculation	AFNI	1deval
Intensity normalization	FSL	fslmaths
Enhanced processing module	**Software**	**Function**
Co-registration/spatial normalization (MNI)	FSL/ANTs	flirt, applywarp, antsRegistration, antsApplyTransform
Smoothing	Workbench	wb_command –volume-smoothing
Temporal filtering	FSL	fslmaths
Covariance regression	FSL	fsl_glm
Signal extraction module	**Software**	**Function**
Atlas-based BOLD extraction	FSL	fslmeants
Seed-based functional connectivity	AFNI	3dNetCorr

As for the dwMRI pipeline, several important parameters of the fMRI pipeline can also be modified by the end user including selection of the brain atlas or segmentation that is supposed to be in the 2 mm iso-cubic MNI space (6th generation in FSL), possible inclusion of T2w images in the processing workflow, methods for the brain extraction, and methods for spatial normalization. For the enhanced processing, the parameters for smoothing, temporal filtering and covariance regression can be defined by the user. This provides freedom for configuration of the data processing steps and parameters sufficient for investigation of their impact on the processed data and extracted signals.

3.3. Singularity containers

The two discussed pipelines for dwMRI and fMRI processing were containerized into a singularity function (https://sylabs.io). The software packages mentioned above were embodied into the two separate containers together with the corresponding pipeline scripts used for processing (1) dwMRI and (2) fMRI data, respectively. The containerized pipeline was designed, tested and optimized for execution on single-core desktops or notebooks, high-throughput computing local clusters and high-performance computing systems, in particular, on the supercomputer JURECA at JSC of FZJ (Jülich Supercomputing Centre 2018) which uses the Slurm workload manager (https://slurm.schedmd.com/). The corresponding batch submission scripts with an optimized configuration are also included. Together with the containers, examples of the input files containing the values of pipeline parameters and environment variables as well as the shell scripts used for execution of the processing scripts within the pipeline containers are provided. Detailed instructions explaining the steps necessary for initial setting and start of the pipelines are also included as Read.me files.

4. Results

In this section we present the results of the application of the discussed dwMRI and fMRI pipelines to processing of the neuroimaging data. We illustrate how the human structural and the resting-state functional connectomes can be extracted by the developed pipelines when some parameters of the data processing vary. The obtained results were compared with those of other pipelines available from the literature. We also illustrate the scalability of the pipelines on the JURECA supercomputer at FZJ (Jülich Supercomputing Centre 2018).

4.1. Structural connectivity pipeline

The dwMRI pipeline receives T1w image, dwMRI data, b-vectors and b-values corresponding to the dwMRI of a given subject as input and calculates the WBT and SC matrices as main output. The latter include two connectivity matrices of streamline counts (also denoted as SC) and average path lengths (PL) consisting of the number of WBT streamlines connecting any two brain regions from a given brain parcellation and the corresponding average streamline lengths, respectively. Examples of WBTs calculated by the current pipeline for 10K and 100K total number of streamlines used as a pipeline parameter for WBT extraction are illustrated in Figure 3, where the difference in the WBT density between the two conditions are evident.

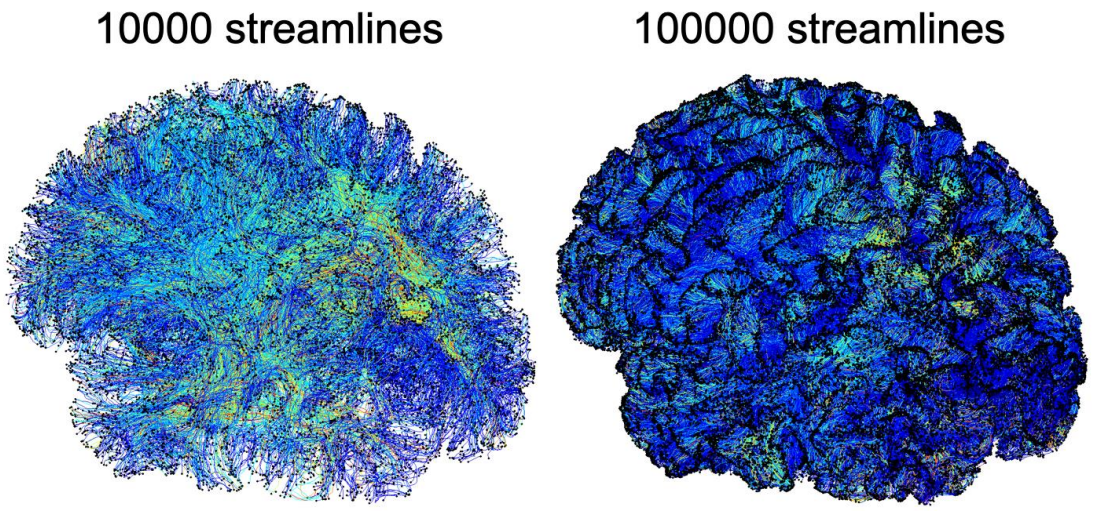

Figure 3: Whole-brain tractography (WBT) with different total numbers of streamlines as indicated in the plots. The thin lines depict the individual streamlines, and the color reflects their length ranging from blue (short streamlines) to red (long streamlines). The black dots indicate starting or ending points of streamlines, which are on the boundary between the cortical gray matter and the white matter.

The WBT density influences the properties of SC and PL matrices that can be extracted by compressing the calculated WBT into inter-region connectivity based on a given brain parcellation/atlas. An important step in this procedure is a conversion of the considered brain atlas from the standard MNI space to a native diffusion-weighted image space of WBT, which is performed in module 3 (Figure 1). The results of such a conversion are illustrated in Figure 4 for the Schaefer atlas with 100 cortical parcels (Schaefer et al. 2018) and the Harvard-Oxford atlas with 96 cortical regions thresholded by 0.25 tissue probability (Desikan et al. 2006). Then the structural connectome as represented by the matrices SC and PL is extracted by projecting the WBT (Figure 3) onto the converted atlas and compressing all whole-brain streamlines into the groups of inter-regional connections as illustrated in Figure 5 for the two examples of WBTs (Figure 3). As expected, the WBT density is clearly reflected in the density of the SC and PL matrices.

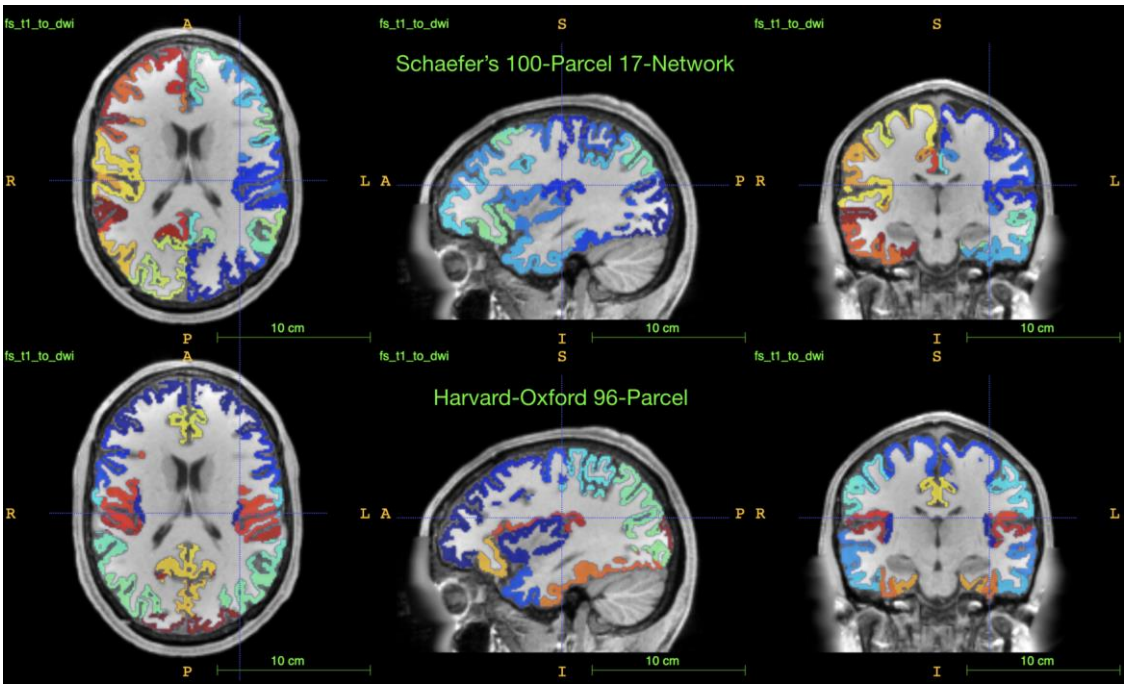

Figure 4: Results of atlas transformation from the MNI standard space (6th generation in FSL) to a native diffusion-weighted image space for the Schaefer 100-Parcel atlas and the Harvard-Oxford 96 cortical regions thresholded by 0.25 tissue probability. Different brain regions parcellated according to the two atlases are indicated by different colors.

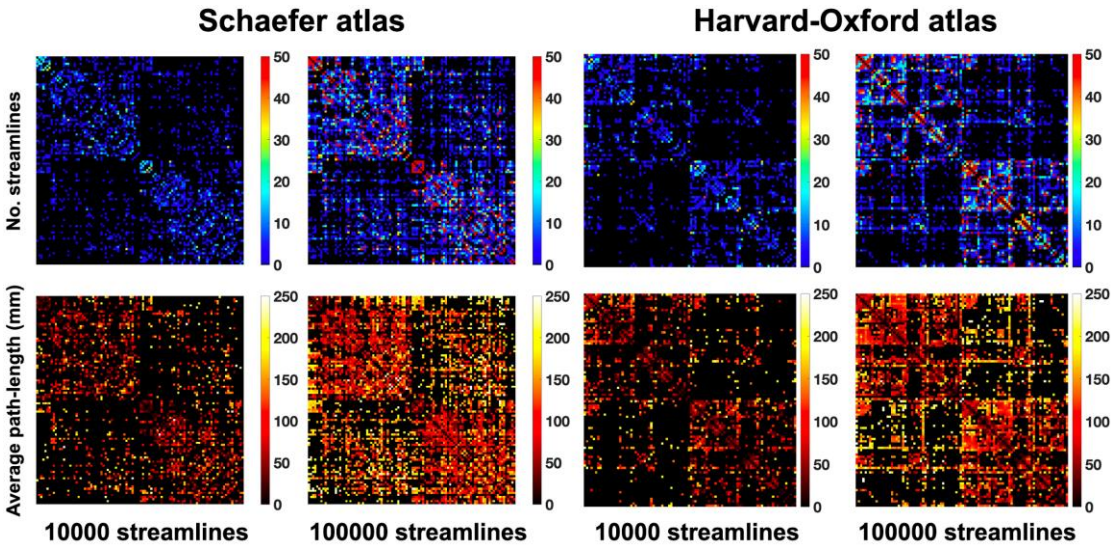

Figure 5: Results of reconstruction of SC matrices (upper row) and PL matrices (lower row) from 10K and 100K whole-brain streamlines of WBTs illustrated in Figure 3 as indicated under the plots for the Schaefer 100-Parcel atlas (two left columns) and the Harvard-Oxford 96 cortical regions thresholded by 0.25 tissue probability (two right columns). Every dot in the plots stands for the corresponding inter-regional connection, and the color reflects the number of inter-region streamlines for SC matrices and average streamline length for PL matrices with the scales on the side color bars.

By comparing the SC and PL matrices extracted for different WBT densities we evaluate the impact of the number of WBT streamlines on the properties of the structural connectome. Figure 6 shows results of such a comparison, where the similarity (Pearson correlation) between SC and PL matrices calculated for varying numbers of total streamlines and those calculated for WBT with 10M streamlines is illustrated. SC matrices with a large number of streamlines are similar to each other stable until 50K streamlines (Figure 6, left bars) However, PL matrices show low similarities across different WBT densities starting already from 2M streamlines (Figure 6, right bars).

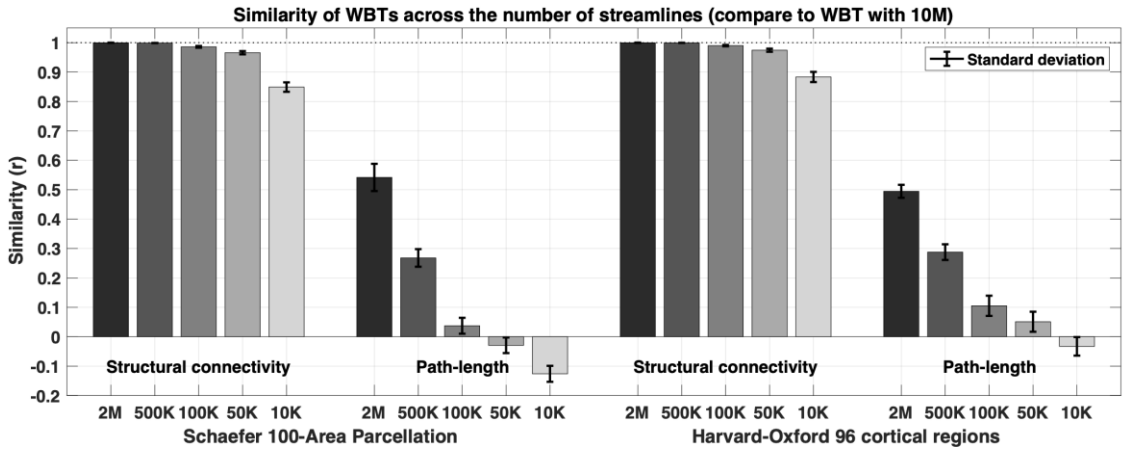

Figure 6: Similarity between the SC and PL matrices extracted from WBT with different numbers of total streamlines indicated on the horizontal axes and those calculated for 10M streamlines. The similarity (Pearson correlation) is calculated for the Schaefer and the Harvard-Oxford atlases and averaged over 272 subjects from the HCP dataset. The error bars indicate the standard deviation.

The results of the discussed dwMRI pipeline developed as a collection of shell scripts were compared with the data processing results of a similar pipeline developed with the employment of Matlab, where the function of the statistical parametric mapping (SPM) toolbox (https://www.fil.ion.ucl.ac.uk/spm/) for the bias field correction of T1w images was used. Otherwise the Matlab scripts and the shell scripts are constituted with the same back-bone of the current dwMRI pipeline. In addition, one more pipeline for dwMRI processing of the paper (Bajada, Schreiber, and Caspers 2019) was involved in the comparison. The reconstructed SC and PL matrices calculated for 23 HCP subjects from a minimally preprocessed diffusion-weighted HCP data (Glasser et al. 2013) by the three mentioned pipelines were pairwise compared with each other. All three pipelines demonstrate high similarity (Pearson correlation close to 0.9 or larger) when comparing SC matrices (Figure 7, left). Somewhat lower similarities of approximately 0.7 correlation were observed for PL matrices (Figure 7, right) which show higher vulnerability as expected from the results depicted in Figure 6.

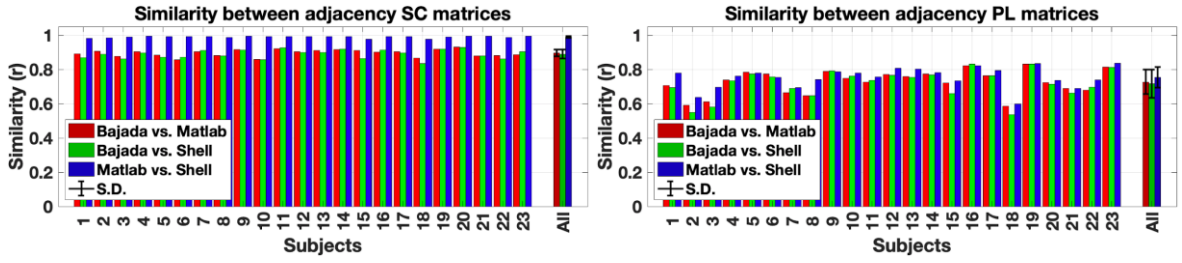

Figure 7: Similarity (Pearson correlation) between SC matrices (left plot) and PL matrices (right plot) extracted by different dwMRI pipelines for 23 HCP subjects based on shell scripts (current pipeline), Matlab and that from the paper (Bajada, Schreiber, and Caspers 2019) as indicated in the legends. S.D. denotes standard deviation.

The containerized dwMRI pipeline was tested and optimized by using the HCP and the enhanced Nathan Kline Institute-Rockland Sample (eNKI, Usage agreement: http://www.nitrc.org/include/glossary.php#546) data. In particular, the pipeline was optimized for execution on the JURECA in JSC (FZJ). Computation time of the preprocessing and the atlas transformation modules was obtained by increasing the number of subjects to measure speed-ups, because functions of FSL do not support multi-threading. The tractography and the reconstruction modules were designed for multi-threading, therefore, computation time of them was obtained by

increasing the number of threads to calculate speed-ups. Figure 8 shows the calculated speed-ups across the number of subjects (for the preprocessing and the atlas transformation modules) and the number of threads (for the tractography and the reconstruction modules). As a result, the optimal number of HCP subjects suggested to be simultaneously processed on the same JURECA node was found to be 8 for the preprocessing module and 12 for the atlas transformation module limited by the amount of working memory of the computational node. The optimal number of threads per subject for the tractography module is 48 which can be used for the reconstruction module as well. Consequently, the optimal configuration of the dwMRI pipeline on JURECA is 8 subjects per node for the preprocessing and the atlas transformation modules and 48 threads per subject for the tractography and the reconstruction modules.

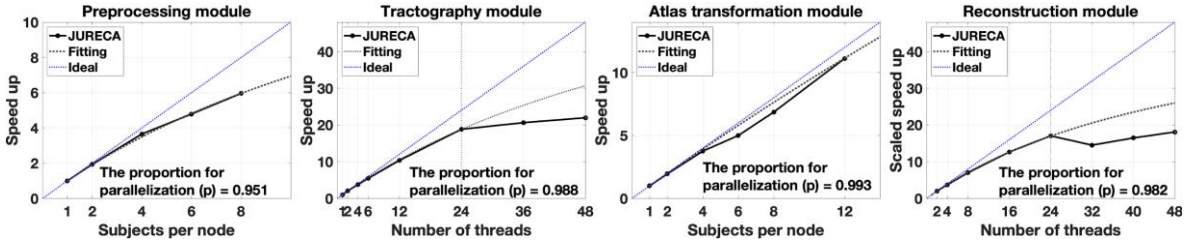

Figure 8: Results of the scaling optimization and speed-ups of the dwMRI pipeline for HCP subjects on the JURECA. The optimization was performed separately for each pipeline module indicated in the titles of the plots. The proportion of parallelization was calculated by Amdahl's law and indicated in the plots, where the curves reflecting the computation time on JURECA, fitting to Amdahl's law and ideal speedup are shown as indicated in the legends.

4.2. Functional connectivity pipeline

The developed fMRI pipeline was used to process the raw resting-state fMRI data and extract of BOLD signals and FC for subjects from the 1000BRAINS cohort (Caspers et al. 2014). Examples of FC matrices for one subject are illustrated in Figure 9, which were calculated using the current fMRI pipeline (Figure 9A) and two other well-known pipelines that are provided by the software packages FSL (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki) (Figure 9B) and SPM (https://www.fil.ion.ucl.ac.uk/spm/) (Figure 9C). In such a way we can compare the performance of the pipeline presented here with the state-of-the-art. The similarity between the obtained FC matrices Figure 9 was quantified by calculating pairwise similarity between the matrices as given by the Pearson correlation coefficient as illustrated in Figure 10 for 30 subjects. We found a relatively high correspondence between the patterns of the resting-state FC for all subjects when comparing the current pipeline with FSL pipeline (Figure 10A) and SPM pipeline (Figure 10B). These similarities are comparable to the correlations between results from the pipelines of FSL and SPM (Figure 10C). It can further be strengthened if the same data processing steps were performed in each of the compared pipelines as illustrated in Figure 10D for the comparison between the current pipeline and FSL.

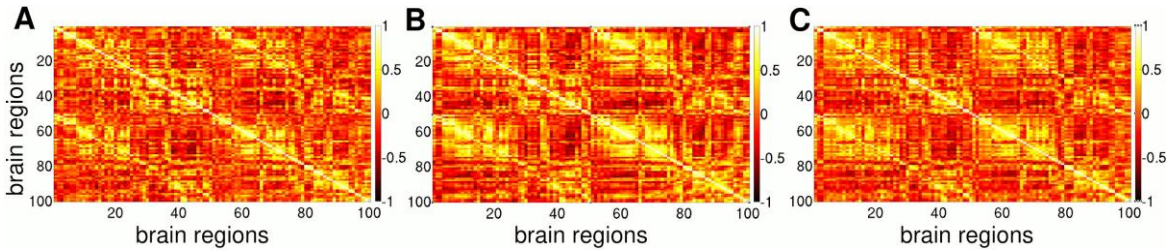

Figure 9: Examples of the resting-state FC calculated for the same subject by (A) current fMRI pipeline, (B) FSL and (C) SPM resting-state pre-processing pipelines in default settings. The Schaefer atlas with 100 cortical parcels was used for the brain parcellation. The Pearson correlation coefficients between the BOLD signals of the brain regions indicated on the axes are depicted by color ranging from -1 (black) to 1 (white).

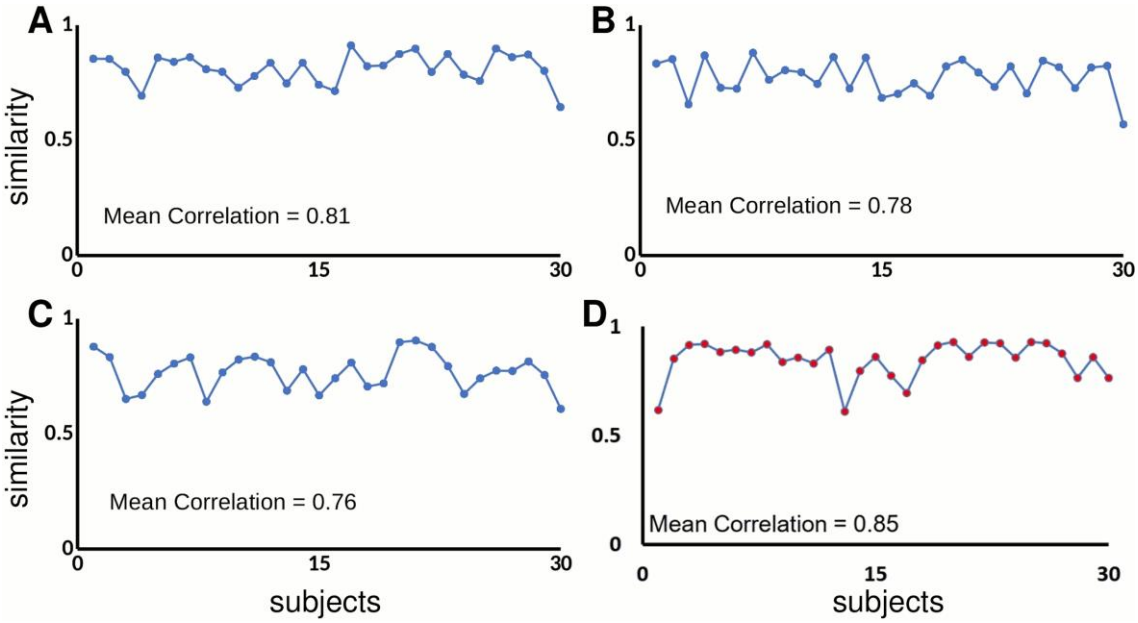

Figure 10: Similarity (Pearson correlation) of the resting-state FC patterns compared between the following two pipelines used of their calculation: (A) Current pipeline and FSL, (B) current pipeline and SPM, (C) FLS and SPM, and (D) current pipeline and FSL with the same steps of data processing. The mean correlation across subjects is also indicated, and the parcellation as in Figure 9.

The quality of the pre-processed resting-state fMRI data can be evaluated by calculating DVARS measures for diagnosis of bad scans, i.e., time points with an enhanced variability as compared to the rest of the data (Afyouni and Nichols 2018). Examples of the time courses of one of such variables are shown in Figure 11, where $\Delta\%D - \mathrm{var}$ reflects an excess variance shown as a spike as a percentage of average variance in the entire data. This variable can be used as a standardized measure for the "practical significance" of the spikes that are still needed to be cleaned up by data processing methods (Afyouni and Nichols 2018). Such significant spikes/scans are marked by red stripes in Figure 11. In the illustrated examples, the current fMRI pipeline performed relatively well (Figure 11A) and exhibited less spikes as compared to FSL pipeline (Figure 11B) or SPM pipeline (Figure 11C). This result was also confirmed when averaging over 30 subjects of 1000BRAINS cohort. The resting-state fMRI data processed by the current pipeline contained on average 5.4 significant scans, whereas the FSL and SPM pipelines missed in average 7 and 10.3 significant spikes, respectively.

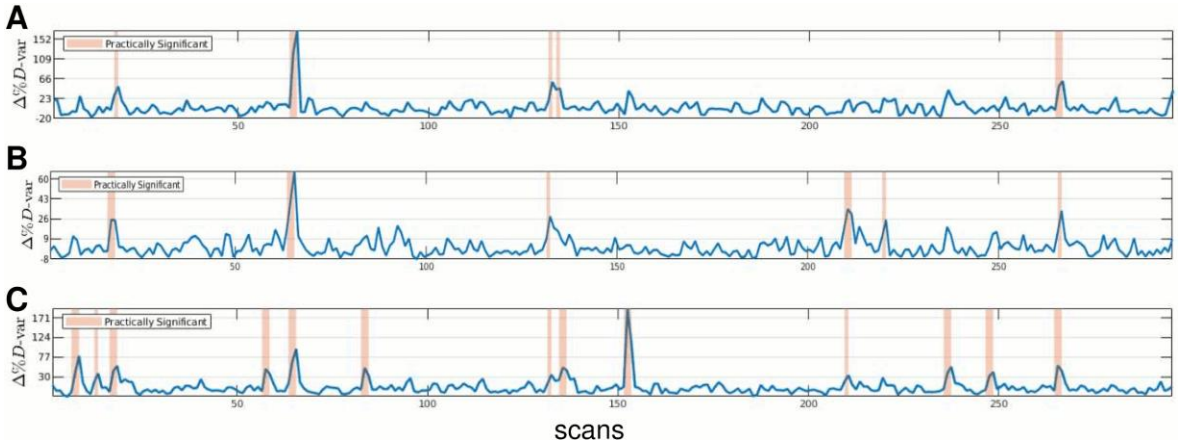

Figure 11: Time courses of DVARS $\Delta\%D - \mathrm{var}$ reflecting an excess fast variance as percentile of the average variance of the resting-state fMRI data (Afyouni and Nichols 2018) pre-processed for the same subject by (A) current pipeline, (B) FSL pipeline and (C) SPM pipeline. The fMRI scans with significant variance are marked by red stripes.

The containerized fMRI pipeline was tested and optimized by using the 1000BRAINS data (Caspers et al. 2014) for deployment on the JURECA supercomputer (Jülich Supercomputing Centre 2018) in JSC (FZJ). Computation time of the four modules (Figure 2) was optimized by increasing the number of subjects for simultaneous processing on the same computation node to measure speed-ups, because the functions of FSL used in all modules (Table 2) do not support multi-threading. The results are presented in Figure 12arallelization is also obtained by fitting the obtained computation time to Amdahl's law. The parallelization is high for the modules of the minimal processing and signal extraction, however, the speed-up is less pronounced for the modules of structural and enhanced processing. Nevertheless, we found that the optimal number of subjects suggested to be simultaneously processed on the same JURECA node was found to be 24 for all modules limited by the amount of working memory of the computational node.

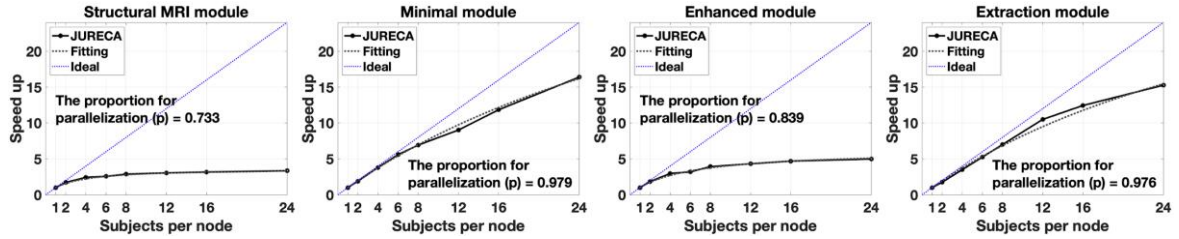

Figure 12: Results of the scaling optimization and speed-ups of the fMRI pipeline for 1000BRAINS subjects on JURECA in JSC (FZJ). The optimization was performed separately for each pipeline module indicated in the titles of the plots. The proportion of parallelization was calculated by Amdahl's law and indicated in the plots, where the curves reflecting the computation time on JURECA, fitting to Amdahl's law and ideal speedup are shown as indicated in the legends.

5. Conclusion, next steps

In this deliverable we provided the full-scope containerized pipeline for processing of dwMRI and resting-state fMRI data with the optimized configuration on the HPC environment (JURECA supercomputer at FZJ). The pipeline is a set of open-source shell scripts based on well-established freely available software packages. The modular structure of the pipeline allows for a flexible pipeline application for processing of the neuroimaging data, its re-usage and recalculation for other values of the pipeline parameters separately for every module. The pipeline parameters can be varied at the moment of pipeline execution by the end user. The designed freedom with modification of several

important parameters of the data processing can facilitate the pipeline application for scientific purposes. In addition, the containerized form of the pipeline distribution supports its cross-platform usage that was tested for an optimal deployment on single-core desktops and notebooks and on supercomputers with parallelized processing of many subjects. We demonstrated that the developed pipeline can effectively process and denoise the neuroimaging data and extract the corresponding signals. We showed that the processing results are comparable with those from the other well-known pipelines available in the neuroimaging community.

The next steps may include further optimization of the fMRI pipeline for parallelized deployment, extending its application spectrum to the task-based fMRI data with the first- and second-level analyses, incorporation of further denoising methods based on, e.g., ICA+FIX and AROMA algorithms and surface-based approaches. Both pipelines can be adapted for individualized data processing, where the whole-brain connectomes will be generated based on individual images in native spaces.

6. References

Afyouni, S., and T. E. Nichols. 2018. 'Insight and Inference for DVARS'. Neuroimage 172 (May): 291–312. https://doi.org/10.1016/j.neuroimage.2017.12.098.

Bajada, Claude J., Jan Schreiber, and Svenja Caspers. 2019. 'Fiber Length Profiling: A Novel Approach to Structural Brain Organization'. NeuroImage 186 (February): 164–73. https://doi.org/10.1016/j.neuroimage.2018.10.070.

Bassett, Danielle S., Jesse A. Brown, Vibhas Deshpande, Jean M. Carlson, and Scott T. Grafton. 2011. 'Conserved and Variable Architecture of Human White Matter Connectivity'. NeuroImage 54 (2): 1262–79. https://doi.org/10.1016/j.neuroimage.2010.09.006.

Caspers, S., S. Moebus, S. Lux, N. Pundt, H. Schutz, T. W. Muhleisen, V. Gras, et al. 2014. 'Studying Variability in Human Brain Aging in a Population-Based German Cohort-Rationale and Design of 1000 BRAINS'. Front. Aging Neurosci. 6 (July): 149. https://doi.org/10.3389/fnagi.2014.00149.

Ciric, R., D. H. Wolf, J. D. Power, D. R. Roalf, G. L. Baum, K. Ruparel, R. T. Shinohara, et al. 2017. 'Benchmarking of Participant-Level Confound Regression Strategies for the Control of Motion Artifact in Studies of Functional Connectivity'. Neuroimage 154 (July): 174–187. https://doi.org/10.1016/j.neuroimage.2017.03.020.

Cox, Robert W. 1996. 'AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages'. Computers and Biomedical Research 29 (3): 162–73. https://doi.org/10.1006/cbmr.1996.0014.

Dale, Anders M, Bruce Fischl, and Martin I Sereno. 1999. 'Cortical Surface-Based Analysis', 16.

Desikan, R. S., F. Segonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, et al. 2006. 'An Automated Labeling System for Subdividing the Human Cerebral Cortex on MRI Scans into Gyral Based Regions of Interest'. Neuroimage 31 (3): 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021.

Esteban, O., C. J. Markiewicz, R. W. Blair, C. A. Moodie, A. I. Isik, A. Erramuzpe, J. D. Kent, et al. 2019. 'FMRIPrep: A Robust Preprocessing Pipeline for Functional MRI'. Nat. Methods 16 (1): 111–+. https://doi.org/10.1038/s41592-018-0235-4.

Glasser, Matthew F., Stamatios N. Sotiropoulos, J. Anthony Wilson, Timothy S. Coalson, Bruce Fischl, Jesper L. Andersson, Junqian Xu, et al. 2013. 'The Minimal Preprocessing Pipelines for the Human Connectome Project'. NeuroImage 80 (October): 105–24. https://doi.org/10.1016/j.neuroimage.2013.04.127.

Honey, C. J., O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli, and P. Hagmann. 2009. 'Predicting Human Resting-State Functional Connectivity from Structural Connectivity'. Proc. Natl. Acad. Sci. U. S. A. 106 (6): 2035–2040. https://doi.org/10.1073/pnas.0811168106.

Jeurissen, Ben, Jacques-Donald Tournier, Thijs Dhollander, Alan Connelly, and Jan Sijbers. 2014. 'Multi-Tissue Constrained Spherical Deconvolution for Improved Analysis of Multi-Shell Diffusion MRI Data'. NeuroImage 103: 411–26. https://doi.org/10.1016/j.neuroimage.2014.07.061.

Jülich Supercomputing Centre. 2018. 'JURECA: Modular Supercomputer at Jülich Supercomputing Centre'. Journal of Large-Scale Research Facilities 4: A132. https://doi.org/10.17815/jlsrf-4-121-1.

Lindquist, Martin. 2020. 'Neuroimaging Results Altered by Varying Analysis Pipelines'. Nature 582 (7810): 36–37. https://doi.org/10.1038/d41586-020-01282-z.

Maier-Hein, Klaus H., Peter F. Neher, Jean-Christophe Houde, Marc-Alexandre Côté, Eleftherios Garyfallidis, Jidan Zhong, Maxime Chamberland, et al. 2017. 'The Challenge of Mapping the Human Connectome Based on Diffusion Tractography'. Nature Communications 8 (1): 1349. https://doi.org/10.1038/s41467-017-01285-x.

Parkes, Linden, Ben Fulcher, Murat Yücel, and Alex Fornito. 2018. 'An Evaluation of the Efficacy, Reliability, and Sensitivity of Motion Correction Strategies for Resting-State Functional MRI'. NeuroImage 171 (May): 415–36. https://doi.org/10.1016/j.neuroimage.2017.12.073.

Popovych, Oleksandr V., Thanos Manos, Felix Hoffstaedter, and Simon B. Eickhoff. 2019. 'What Can Computational Models Contribute to Neuroimaging Data Analytics?' Front. Syst. Neurosci. 12: 68. https://doi.org/10.3389/fnsys.2018.00068.

Qi, Shouliang, Stephan Meesters, Klaas Nicolay, Bart M. ter Haar Romeny, and Pauly Ossenblok. 2015. 'The Influence of Construction Methodology on Structural Brain Network Measures: A Review'. Journal of Neuroscience Methods 253 (September): 170–82. https://doi.org/10.1016/j.jneumeth.2015.06.016.

Salimi-Khorshidi, G., G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith. 2014. 'Automatic Denoising of Functional MRI Data: Combining Independent Component Analysis and Hierarchical Fusion of Classifiers'. Neuroimage 90 (April): 449–468. https://doi.org/10.1016/j.neuroimage.2013.11.046.

Schaefer, A., R. Kong, E. M. Gordon, T. O. Laumann, X. N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. Thomas. 2018. 'Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI'. Cereb. Cortex 28 (9): 3095–3114. https://doi.org/10.1093/cercor/bhx179.

Schirner, M., S. Rothmeier, V. K. Jirsa, A. R. McIntosh, and P. Ritter. 2015. 'An Automated Pipeline for Constructing Personalized Virtual Brains from Multimodal Neuroimaging Data'. Neuroimage 117 (August): 343–357. https://doi.org/10.1016/j.neuroimage.2015.03.055.

Smith, Stephen M., Mark Jenkinson, Mark W. Woolrich, Christian F. Beckmann, Timothy E.J. Behrens, Heidi Johansen-Berg, Peter R. Bannister, et al. 2004. 'Advances in Functional and Structural MR

Image Analysis and Implementation as FSL'. NeuroImage 23 (January): S208–19. https://doi.org/10.1016/j.neuroimage.2004.07.051.

Soares, José M., Paulo Marques, Victor Alves, and Nuno Sousa. 2013. 'A Hitchhiker's Guide to Diffusion Tensor Imaging'. Frontiers in Neuroscience 7. https://doi.org/10.3389/fnins.2013.00031.

Sporns, Olaf, Giulio Tononi, and Rolf Kötter. 2005. 'The Human Connectome: A Structural Description of the Human Brain'. PLoS Computational Biology 1 (4): e42. https://doi.org/10.1371/journal.pcbi.0010042.

Tournier, J-Donald, Robert Smith, David Raffelt, Rami Tabbara, Thijs Dhollander, Maximilian Pietsch, Daan Christiaens, Ben Jeurissen, Chun-Hung Yeh, and Alan Connelly. 2019. 'MRtrix3: A Fast, Flexible and Open Software Framework for Medical Image Processing and Visualisation'. NeuroImage 202 (November): 116137. https://doi.org/10.1016/j.neuroimage.2019.116137.

Tustison, Nicholas J, Brian B Avants, Philip A Cook, Yuanjie Zheng, Alexander Egan, Paul A Yushkevich, and James C Gee. 2010. 'N4ITK: Improved N3 Bias Correction'. IEEE Transactions on Medical Imaging 29 (6): 1310–20. https://doi.org/10.1109/TMI.2010.2046908.