

Mitigating head motion artifact in functional connectivity MRI

Rastko Ciric¹, Adon F. G. Rosen¹, Guray Erus², Matthew Cieslak^{1,2}, Azeez Adebimpe^{1,2}, Philip A. Cook³, Danielle S. Bassett⁴, Christos Davatzikos^{2,3}, Daniel H. Wolf^{1,2}, Theodore D. Satterthwaite^{1,2*}

Participant motion during functional magnetic resonance image (fMRI) acquisition produces spurious signal fluctuations that can confound measures of functional connectivity. Without mitigation, motion artifact can bias statistical inferences about relationships between connectivity and individual differences. To counteract motion artifact, this protocol describes the implementation of a validated, high-performance denoising strategy that combines a set of model features, including physiological signals, motion estimates, and mathematical expansions, to target both widespread and focal effects of subject movement. This protocol can be used to reduce motion-related variance to near zero in studies of functional connectivity, providing up to a 100-fold improvement over minimal-processing approaches in large datasets. Image denoising requires 40 min to 4 h of computing per image, depending on model specifications and data dimensionality. The protocol additionally includes instructions for assessing the performance of a denoising strategy. Associated software implements all denoising and diagnostic procedures, using a combination of established image-processing libraries and the eXtensible Connectivity Pipeline (XCP) software.

Introduction

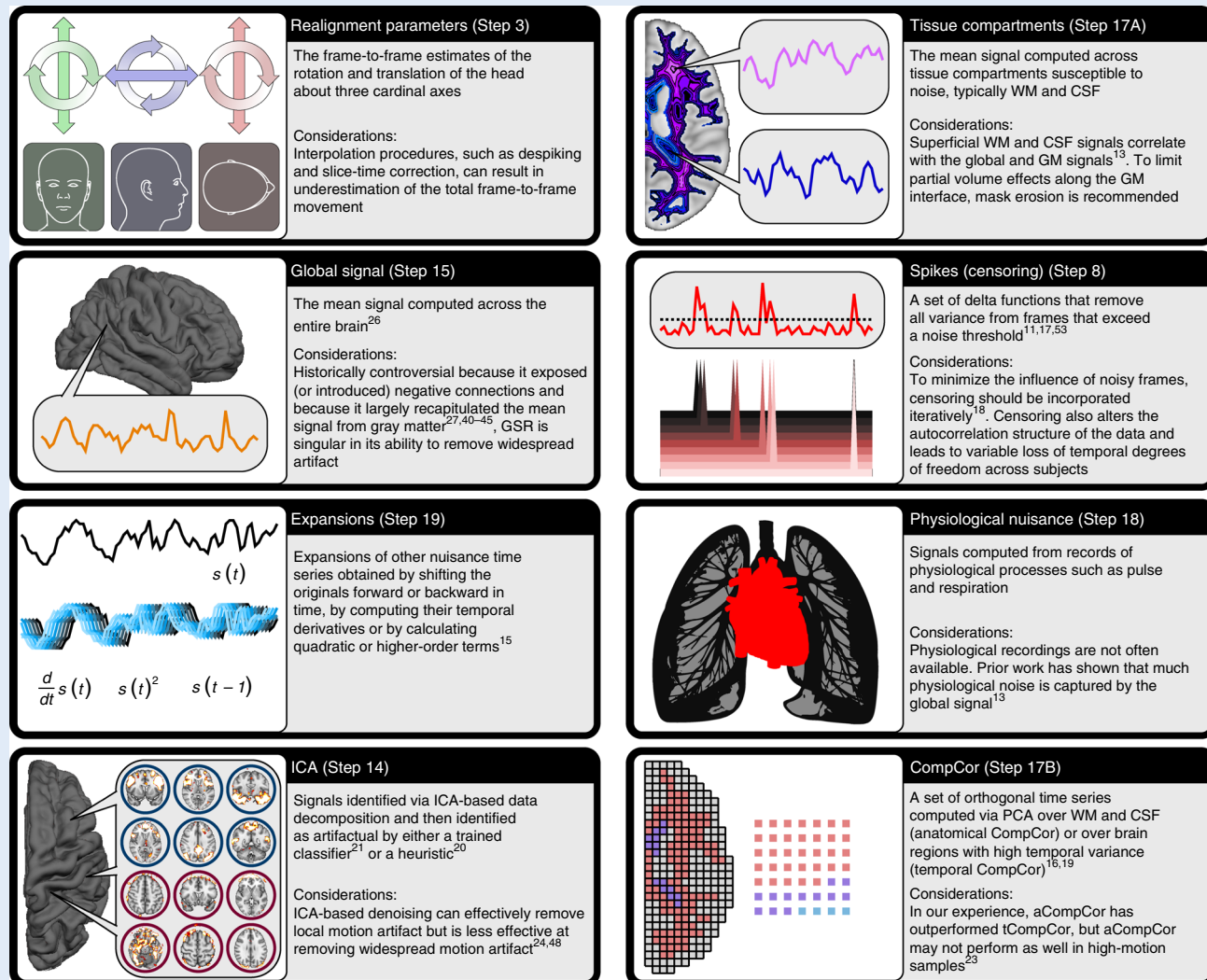
Functional connectivity analysis has yielded important insights into the intrinsic organization of the human brain^{1,2}. Whereas task fMRI examines the functional responses to specific cognitive, perceptual, or motor manipulations, functional connectivity MRI can provide detailed maps of major functional networks without requiring the use of a task^{3–5}. Prior studies have shown that important individual differences, including chronological age across the lifespan, cognitive performance, and psychiatric diagnoses, are associated with differences in functional connectivity^{6–9}. However, many measures of interest in neuroimaging studies are also correlated with the movement of subjects in the scanner. For instance, children and subjects from clinical populations tend to move more when they are scanned than do healthy adults¹⁰. In-scanner motion induces large signal fluctuations in fMRI time series data, which can systematically alter observed patterns of functional connectivity^{10–13}.

If data processing fails to account for motion-related signal variance, artifacts can easily confound inference. For example, initial reports that brain development is associated with strengthening of long-range connections and weakening of short-range connections have been shown to be markedly inflated by the presence of motion artifact in younger children^{8,14}. Accordingly, investigators have developed numerous denoising strategies to mitigate the influence of motion on functional connectivity, and thus disentangle motion effects from effects of interest (Box 1). However, not all of these strategies are equally efficacious. Furthermore, the rapid proliferation of denoising approaches has introduced uncertainty among researchers as to which approach has the greatest utility for their data.

Although the development of denoising strategies will continue into the foreseeable future, an emerging corpus of evidence suggests that the most successful strategies share a number of common elements that target particular features of motion artifact (Box 2). Here, we describe a validated, high-performance protocol for removal of motion artifact from functional MRI data. We also provide a theoretical and historical context for our protocol, its development, and its utility. Finally, we

¹Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA. ²Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, USA. ³Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA. ⁴Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA. *e-mail: sattertt@pennmedicine.upenn.edu

Box 1 | Overview of confound models for motion correction



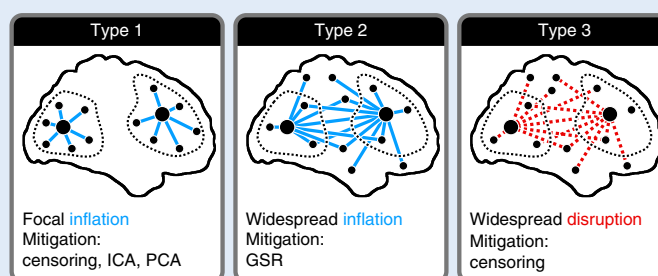
Numerous strategies have been used for removing motion-related variance from a functional MR time series. This box summarizes several denoising options, along with the steps of the protocol at which they can be added to the confound model, and discusses potential pitfalls and limitations of each approach. The design of a confound model suited to a particular dataset and scientific objective is discussed in the 'Experimental design' section.

introduce a software package that implements common denoising protocols and that provides tools for assessing the efficacy of denoising.

Development of the protocol

Although the susceptibility of fMRI to motion has been acknowledged since 1996¹⁵, it was not until 2012 that the influence of motion artifact on functional connectivity was systematically investigated^{10–12}. Three studies published by independent groups not only reported the systematic contamination of functional connectivity measures by motion artifact, but also introduced measures to control for this artifact^{10–12}. By and large, most approaches to denoising functional connectivity data have expanded on a family of strategies that had previously been used to denoise functional activation data: confound regression¹⁵.

Confound regression aims to separate signal from noise by modeling artifactual processes as time series. These time series might include, for instance, frame-to-frame estimates of how much the head has moved, or signals from noise-prone tissue compartments such as the ventricles. These 'artifact'

Box 2 | Taxonomy of motion artifacts in BOLD time series

Motion artifact impacts functional connectivity data in three primary ways. This taxonomy, first introduced by Power and colleagues²⁸, is graphically summarized in the schematic shown. Type 1 effects occur when a movement drives the signal in proximal voxels homogeneously, resulting in spuriously inflated correlations among nearby regions. Type 2 effects, by contrast, occur when a movement globally drives the BOLD signal in a homogeneous manner, inducing widespread inflation of correlations. Finally, Type 3 effects occur when a movement induces heterogeneous signal fluctuations across the brain, disrupting correlations, particularly those between distal regions. In general, Type 2 artifacts are the most common and Type 3 are the least common. Notably, different confound regression strategies are more effective at targeting different features of motion artifact. For instance, GSR models the widespread signal fluctuations that characterize Type 2 artifact. By contrast, strategies that model spatially localized artifactual variance, such as ICA and PCA, may be better suited for cleaning Type 1 artifact. Censoring approaches, which remove entire frames that have been contaminated by motion, remove the spatially varied Type 1 and Type 3 artifacts. (More aggressive censoring can also be effective against Type 2 artifacts¹⁸.)

The contributions of different artifact types can also account for why motion artifact exhibits a distance-dependent spatial profile^{10,11,17}: the connectivity among nearby regions of the brain is inflated by a combination of Type 1 and Type 2 effects, whereas the connectivity among more distant regions is inflated by Type 2 effects (and, to a lesser extent, is disrupted by Type 3 effects). Consequently, motion, on average, inflates short-distance connections more than it does long-distance connections. Because GSR is highly effective at mitigating common Type 2 motion effects, its residual artifact profile is dominated by Type 1 effects; this is observed as elevated distance dependence^{22,28,56}.

time series are fitted to the observed blood oxygen level-dependent (BOLD) time series using a general linear model (the confound model). The residuals of the fit, corresponding to the variance that the confound model cannot explain, can then be analyzed as ‘cleaned’ data, whereas the BOLD variance explained by the confound model can be discarded from further analyses.

Confound regression remains the most prevalent method for removing head motion artifact in functional connectivity MRI. However, the performance of any confound regression strategy will depend on the constituent time series of its confound model. A more effective model will explain more artifactual variance in the data (and thus remove more artifact), whereas a more efficient model will minimize false positives (and thus not mistakenly identify the signal of interest as artifactual in origin). The constitution of the optimal confound model has been hotly debated in the field, with numerous studies proposing a new, ostensibly better model each year^{16–21} (summarized in Box 1).

More recently, several studies have attempted to cope with the proliferation of confound models by establishing a standard set of benchmarks to evaluate their efficacy and efficiency (Tables 1 and 2). These benchmarking studies have yielded remarkably convergent results^{22–25}. Taken together, these studies indicate that models that rely exclusively on estimates of frame-to-frame head movement fail to correct for motion artifact. By contrast, models that remove the time course of the average BOLD signal across the entire brain^{26,27} (‘global signal regression’ (GSR)) and/or time series obtained via signal decomposition techniques^{16,19–21} (principal component analysis (PCA) or independent component analysis (ICA)) exhibit markedly improved performance. Such models can be profitably augmented with temporal censoring operations (e.g., spike regression, scrubbing)^{11,17,18} that specifically remove the influence of volumes corrupted by artifact. Most top-performing strategies combine several techniques in order to mitigate the influence of both local and global features of motion artifact²⁸ (Box 2). Here, we provide a protocol for implementation of a high-performance denoising strategy, while acknowledging that different strategies have specific strengths and weaknesses. Ultimately, the specific choice of confound regression strategy should be dictated by the hypothesis of interest, as discussed in more detail in the ‘Experimental design’ section.

Table 1 | Summary of subject-level quality indices

Subject-level index	Description	Stand-alone implementations
Framewise displacement (FD) ^{11,12,72}	An estimate of the subject's head movement from one frame of the time series to the next	<ul style="list-style-type: none"> FSL: <code>fsl_motion_outliers</code> (FD_{Jenkinson} and FD_{Power}) FSL: <code>mcflirt</code> (FD_{Jenkinson}) XCP: <code>fd.R</code> (FD_{Power})
DVARs ^{11,74,75}	The temporal derivative of the root-mean-square intensity, an index of the frame-to-frame change in signal intensity across the brain	<ul style="list-style-type: none"> FSL: <code>fsl_motion_outliers</code> XCP: <code>dvars</code> (standardized)
Outlier count	An index of the number of outlier values over all voxelwise time series within each frame	AFNI: <code>3dToutcount</code>
'Quality index'	A measure of the dissimilarity of a frame with respect to the median value over all frames	AFNI: <code>3dTqual</code>
FD-DVARs correlation ¹⁹	The correlation between FD and DVARs score; indexes the extent to which signal fluctuations relate to subject movement	XCP: <code>featureCorrelation.R</code>
Voxelwise displacement ¹⁷	An estimate of each voxel's movement between consecutive frames of the time series	—
Spike count (number of superthreshold frames)	The number or percentage of frames in a time series that exceed a motion threshold	—
Loss of temporal degrees of freedom (tDOFs) ^{25,88}	The number of temporal degrees of freedom lost during the denoising process, typically estimated as the sum of the number of nuisance regressors and the number of censored frames	—
Variance explained by nuisance time series and motion-BOLD contrasts ^{23,25,88}	Voxelwise maps or summary values indicating the fraction of variance in the data that can be explained by each term in the confound model, or the regression coefficients of a linear model fitting nuisance time series to BOLD time series	<ul style="list-style-type: none"> AFNI: <code>3dTfitter</code> FSL: <code>fsl_glm</code>
Carpet plot/voxel plot ⁵⁷	A time-by-space matrix containing all values in the time series, often plotted alongside quality index time series such as FD or DVARs	<ul style="list-style-type: none"> <code>plotdemo.m</code> XCP: <code>voxts.R</code>
Network identifiability ^{17,22,56,83,84}	The extent to which subnetwork structure can be resolved in the connectome; can be estimated as the modularity quality Q	—
Test-retest reliability ^{23,25}	An estimate of the replicability of motion and functional connectivity estimates across repeated measures from the same subject	—

Table 2 | Summary of sample-level quality indices

Sample-level index	Description	Stand-alone implementations
QC-FC correlations ^{10,11,22,23,28,56}	Correlations between mean framewise displacement and the functional connectivity between each pair of regions, computed across subjects	XCP: <code>qcfc.R</code>
Network-level QC-FC correlations ^{22,44}	Correlations between mean framewise displacement and network measures (such as modularity), computed across subjects	XCP: <code>qcfc.R</code>
QC-FC distance-dependence ^{11,17,23,28,56}	The second-order correlation between edgewise QC-FC correlations and the Euclidean separation between region pairs	XCP: <code>qcfcDistanceDependence</code>
High- versus low-motion (HLM) contrasts ^{23,88}	Group-level comparisons between matched high- and low-motion subject bins	—
Average tDOFs loss, variability in tDOFs loss ^{22,23,88}	The mean and variance in the loss of tDOFs across subjects	—
Discriminability ^{23,56}	Post-denoising sensitivity to between-group or individual differences	—

Comparison with other methods

Despite its ubiquity, confound regression is not the only means of correcting for head motion artifact in functional connectivity MRI. Two additional families of motion correction approaches that have gained traction are prospective and group-level motion correction strategies. Prospective approaches

operate during data acquisition—for instance, by using behavioral interventions to pre-emptively minimize head movement²⁹, by leveraging MR sequences designed to facilitate separation of signal from noise^{30–32}, or by ensuring that some minimal number of low-noise snapshots of the brain are available for each participant³³. Although minimization of motion during acquisition is critical for all studies, novel techniques for aiding prospective motion correction cannot be applied to the vast majority of datasets that have already been acquired. Furthermore, even in datasets in which prospective approaches have been applied, it is likely that data quality can be further improved via retrospective denoising approaches. For instance, multi-echo acquisition techniques support improved identification and removal of focal motion artifact, but denoising can be optimized by combining the advantages offered by this prospective approach with global artifact removal using retrospective denoising strategies such as GSR³⁴.

By contrast, group-level retrospective approaches are applied after image processing, during group-level analysis. They often take the form of a nuisance covariate that is included in explanatory models alongside variables of interest. Importantly, correction at the group level is not equivalent to participant-level correction³⁵, nor is it mutually exclusive. On the one hand, it can reduce the sensitivity of analyses to any connections that are both related to the variable of interest and susceptible to motion artifact³⁶. For example, including both age and motion in a regression model as part of a study of brain development will by definition remove any shared variance. On the other hand, inclusion of a single motion covariate in a group-level linear model is insufficient to address either nonlinear effects or interactive effects of motion with variables of interest such as age. In summary, although group-level approaches can often supplement participant-level confound regression, they may be inadequate as a stand-alone approach. In general, best practices minimize motion during acquisition, use a high-performance confound regression model during participant-level denoising, and evaluate the impact of motion in group-level statistical models.

Applications of the protocol

Functional connectivity analysis is a routine component of many neuroimaging studies, and a growing number of large studies are organized around the explicit aim of mapping the functional connectome in humans^{2,37}. Any MRI study of the brain's functional connectivity can reasonably be expected to include some form of motion denoising as part of image processing. However, denoising is most critical in cases in which the scientific hypothesis is focused around individual differences, especially age, clinical diagnosis, symptom burden, or cognition. The strengths of the protocol we present have been demonstrated in direct studies of motion artifact^{10,17,35}, in studies of individual differences^{38,39}, and most recently in a pair of head-to-head evaluations of motion correction strategies according to standardized benchmarks^{22,23}. In particular, benchmarking efforts in large datasets have demonstrated that the protocol effectively reduces the relationship between motion and functional connectivity up to a 100-fold relative to minimal-processing strategies^{22,23}. Furthermore, the protocol reduces the distance dependence of motion artifact.

Limitations

The implementation of confound regression strategies has not been without controversy. One of the most heated controversies in functional neuroimaging has concerned the appropriateness of using the global mean signal in confound regression^{27,40–45}. A primary concern raised regarding GSR is the possibility that the mean global signal might have neural origin, or that it could reflect the processes of interest^{46,47}. However, accumulating evidence suggests that the global signal is in fact dominated by signals of nonneural origin: the content of motion-related and respiratory noise in the global signal is demonstrably high across datasets and scanning platforms¹³. Furthermore, a major limitation of alternative approaches based on spatially specific nuisance models (including ICA) is their inability to remove widespread, globally distributed artifact that is the primary result of motion ('Type 2' artifact, in Box 2)^{24,48}.

Although GSR is a highly effective denoising strategy, it is a mathematical necessity that data processed using GSR will include negatively correlated time series, even if no negative correlations were present before the application of GSR^{27,41}. Whereas the distribution of connection strengths in data processed without GSR typically has a positive center and negative skewness, GSR-based processing often results in a distribution that is zero-centered and relatively symmetrical⁴¹. Without artifact-free data available to provide a ground truth, it is unknown whether negative correlations in the connectome reflect a biologically meaningful process or an artifact of processing. However, even

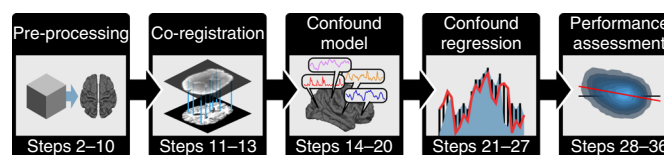


Fig. 1 | Workflow for motion correction of functional connectivity MRI data. The processing stream begins by estimating subject movement and minimally pre-processing the functional data. This minimal pre-processing starts with the image collected from the scanner and produces outputs appropriate for denoising and co-registration; it includes realignment of frames, brain boundary estimation, removal of trends and outliers, and motion estimation (Steps 2–10). Next, co-registration aligns the pre-processed functional MR image to a high-resolution anatomical MR image, supporting the use of signals from specific tissue classes in the confound model (Steps 11–13). After co-registration, the processing stream builds a model of the artifact in the functional time series by combining motion estimates, the global signal, signals from white matter and cerebrospinal fluid, time-series expansions, spike regressors, and time series derived from signal decompositions (Steps 14–20). In confound regression, the confound model is fitted to the pre-processed functional time series. Any variance explained by the confound model is discarded from the data, while the residuals are retained as the denoised functional time series (Steps 21–27). Finally, subject- and group-level diagnostic measures are produced to facilitate assessment of model performance and transparent reporting of denoising efficacy. Performance assessment begins by computing the subject's connectome across regions defined in a functional brain atlas, which requires warping the atlas into the coordinate space of the functional time series. Any variance in the estimated connectome that can be explained by motion is used as a proxy for estimating the residual motion artifact (Steps 28–36).

when data are processed without GSR, lower-motion subjects tend to exhibit a greater number of negative edges and a distribution of connection strengths that is closer to zero-centered in comparison with higher-motion subjects¹⁸. Finally, it should be noted that GSR tends to increase the distance dependence of motion artifact, especially when it is not supplemented with either temporal censoring or spatially specific nuisance regression¹⁷ (Box 2). However, it should be emphasized that this relationship is due to differential denoising efficacy for long-range over short-range connections. Many alternative models that do not include GSR have less distance dependence, but this insensitivity is often due to ineffective performance at connections of all lengths.

In addition to GSR, another highly effective technique is temporal censoring. However, denoising strategies that include temporal censoring can remove a substantial proportion of temporal degrees of freedom (tDOFs) from the data; because of this loss of tDOFs, these strategies are effective but not always efficient (Table 1). As a result, the number of frames censored—and thus the cost in tDOFs—will vary across subjects, leading to the possibility that group differences will be introduced because of the resultant sampling differences. Furthermore, temporal censoring clearly disrupts the temporal dynamics of the time series. However, objections to temporal censoring should be tempered by the observation that the signal being removed in censoring may be of questionable value in the first place. Thus, motion-corrupted data do not have the same tDOFs or temporal dynamics as artifact-free time series. Moreover, investigators can randomly censor noncorrupted data in order to balance the tDOFs between groups, albeit at the cost of losing true observations.

Another open problem in motion mitigation is the quality and nature of frame-to-frame motion estimates. Typically, subject movement is estimated post hoc, by aligning each frame to a reference and using the alignment matrix to estimate that frame's translational and rotational displacement from the reference. Real-time, optical measures provide an alternative option that may better estimate true head motion. Although such measures are under active development, they have yet to be adopted broadly, in part because the requisite technology is at present unavailable at most MRI centers. Furthermore, although the motion estimates obtained via post hoc realignment may be imperfect, prior work indicates convergence between these estimates and more direct optical measures⁴⁹.

Overview of the procedure

In this protocol, we provide a step-by-step guide to implementing a high-performance denoising pipeline (Fig. 1). In addition to the confound regression protocol itself, an equally important dimension of motion artifact correction is transparent reporting of data quality indicators (Tables 1 and 2). Such indicators provide a measure of the artifact initially present in each image and across the entire sample, and also of the residual artifact after images have been denoised. These measures can be leveraged as proxies for denoising performance. Thus, we additionally describe how to quantify and report (i) the presence of artifact in a functional brain image and (ii) the efficacy and efficiency of

Box 3 | The XCP cohort file

A pipeline cohort file defines the experimental sample, the set of subjects that the pipeline should process. In the XCP system, the cohort file is formatted as a .csv file and contains

- A column for each identifier variable
- A column for each pipeline input
- A header naming each category of input
- A row corresponding to each subject

An example is provided below:

```
id0,id1,img,antsct
```

```
ACC,001,rawData/ACC_001_rest.nii.gz,processedData/ACC_001_antsct
```

```
DSQ,001,rawData/DSQ_001_rest.nii.gz,processedData/DSQ_001_antsct
```

```
DSQ,002,rawData/DSQ_001_rest.nii.gz,processedData/DSQ_002_antsct
```

```
CAT,001,rawData/CAT_001_rest.nii.gz,processedData/CAT_001_antsct
```

The header for identifier columns should begin with `id` and end with a nonnegative integer. Identifiers should be ordered hierarchically; for instance, `id0` could correspond to the subject identifier, `id1` to the session identifier, and `id2` to the scan identifier within the session. The primary image to be processed should be listed under the header `img`, and the data directory that directly contains all anatomical inputs (anatomical image, transforms, and segmentation) should be listed under the header `antsct`.

If the data are to be processed using the CBICA IPP, then the paths in the cohort file should not include any directories, as those directories will not be uploaded to IPP. For the archived and compressed anatomical processing directory, the extension (i.e., `.tar.gz`) should not be included in the path that is specified in the cohort.

the motion correction protocol for an individual subject and for the sample as a whole. Code for all stages of the protocol is freely and publicly available online (<https://github.com/PennBBL/xcpEngine>).

Each step of the protocol both describes the objective of that step and includes a specific implementation in code. A variety of software libraries are available for image processing, and in many cases, alternatives to the specific implementation that we provide are available. A description of the general objective of each processing step is included so that investigators can select the best tool tailored to their own specific needs.

The denoising protocol can be implemented step by step by following the procedure outlined here, but we also provide a software package that flexibly implements the protocol in an automated manner. XCP provides a platform for the processing of neuroimaging data that packages common processing steps into configurable modules, and then combines the modules into processing streams, with each stream corresponding to an imaging modality or an analytic objective. Each module uses common tools, enforces consistent output directory structures, and utilizes identical naming conventions. Metadata and provenance of any image derivatives are internally tracked, allowing the pipeline to process each derivative in the most appropriate way.

The XCP system is powered by Analysis of Functional NeuroImages (AFNI), Advanced Normalization Tools (ANTs), and the FMRIB Software Library (FSL)^{50–52} but also introduces multiple stand-alone image-processing utilities. XCP supports processing streams for a number of sequences and analytic modalities, including structural morphometry, task-based fMRI, and arterial-spin labeled MRI. However, in this protocol, we focus on applications to functional connectivity MRI. At present, XCP includes default configurations for three high-performance confound regression strategies: an anatomical CompCor model based on PCA¹⁹, an ICA-AROMA (automatic removal of motion artifacts) model based on ICA²⁰, and a high-parameter model based on noisy tissue compartments and motion estimates¹⁷. All streams include GSR²⁶, and all can optionally incorporate censoring^{11,17,18,53}. A version of each stream that does not incorporate GSR is also available to facilitate assessment of the robustness of results to alternative processing schemes, but we do not recommend these alternatives for most use cases.

When processing data with XCP Engine, a design file and a cohort file are required. The design file parameterizes the processing stream, whereas the cohort file characterizes the input sample. A configuration script and GitHub repository are available for easy configuration of the design file (<https://github.com/PennBBL/xcpConfig>). A brief overview of cohort file specifications is included in Box 3, as is an example. More extensive documentation of the design and cohort files, including examples, format specifications, and instructions for cohort file setup are available online (<https://pipedocs.github.io/config/index.html>). XCP can be run either from a local installation or using the publicly available Image Processing Portal (IPP, <https://ipp.cbica.upenn.edu>) from the Center for Biomedical Image Computing and Analytics (CBICA) at the University of Pennsylvania, which provides the scientific community with the resources of a high-performance computing

cluster for processing data with XCP and also provides automatic access to future extensions of XCP functionality.

Functional connectivity processing generally requires that each participant's anatomical data be processed first, and in the protocol detailed here, we assume that the user has already executed an anatomical pipeline to completion. Outputs of anatomical processing should minimally include (i) a high-resolution anatomical image with nonbrain tissue stripped and with any intensity bias artifact removed; (ii) a segmentation of the anatomical image into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF); and (iii) a set of transforms that map between coordinates in the subject's anatomical image and coordinates in a standard reference space that is the same for all subjects, such as the Montreal Neurological Institute (MNI) template. The segmentation of the structural image is required to isolate nuisance signals from WM and CSF, and it can additionally be used to improve co-registration quality. Transforms to template space (also called 'image normalization' or 'registration') are a prerequisite for generating functionally homologous connectomes across subjects; they are used to move a standard brain parcellation into each subject's coordinate space so that regional signals can be computed and correlations between them can be estimated. XCP requires that normalizations to template space use warps generated by the top-performing registration procedures included in ANTs^{51,54}. ANTs transforms are saved as an affine matrix and a Neuroinformatics Technology Initiative (NIFTI)-formatted vector field. These transforms and all other anatomical prerequisites can be easily obtained using the ANTs Cortical Thickness pipeline (ANTsCT; <https://github.com/ANTsX/ANTs>)⁵⁵. As an alternative to the ANTsCT pipeline, XCP supports several different levels of anatomical processing to obtain the requisite anatomical derivatives; these options are detailed in the online documentation.

Experimental design

The design of a confound model is subject to a number of considerations. Principal among these are the research hypothesis under consideration and the nature of the dataset being processed⁵⁶. Because motion can confound studies of individual or group differences (especially in clinical and developmental samples), a confound model's efficacy in reducing the motion–connectivity relationship is a primary concern in these studies. For basic research in connectomics, the integrity of network topology might be of equal importance; in this case, the distance dependence of residual artifact could be a critical indicator.

Although top-performing models can be expected to generalize reasonably well across datasets, there will be some variability in their performance. For example, it should be noted that these processing streams were developed and benchmarked on fMRI data with standard temporal resolution (TR) (i.e., TRs of 2–3 s), and their performance on multiband fMRI data with high temporal resolution has not been fully evaluated. As such, if performance benchmarking is a feasible option, then evaluating several motion correction strategies before processing the dataset can inform the decision as to which model to use. Otherwise, the set of existing benchmarking studies suggest first that top-performing models include GSR. Second, framewise censoring approaches provide a highly effective complement to GSR that can mitigate the distance-dependent profile of motion artifact but that can also incur considerable costs in the data's DOFs and autocorrelation structure. Third, although they might not be as effective as censoring, signal decomposition strategies such as ICA-AROMA and anatomical CompCor (aCompCor) provide an alternative mechanism for targeting spatially focal artifact that can be missed when applying GSR alone^{22–25}. Importantly, denoising elements are often combined to produce an effective confound model (Boxes 1 and 2).

To evaluate whether denoising is successful, investigators can examine measures computed at both the subject and the sample level (Tables 1 and 2). At the subject level, we suggest assessing the outcome of the denoising procedure by visualizing the entire dataset with a voxelwise carpet plot⁵⁷ (Fig. 2). This plot can qualitatively illuminate frame-to-frame relationships between movement and the BOLD signal and can be used to identify BOLD signal fluctuations that co-occur with subject movement. At the sample level, the residual correlation between motion and functional connectivity across subjects,^{10,11,17,28} or quality control–functional connectivity (QC–FC) correlation, provides a quantitative measure of denoising success (Fig. 3). Strong QC–FC correlations indicate that functional connectivity was substantially impacted by motion. Beyond QC–FC relationships, it is useful to measure QC–FC distance dependence, or the extent to which QC–FC correlations are dependent on the Euclidean distance between the centers of mass of any pair of regions^{10,11,17}. QC–FC distance dependence provides an informative diagnostic of the spatial profile of residual artifact (Box 2). In

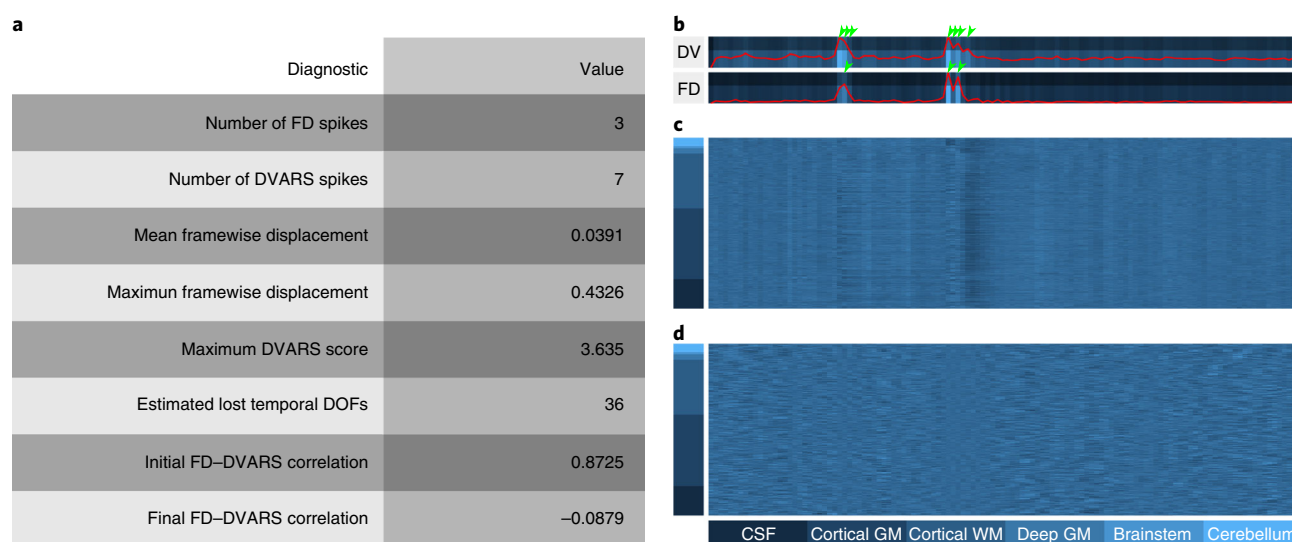


Fig. 2 | Summary of subject-level performance diagnostics and anticipated results (Steps 28–33). The illustrated results are from a single subject from the Philadelphia Neurodevelopmental Cohort (PNC)⁸⁷ and were processed using a 36-parameter stream that combines six realignment parameters, the mean WM and CSF time series, the mean global time series, and derivative and quadratic expansions. **a**, Summary of quantitative diagnostics shows a modest spike count (3 out of 120 frames flagged for high framewise displacement (FD); low mean framewise displacement indicates that this subject remained relatively still, apart from a few brief epochs of high movement. Refer to Table 1 for interpretation of the reported values. Notably, the GSR-based processing stream abolishes the strong FD–DVARS ‘D’ refers to temporal derivative of time courses, ‘VARS’ refers to RMS variance over voxels) correlation¹⁹. **b–d**, Visual aids for performance assessment⁵⁷ produced by the `voxels.R` command. **b**, Traces of the subject’s FD and DVARS score (DV). The plot ranges from 0 to the maximum value present in the data (in the example shown here, $FD_{max} = 0.432548$ and $DV_{max} = 3.635113$). FD and DV values are plotted as red lines. For this analysis, a frame was flagged as a spike if the FD value exceeded 0.25 or if the standardized DVARS score exceeded 2. The superthreshold region of the trace is demarcated by a darkened rectangle that covers the top fraction of each trace, allowing identification of flagged frames. In the traces shown, three frames exceed the FD threshold and seven exceed the DVARS threshold (green arrows). **c**, A voxelwise carpet plot of the subject’s BOLD activity, computed over the minimally pre-processed image. Time is plotted on the abscissa and is synchronized to the quality traces at the top. Space is plotted on the ordinate, with voxels sorted according to their membership in tissue compartments (bottom right). Within each tissue compartment, time series are sorted from most superficial (bottom) to deepest (top). In this subject, movements are associated with global bands of signal loss (type 2 artifact in Box 2), which is reflected in functional connectivity as global coupling. **d**, After motion correction, the same subject’s blood oxygen level-dependent (BOLD) signal no longer exhibits global bands, although some loss of signal variance is visible in the most contaminated frames. All participants in the example datasets used in this figure gave consent in accordance with a protocol approved by the Institutional Review Board at the University of Pennsylvania. DOF, degree of freedom.

addition to QC–FC and distance dependence, another potentially useful standard for the assessment of denoising methods is test–retest reliability. The Consortium for Reliability and Reproducibility (CoRR⁵⁸) hosts a public repository of datasets with repeated measures that provide an ideal resource for studies assessing test–retest reliability. However, investigators using this metric are cautioned that motion artifact itself exhibits test–retest reliability^{25,58}.

Another consideration is the decision to include or exclude censoring in the confound model. Censoring is a denoising approach whereby all variance in a motion-contaminated frame is completely removed from the time series, either via deletion of the frame or via spike regression^{17,18,53} (Box 1). Modern censoring approaches are designed to minimize the influence of censored frames by incorporating the censoring procedure into any intermediate stages of the analysis, such as detrending and temporal filtering, leaving the autocorrelation structure of the data intact for as long as possible^{18,59,60}. Two large-scale benchmarking studies have ranked censoring in combination with GSR as one of the most effective families of retrospective participant-level motion correction strategies^{22,23}.

However, a frequent argument against the use of censoring is that its application results in a large and variable loss of tDOFs in the data (see ‘Limitations’ section)^{18,44}. In general, we advocate censoring except in cases in which (ii) the input time series is short enough that removal of any additional frames could result in insufficient DOFs or (ii) the analytic objective requires the autocorrelation structure of the data to remain intact. For instance, sliding-window connectivity analysis depends on the autocorrelation of adjoining frames, and investigators using this technique in combination with censoring may need to account for any disruptions that censoring incurs to the autocorrelation structure. Although our research group has historically favored a model that combines censoring (either framewise or spatially adaptive despiking) with GSR, tissue-based regressors,

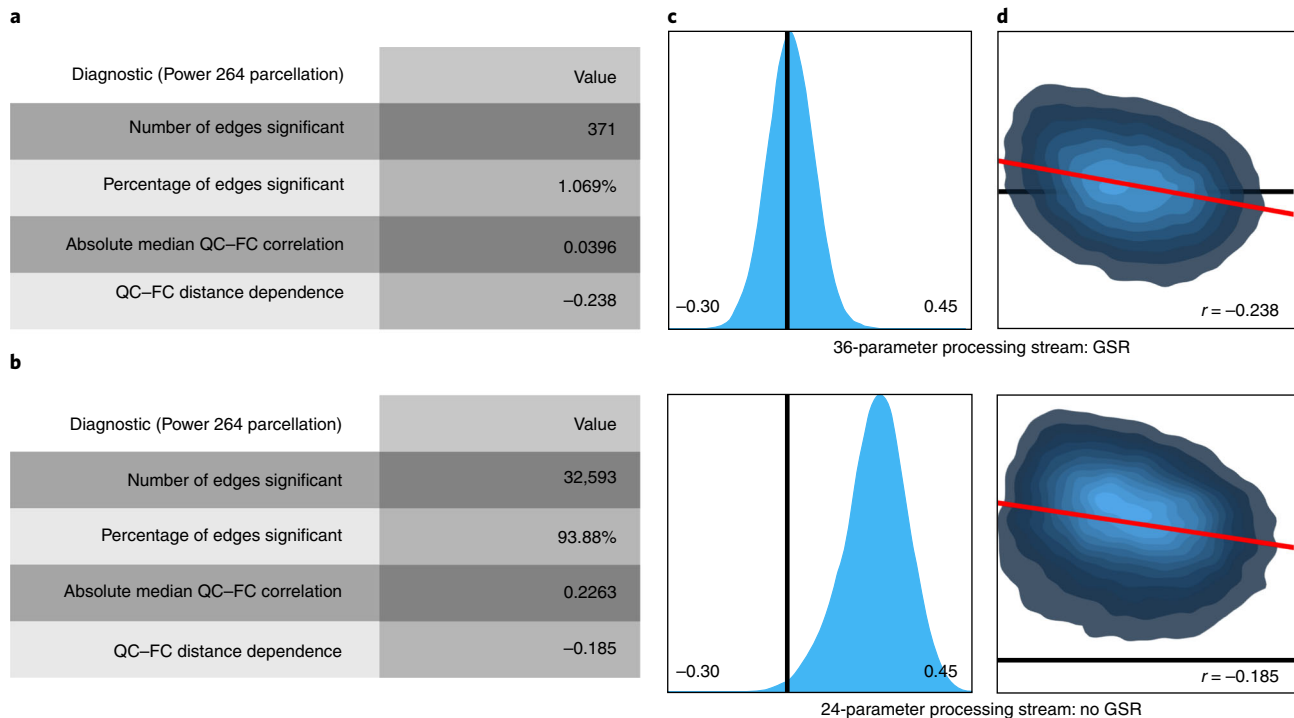


Fig. 3 | Summary of group-level performance diagnostics and anticipated results (Steps 34–36). **a–d**, The illustrated results from 500 low-motion subjects randomly sampled from the Philadelphia Neurodevelopmental Cohort (PNC)⁸⁷. The top panels show performance diagnostics when the data are processed using a 36-parameter stream that combines the mean global time series, six realignment parameters, the mean WM and CSF time series, and derivative and quadratic expansions. The bottom panels show analogous results when the same data are processed using a 24-parameter stream that uses only the six realignment parameters with derivative and quadratic expansions. **a–d**, Summaries of quantitative diagnostics (**a,b**); visual aids for performance assessment²² (**c,d**). Results are shown for the 264-node Power parcellation³. **a**, Results that show effective motion correction leave only a small percentage of edges that are significantly related to motion, with a weak absolute median QC-FC correlation of -0.04. **b**, The 24-parameter model performs poorly by comparison, leaving a marked absolute median QC-FC correlation of 0.2263, with nearly all edges exhibiting a significant relationship with motion. **c,d**, `qcfc.R` plots the distribution of QC-FC correlations across the two denoising schemes. **c** (Top), In a high-performance processing stream such as this, the distribution is narrow and centered at approximately zero. (Bottom) An ineffective stream, by contrast, has a QC-FC distribution that is broader and positively centered. **d**, QC-FC distance-dependence plots, with trend lines indicated in red and QC-FC distance-dependence estimates reported as Pearson's *r*. Euclidean distance is plotted on the abscissa and QC-FC is plotted on the ordinate; the horizontal black line corresponds to 0 QC-FC. (Top) A degree of QC-FC distance-dependence is unmasked by the GSR-based processing stream, as is evident in this visualization produced by the `featureCorrelation.R` command. (Bottom) In comparison with the GSR-based stream, motion artifact in the RP-only stream is less strongly related to the Euclidean distance between network nodes. A processing stream that augments GSR with either censoring or signal decomposition techniques will typically exhibit less distance dependence (Box 2).

motion estimates, and temporal and quadratic expansions, current indications suggest that models that combine GSR with either ICA-AROMA²⁰ or anatomical CompCor^{16,19} are very reasonable alternatives^{22,23}.

Alternative software packages to the XCP pipeline are available for automation of many protocol steps. A selection of alternatives is listed here, and their application is described. Consult the relevant manuals for detailed documentation and installation instructions.

- The Python-based software package C-PAC⁶¹ implements most subject-level motion correction strategies.
- The Python-based software package QAP⁶², available from the Preprocessed Connectomes Project, computes many subject-level diagnostics of motion artifact.
- The script `afni_proc.py` bundled with AFNI provides similar functionality.
- The MATLAB-based software packages CONN⁶³ and Connectome Computation System⁶⁴ implement many motion correction strategies and subject-level diagnostics.
- The Python-based software package `fmriprep` implements robust pre-processing, co-registration, and confound modeling (Steps 2–20) with minimal user overhead for datasets in the BIDS format⁶⁵.
- The rs-fMRI repository used in a recent benchmarking study²³ (<https://github.com/lindenmp/rs-fMRI>) provides a MATLAB-based implementation of the complete protocol that requires some manual editing by users to run on new datasets.

Level of expertise needed to implement the protocol

Designing a motion denoising pipeline de novo is a demanding programming task, requiring competence in time-series analysis and image processing, and a considerable familiarity with the range of software suites available for processing of brain images. By contrast, executing the automated processing stream provided via XCP is substantially more accessible. To use XCP, users should have attained a minimal level of competence with command-line usage, and should have access to a computing environment in which AFNI, ANTs, and FSL can be installed. We anticipate that most users of existing fMRI tools will be able to use this software successfully. In addition, familiarity with statistical modeling and an understanding of functional connectivity is required to design and interpret QC-FC models.

Materials

Equipment

Software

- A computer or virtual machine running a terminal emulator with Bash shell v.4.0 or higher
- **▲ CRITICAL** The following standard software libraries are needed for processing of neuroimaging data.
 - FSL⁵² (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslInstallation>)
 - AFNI⁵⁰ (https://afni.nimh.nih.gov/pub/dist/doc/html/doc/background_install/install_instructs/index.html)
 - ANTs⁵¹ (<http://stnava.github.io/ANTs/>)
 - C3D (<http://www.itksnap.org/pmwiki/pmwiki.php?n=Downloads.C3D>)
- The statistical language R, along with image-processing packages available from CRAN (filter, pracma, RNifti, and optparse) (<https://cran.r-project.org/>). To support graphical renderings of denoising diagnostics, it is recommended that the R packages ggplot2, reshape2, svglite, and grid also be installed. To support modeling of repeated measures from each subject, the R package lme4 must be installed.
- XCP system (<https://github.com/PennBBL/xcpEngine>)
- (Optional) ICA-AROMA²⁰ software implementation (<https://github.com/rhr-pruim/ICA-AROMA>) and Python 2.7 or higher (<https://www.python.org/download/releases/2.7/>), with Python modules os, argparse, commands, numpy, and random

Input data

- BOLD-weighted functional MR images of each subject's brain. These are the four-dimensional (4D) time series to be denoised.

Example use cases for denoising of subject data with XCP and for assessment of denoising performance are downloadable from FigShare (<https://figshare.com/s/d0161bac47f98eb1830b> and <https://figshare.com/articles/motion-test-data/6225968>), together with anticipated results. The provided examples include input images, a cohort file, expected outputs from subject-level denoising, and a set of 101 functional brain networks that can be used to validate the functionality of `qcfc.R` and `qcfcDistanceDependence.R`. Detailed documentation is provided with the example data.

! CAUTION Acquisition and analysis of neuroimaging data should be approved by an institutional ethical review committee. If new data are acquired locally, then all participants must provide informed consent to the acquisition and use of their data. Analysis of public datasets may require approval of a data-sharing agreement. All participants in the example datasets used in this protocol consented in accordance with a protocol approved by the Institutional Review Board at the University of Pennsylvania. **▲ CRITICAL** All images should be converted to NIfTI format prior to processing. **▲ CRITICAL** Because the results of fMRI processing are impacted by data integrity, both functional and anatomical images should be examined, either visually or using quantitative metrics^{62,66}, to evaluate suitability for processing. **▲ CRITICAL** The quality of image registration will be improved by correcting for magnetic field distortions incurred during functional image acquisition. If the acquired sequences support distortion correction of functional images (e.g., via FSL's FUGUE or TOPUP routines), this procedure should be performed before motion realignment (Step 5).

Anatomical and atlas input

- A high-resolution anatomical MR image for each subject, with skull and nonbrain tissue stripped from the image **! CAUTION** Acquisition and analysis of neuroimaging data should be approved by an institutional ethical review committee. If new data are acquired locally, then all participants must

provide informed consent to the acquisition and use of their data. Analysis of public datasets may require approval of a data-sharing agreement. All participants in the example datasets used in this protocol consented in accordance with a protocol approved by the Institutional Review Board at the University of Pennsylvania.

- Anatomical segmentations of the subject's brain into GM, WM, and CSF
- Transforms that map between coordinates in subject-anatomical space and coordinates in a template space that can be used as a standard across all subjects. This template space could be either the MNI152 template or a sample-specific template. For users of XCP Engine, XCP assumes that the transforms are in ITK (ANTs) format.
- If anatomical segmentations or ITK-formatted transforms between subject and template space are not available, these can be obtained by running the complete ANTsCT pipeline⁵⁵ for each subject. Alternatively, XCP implements several options for preparing all necessary anatomical input; refer to the online documentation for details. Online documentation is accessible at <https://pipedocs.github.io>.
- An atlas or parcellation that defines regions of interest across the brain **▲ CRITICAL** The atlas should ideally be designed as a functional parcellation that is suitable for connectomic analysis. Commonly used parcellations include the 264-region Power parcellation³, the 333-region Gordon areal parcellation⁶⁷, the Yeo network parcellation⁴, the multiscale Lausanne parcellations⁶⁸, data-driven functional parcellations⁶⁹, and the multiscale local-global Schaefer parcellations⁷⁰. **▲ CRITICAL** If manually processing the data without XCP, take note of the coordinate space in which the parcellation is defined. Many atlases are available in the MNI152 coordinate space, for instance. The transforms above should be able to map, or 'warp', the parcellation into subject-anatomical space using an appropriate warping command, such as `antsApplyTransforms` (with ANTs) or `applywarp` (with FSL). If they are not, then it will be necessary to either (i) compute new transforms from subject-anatomical space to the parcellation space and use those as the input transforms, or (ii) obtain a parcellation in a space that is accessible to the existing transforms. **▲ CRITICAL** If processing the data using XCP, the connectome will be computed over the Power parcellation by default. To run additional parcellations, download them from the brainspaces database on GitHub (<https://github.com/brainspaces>) and then run, from the XCP install directory:

```
utils/atlasMetadata -a <parcellation name> -d <downloaded repository>
```

Equipment setup

ANTs, AFNI, FSL, and R **▲ CRITICAL** Instructions for installation and setup of all software packages are available at the websites of these packages.

- ANTs: <http://stnava.github.io/ANTs/>
- AFNI: https://afni.nimh.nih.gov/pub/dist/doc/html/doc/background_install/install_instructs/index.html
- FSL: <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FslInstallation>
- R can typically be installed using the operating system's package manager. The required R packages can be installed from within R by calling the `install.packages()` command with the package name as argument. To verify successful installation of all dependencies, run the XCP command-line utility `core/checkDependenciesXCP`.

XCP utilities

To run XCP utilities, define the environmental variable `XCPEDIR` to point to the XCP install directory:

```
export XCPEDIR=<full path to install directory>
```

Including this definition statement in the user's `.bash_profile` or `.bashrc` is recommended.

Running XCP Engine

XCP automates all motion correction and quality assessment steps described below through its front end, `xcpEngine`. This automation is powered by a user-provided cohort file (subject list) and design file (which toggles the optional steps specified below). The design file can be configured by running the script `xcpConfig` in the XCP install directory, after defining `XCPEDIR` as above, and then selecting 'Functional connectivity'. Consult the software manual for detailed documentation (<https://pipedocs.github.io>).

Running XCP Engine on a computing cluster, without installing dependencies

In lieu of installing all dependencies, it is possible to complete the motion correction protocol using the IPP (<https://ipp.cbica.upenn.edu/>) at the CBICA at UPenn, which implements pre-processing in XCP. For input to IPP, prepare a design file as described above in the ‘Running XCP Engine’ section. Upload functional input data in compressed NIfTI format (.nii.gz), and upload any processed anatomical data, first archiving and compressing the directory containing all anatomical inputs (e.g., the ANTsCT output directory) in .tar.gz format. If anatomical data were processed in XCP, each subject’s `struc` subdirectory should be archived and compressed as .tar.gz.

Running XCP Engine inside a Docker or Singularity container

Docker and Singularity are technologies that enable the packaging of software and its dependencies into images and running them in containers. A Docker image is available for running XCP Engine, or to convert to a Singularity image. To run the motion correction protocol on a system with either Docker or Singularity installed, consult the XCP container documentation at <https://pipedocs.github.io/containers>. For input, prepare a design file as described above in the ‘Running XCP Engine’ section. Detailed instructions can be found in the XCP container documentation on how to bind local file system directories so they can be accessed from the container. Note that the Docker image contains XCP and all its dependencies.

Input data

Structural and functional images should be named in a systematic manner. The BIDS standard⁷¹ presents one recommended way to organize data in a consistent manner. Details of the BIDS specification for data organization and naming conventions are available on the BIDS home page at https://bids.neuroimaging.io/bids_spec.pdf.

Procedure**Preparing a subject list ● Timing 1–10 min**

▲ **CRITICAL** When using XCP Engine, a subject list (or ‘cohort file’) supports iteration and uniform processing across the entire sample. An example is provided in Box 3, along with basic specifications. Detailed specifications are available in the online documentation (<https://pipedocs.github.io/config/cohort.html>).

- 1 To prepare a subject list, associate each subject’s functional time series with that subject’s processed anatomical data (segmentation, transforms, and brain-extracted, bias-corrected image). For each subject, write a comma- or tab-delimited row of values to the subject list file. Each subject’s values can include any information that could facilitate processing of the subject’s data, including subject identifiers and input paths. However, values should be consistent across subjects—if the third column (value) for subject 1 contains the path to subject 1’s functional time series, then the third column for all subjects should contain the path to those subjects’ functional time series. All processing steps described below under the headings ‘Minimal pre-processing’, ‘Co-registration’, ‘Confound modeling’, ‘Confound regression’, and ‘Performance assessment: subject level’ should be performed separately for each subject, whereas those under ‘Performance assessment: sample level’ should be performed for the analytic sample as a whole by pooling outputs from all subjects.

Minimal pre-processing ● Timing 5–15 min per subject

▲ **CRITICAL** XCP’s `prestats` module implements minimal pre-processing.

- 2 Discard the initial volumes of the functional time series, using the `fslroi` command and specifying the functional time series as `<input>`, the number of volumes to be discarded as `<tmin>`, and `-1` (indicating retention of all remaining volumes) as `<tsize>`:

```
fslroi <input> <output> <tmin> <tsize>
```

▲ **CRITICAL STEP** The number of volumes to discard will depend on the acquisition parameters of the functional data and on the specifications of the scanner. This step is unnecessary if initial volumes have already been discarded as part of the acquisition protocol.

- 3 Estimate framewise head motion by computing the realignment parameters (RPs) and the overall framewise displacement (FD) of the head. Estimating the RPs will require selection of an exemplar volume from the 4D time series to serve as a reference for realignment. By default, this will be the

midpoint of the time series. Compute the RPs, using FSL's `mcflirt` command and specifying the time series as input⁷²:

```
mcflirt -in <input> -out <output root> -reffile <reference for re-
alignment> -plots -rmsrel -rmsabs -spline_final
```

Toggle the `rmsrel` and `rmsabs` flags in order to compute additional summary measures of the head's framewise and overall displacement. The `plots` flag configures the program to generate a visualization of motion estimates, whereas `spline_final` specifies the interpolation approach. The output with suffix `.par` contains the six RPs in six columns, and the output suffixed `.rel.rms` contains the FD in a single column.

▲ CRITICAL STEP By convention, the displacement of the head in the first acquired frame is defined to be 0. By default, however, the `mcflirt` command does not write out the FD for the first frame. To ensure that temporal masking (Step 8) selects the correct frames for removal, the default output from `mcflirt` should be padded with a leading 0.

- 4 (Optional) *Apply correction for timing of slice acquisition.* Use FSL's `slicetimer` program, consulting the acquisition parameters of the data to determine the most appropriate call to `slicetimer`.

▲ CRITICAL STEP Interpolation steps, such as slice-time correction and despiking, affect estimates of RPs⁷³. To ensure that RPs are computed independently of these procedures, the procedures should not be run until after the RPs have been estimated. However, proper slice-time correction requires information about the 2D slice in which each voxel was acquired. Because motion realignment reslices the time series, slice-time correction should be performed before realignment but after estimation of RPs. Accordingly, the input to slice-time correction should be the output of Step 2, and not the output of Step 3.

- 5 *Realign all volumes relative to a reference.* As in Step 3, the `mcflirt` command should be called:

```
mcflirt -in <input> -out <output root> -reffile <reference for realign-
ment> -mats -spline_final
```

Use an `<output root>` different from the `<output root>` in Step 3 to ensure that the RPs are not overwritten. Because RPs have already been computed, it is also possible to select a reference volume on the basis of low framewise head motion or to use the average of all volumes as the reference for this iteration of realignment. Without use of slice-time correction, Steps 3 and 5 can be combined into a single step.

- 6 Classify all image voxels as brain or nonbrain, and delete nonbrain values from the image. This step can be completed by passing the exemplar volume to the FSL command-line tool `bet` to generate a binary brain mask and afterward multiplying the processed functional time series by the brain mask using the `fslmaths` command to produce a skull-stripped time series.
- 7 Compute standardized DVARS score, an index of the frame-to-frame signal change across the brain that can be used to flag outliers^{11,74,75}. Use the skull-stripped time series from Step 6 as the input to the XCP utility script `dvars`:

```
${XCPEDIR}/utils/dvars -i <input> -o <output>
```

DVARS will produce three outputs. The output with suffix `.std.1d` contains the standardized DVARS score.

- 8 (Optional) *Compute temporal masks and generate spike regressors.* Skip this step if censoring is not performed. A temporal mask is a binary-valued time series that is equal in length to the functional time series and that indicates whether each frame of the functional time series should be preserved or excised (censored) (refs^{11,18}). Before proceeding, determine whether temporal censoring is a reasonable processing strategy for the dataset and analytic goals (Experimental design). Next, select criteria and thresholds for censoring (e.g., $FD > 0.2$ mm or standardized DVARS score > 2). Temporal masks can be generated for each criterion using either `ld_tool.py` in AFNI or the XCP utility `tmask.R`:

```
${XCPEDIR}/utils/tmask.R -s <time series> -t <threshold> -o <output>
```

To prepare a temporal mask, supply the criterion `<time series>` (output suffixed `.rel.rms`), the framewise motion estimates from Step 2 and, optionally, the DVARS score estimates from Step 7) and the censoring `<threshold>` for each criterion. Define the final temporal mask as

the union of all single-criterion temporal masks. After obtaining the final temporal mask, derive spike regressors^{17,53} from this mask by using the XCP utility `tmask2spkreg.R`, specifying the temporal mask as `<tmask>`:

```
${XCPEDIR}/utils/tmask2spkreg.R -t <tmask> >> <output spike regres-  
sors>
```

▲ CRITICAL STEP Although censoring has been established as an efficacious motion correction strategy, the exact threshold and criteria for determining whether a frame should be censored vary from study to study. For example, in typical single-band fMRI studies with TRs of 3 s, censoring any volumes with FD > 0.1 mm is very stringent, whereas censoring volumes with FD > 0.25 mm is less stringent but still effective. Furthermore, thresholds for framewise censoring criteria have typically been reported using units per frame (e.g., millimeters per frame for FD). However, it should be emphasized that 0.25-mm FD during a 1-s frame does not reflect the same degree of motion as does 0.25-mm FD during a 3-s frame.

▲ CRITICAL STEP There are several different but highly correlated ways to define FD. Here, we refer to FD as defined by the output of the `mcflirt` command (FD_{Jenkinson}). An alternative formulation of FD, FD_{Power}, can be computed by passing the RPs from Step 2 to the XCP utility `fd.R`; this measure can then be used as a criterion for temporal masking. FD_{Power} and FD_{Jenkinson} are highly correlated, with FD_{Power} equal to approximately twice FD_{Jenkinson}^{25,56}.

? TROUBLESHOOTING

- 9 (Optional) *Despike the time series.* Despiking is a spatially adaptive interpolation approach that identifies voxelwise signal outliers and imputes new values for each of those outliers. Perform despiking using AFNI's `3dDespike` program:

```
3dDespike -prefix <output> -nomask -NEW <input>
```

▲ CRITICAL STEP Note that alternative despiking techniques exist, including wavelet-based despiking⁷⁶. In addition, the relationship between despiking and temporal censoring is not characterized; it is likely that both techniques target the same types of noise using different identification criteria. As a result, despiking is often used as an alternative to temporal censoring, and combining despiking and temporal censoring in the same processing stream is uncommon.

- 10 Remove the mean and any linear or polynomial trends from the functional time series, using the XCP utility `dmdt.R`:

```
${XCPEDIR}/utils/dmdt.R -i <img> -d <order> -o <output> -m <mask> -x  
<mean image> [-t <tmask>]
```

Specify the skull-stripped time series from Step 6 as ``, the brain mask from Step 6 as `<mask>`, and (if applicable) the temporal mask from Step 8 as `<tmask>`. De-meaning and de-trending operates by fitting a polynomial basis to the input time series, using a general linear model and afterward discarding the explained variance. Any frames flagged for censoring in the temporal mask are not considered in the fit¹⁸. The de-trend `<order>` command refers to the highest-order polynomial term included in the model, which can either be defined a priori or estimated using the formula

$$1 + \frac{T_R \times n}{150},$$

where T_R is the sampling time of the functional time series in seconds and n is the number of frames sampled⁵⁰. Toggle the `-x` option to generate a mean functional image, defined as the fit of the constant term of the model, and define an appropriate output path.

▲ CRITICAL STEP Save the last output produced during minimal pre-processing, as useful diagnostic information can be obtained from it.

Co-registration ● Timing 20–45 min per subject

▲ CRITICAL XCP's `coreg` module implements co-registration.

- 11 *Identify the WM boundary.* Use the tissue-class segmentation to obtain a binary mask indicating whether each voxel in the anatomical image represents WM tissue. If necessary, the tissue-specific WM

mask can be generated with the XCP utility `val2mask.R`, using the anatomical segmentation as the reference image and selecting all values that correspond to WM as the values of interest:

```
{XCPEDIR}/utils/val2mask.R -i <input> -v <WM values> -o <output>
```

- 12 Co-register the functional image to a high-resolution anatomical reference acquired for the same subject. Use FSL's `flirt` command to compute the co-registration, specifying either the reference volume from Step 4 or the mean image from Step 10 as the input (<source>) and the subject's anatomical image as the <reference>:

```
flirt -in <source> -ref <reference> -dof <degrees of freedom> -out  
<output image> -omat <output matrix> -cost bbr -wmseg <white matter  
segmentation volume>
```

Set the <degrees of freedom> to 6 (or alternatively 9, if distortion correction has not been performed), and ensure that an <output matrix> path is defined. Specify `bbr` (boundary-based registration) as the cost function, and use the binary WM mask from Step 11 as the <white matter segmentation volume>⁷⁷. After the co-registration is performed, use FSL's `convert_xfm` command to calculate the inverse of the <output matrix> (a .mat file) specifying the <output matrix> from the forward transformation as the argument to the `inverse` flag:

```
convert_xfm -omat <output inverse> -inverse <matrix>
```

▲ CRITICAL STEP Because the co-registration computed here will later be necessary for computing tissue-based confounds, the quality of the co-registration should be assessed before proceeding. The most reliable way to assess co-registration quality is through expert visual inspection of all registered images. However, co-registration quality can also be quantified using spatial cross-correlation and coverage metrics. Cross-correlation can be computed using the `fslcc` command, where coverage can be computed as the percentage of the binarized structural image that also lies within the binarized co-registered image or as the Dice and Jaccard coefficients between the binarized structural and co-registered images. All quality indices can be obtained using the XCP utility `maskOverlap.R`:

```
{XCPEDIR}/utils/maskOverlap.R -m <co-registered functional image>  
-r <anatomical image>
```

▲ CRITICAL STEP After computing these values for all subjects, they can be used to flag outliers for more thorough visual inspection.

- 13 (Optional) Create ITK-compatible versions of the forward and inverse co-registration matrices. Skip this step if using any program other than ANTs to warp images between coordinate spaces. This step will enable use of the transforms with ANTs. Converting the FSL-based co-registration matrices to ITK format can be accomplished using the `c3d_affine_tool` command with the `fsl2ras` and `oitk` flags, specifying each .mat file from Step 12 as the <transform> file, providing source and reference images for each transform as for the `flirt` command, and specifying the output path using the `oitk` flag:

```
c3d_affine_tool -src <source> -ref <reference> <transform> -fsl2ras  
-oitk <output>
```

To verify that the conversion was successful, apply the reformatted transform to the <source> image from Step 12 and evaluate whether the output is identical to the <output image> from Step 12.

Confound modeling ● **Timing** 1–5 min per subject without ICA-AROMA; 20–30 min per subject with ICA-AROMA

▲ CRITICAL STEP XCP's 'confound' module implements confound modeling.

- 14 (Optional) Use ICA-AROMA to decompose the pre-processed data into linearly independent signal sources, identify motion-related components, and remove them from the data²⁰. Additional

documentation and detailed instructions for implementation are available online along with the ICA-AROMA code (<https://github.com/rhr-pruim/ICA-AROMA/blob/master/Manual.pdf>).

- 15 Compute the mean global signal²⁶. Use the `fslmeants` function, providing the functional image as the `<input>` and providing the whole-brain mask from Step 6 as the `<mask>`:

```
fslmeants -i <input> -o <output> -m <mask>
```

- 16 Compute eroded WM and CSF masks. The XCP utility script `erodespare` erodes a mask such that only a user-specified percentage of the deepest tissue in the original mask remains:

```
${XCPEDIR}/utils/erodespare -i <input> -o <output> -r <retention  
criterion> -v <value set>
```

It is recommended that masks be eroded to retain 5–10% of their original size (`<retention criterion>` of 5–10) so as to minimize partial volume effects. Use the anatomical segmentation as the `<input>`. For the `<value set>`, enter the values that correspond to the WM or CSF label in the anatomical segmentation. Finally, warp the output-eroded masks into the same coordinate space as the functional time series.

▲ CRITICAL STEP Because the WM and CSF signals mix with the GM signal near the GM interface, the WM and CSF masks derived from the anatomical segmentation will include influence from adjoining GM voxels^{18,28}. This influence should be mitigated by eroding the WM and CSF masks to exclude boundary voxels; more liberal erosion will result in WM and CSF signal estimates that are more independent of the global signal. However, erosion should leave enough voxels in each mask in order to spatially sample the tissue class of interest.

? TROUBLESHOOTING

- 17 Compute nuisance signals in WM and CSF. Several standard options are available for extracting nuisance signals from the WM and CSF. Follow option A to compute mean signals within the WM and CSF or option B to use `aCompCor`^{16,19}.

(A) Compute mean signals within the WM and CSF, and add these to the motion model

● **Timing** <20 s per subject

- (i) Compute mean signals as in Step 15, substituting the eroded tissue-specific masks from Step 16 for the whole-brain mask.

(B) Use `aCompCor` ● **Timing** 30 s–1 min per subject

▲ CRITICAL `CompCor`^{16,19} decomposes the signals in the eroded WM and CSF masks using PCA.

- (i) Determine the number of components from each tissue class (WM and CSF) to use in the motion model. First, use AFNI's `3dpc` utility to decompose the data in each eroded tissue mask.

```
3dpc -prefix <output> -mask <eroded mask> [-pcsave <number of  
PCs>] <input>.
```

The cumulative variance explained is indexed in the `<output>_eig.1D` file under the 'cumulative fraction' heading, and the principal component time series are saved in an output file called either `<output>_vec.1D` or `<output>.1D`. For each tissue compartment (WM and CSF), find the row in the `<output>_eig.1D` file corresponding to 50% cumulative variance explained¹⁹. The component number of this row will be the number of components to use in the motion model.

- (ii) After determining the correct number of components, rerun `3dpc` as in Step 17B(i) above, this time specifying the number of components to be included, using the `-pcsave` option.

- 18 (Optional) Prepare any custom nuisance time series. These can include any additional nuisance time series available for the current dataset, including recordings of artifactual processes such as pulse and respiration or nuisance time series defined separately for each voxel (as well as expansions of these where appropriate)⁷⁸. If the data are being processed for a task-constrained functional connectivity analysis, the task model can be included in the confound model, depending on analytic objectives⁷⁹.

- 19 (Optional) *Compute time-series expansions*¹⁵. For each time series in the motion model, compute its temporal derivative, its quadratic term, and the temporal derivative of its quadratic term. Note that adding a previous time point to the confound model and adding a temporal derivative to the confound model for the purposes of confound regression are equivalent, as the temporal derivative of a discretely sampled time series can be expressed as the difference between the original time series and the backward-shifted time series (a linear combination). To ensure that the confound model is not overspecified because of collinearity, a single confound model should not include both derivatives and temporal shifts.
- 20 *Build the confound model into a matrix*. Concatenate the time series produced in Steps 3 and 15–19 into a model matrix, using AFNI's `ldcat` utility.
▲ CRITICAL STEP Do not add the spike regressors to the confound model yet, as they should not be subjected to any temporal filtering.
▲ CRITICAL STEP To verify that WM and CSF signals are sufficiently independent from the global signal, it is helpful to estimate the collinearity of all predictor variables. A matrix of correlations among predictor variables can be obtained using the XCP utility `ts2adjmat.R` and specifying the complete motion model as the input time series. (This matrix will be written in the form of a feature vector, which can be converted to a symmetric square matrix using the `squareform` function in SciPy, MATLAB, or the R library `pracma`.)

Confound regression ● Timing 5–10 min per subject without censoring; 60 min per subject with censoring

▲ CRITICAL XCP's `regress` module implements confound regression.

- 21 (Optional) *Interpolate over epochs marked for censoring*. Skip this step if censoring is not performed. Temporal filtering (Step 22) can result in propagation of artifactual variance from frames flagged for censoring into adjoining frames^{18,59,60}. Interpolation of values in contaminated frames can help make the temporal filter robust to noise in censored frames while preserving the overall autocorrelation structure of the data. Currently, the recommended approach¹⁸ is based on the Lomb–Scargle periodogram⁸⁰; this approach generates surrogate data based on the spectral characteristics observed in low-noise (unflagged) frames. This approach is, however, computationally intensive. Less rigorous but faster interpolation can be achieved using neighbor-based approaches, which sacrifice the data's autocorrelation structure^{59,60}. The periodographic approach is currently implemented in the XCP utility `interpolate.R`.
- 22 *Apply a temporal filter to the data*. A commonly used pass band ranges from a high-pass limit of 0.01 Hz to a low-pass limit of 0.08 Hz, above which the power spectra of high- and low-motion subjects diverged, even when a high-performance confound model was used¹⁷.
▲ CRITICAL STEP Ensure that the same filter is applied to both the time-series data and the confound model to prevent spectral misspecification, which can result in reintroduction of artifactual variance into suppressed frequency bands during the model fitting stage⁸¹ (Step 24).
? TROUBLESHOOTING
- 23 Concatenate the spike regressors computed in Step 8 into the filtered model matrix, as in Step 20.
- 24 *Perform confound regression*. Use a general linear model to estimate the contributions of artifactual signals to the BOLD signal at each voxel. The AFNI program `3dTproject` computes the model fit and returns the residuals as `<output>`:

```
3dTproject -input <input> -ort <regressors> [-dsort <voxelwise>]
-prefix <output>
```

Provide the filtered time series as `<input>` and the filtered confound model as `<ort>`. Any voxelwise-defined regressors should be provided as `<dsort>`, and the desired output path should be provided as `<output>`. (It is also possible to combine Steps 22–24 into a single step using `3dTproject`. Refer to AFNI's online `3dTproject` documentation for details: https://afni.nimh.nih.gov/pub/dist/doc/program_help/3dTproject.html.)

- 25 (Optional) *Censor the data*^{11,18}. Skip this step if censoring is not performed. Perform censoring using the XCP utility script `censor.R`, providing the temporal mask as `<tmask>` and the time-series residuals as ``:

```
${XCPEDIR}/utils/censor.R -i <img> -t <tmask> -o <output>
```


(Optional) Re-compute the mean FD after censoring. In Step 35, residual motion artifact is estimated as the correlation between mean FD and connectivity. Using the post-censoring FD in this computation links the outcome value more directly to the residual motion in the dataset but reduces the variance in motion in comparison with the pre-censoring data. Although the XCP pipeline at present uses pre-censoring FD, all performance assessment scripts have been designed to support either alternative.

▲ CRITICAL STEP Censoring will disrupt the autocorrelation structure of the data. Therefore, any processing steps that are sensitive to autocorrelation structure (e.g., temporal filtering or dynamic connectivity analysis) should either be performed before censoring or be performed in a way that accounts for disruptions in the autocorrelation structure.

- 26 (Optional) Re-mean the time series by adding the mean image computed in Step 10 to the residualized data, using the `fslmaths` script:

```
fslmaths <input> -add <mean image> <output>
```

? TROUBLESHOOTING

- 27 (Optional) *Spatially smooth the image data.* Smoothing across tissue boundaries can be reduced by using an adaptive filter that combines smoothing and edge detection. This can be achieved using the FSL program `susan`, providing the processed time series as input⁸²:

```
susan <input> <bt> <kernel> 3 1 1 <usan> <bt> <output>
```

Specify the mean image from Step 10 as the `<usan>` (the image over which tissue boundaries are detected). The `<bt>` (brightness thresholds) can be estimated as 75% of the median intensity value in the `<usan>`. To obtain the median intensity value, use `fslstats` to compute the 50th percentile, limiting the computation to voxels inside the brain by using the brain mask with the `-k` parameter:

```
fslstats <usan> -k <mask> -p 50
```

Alternatively, to completely restrict smoothing across tissue boundaries, warp the anatomical segmentation into the same coordinate space as the processed image and then use this as the `<usan>`, setting the `<bt>` arbitrarily low.

Performance assessment: subject level ● Timing 5–10 min per subject

▲ CRITICAL XCP's `fcon` and `qcfc` modules implement performance assessment.

- 28 *Compute the functional connectome within a selected parcellation.* To do this, first select a parcellation and warp it into the same coordinate space as the processed image. We suggest using multi-label interpolation rather than nearest neighbor (using `antsApplyTransforms`). Next, compute the local time series for each parcel, using the XCP utility `roi2ts.R`, providing the parcellation as the `roi` and providing a list of regions as `labels` if one is available. Finally, obtain the adjacency matrix for the connectome, using the XCP utility `ts2adjmat.R` and providing the local time series as the `ts` argument.
- 29 *Prepare a depth map for the anatomical segmentation.* A depth map is a voxelwise map of the brain in which the value of each voxel indicates the voxel's depth within its assigned tissue class; this map will be used to create a voxelwise summary plot of the data (Fig. 2). (In the depth map, the first digit of the voxel's value indicates its tissue class, and the last two digits indicate its depth in that tissue class, as a percentile. For instance, if WM is assigned a value of 3 in the anatomical segmentation, a superficial WM voxel at the 30th percentile of depth would be assigned a value of 330, and a deep WM voxel at the 98th percentile of depth would be assigned a value of 398.) This depth map can be generated using the XCP utility `layerLabels` and defining the input label set as the segmentation:

```
${XCPEDIR}/utils/layerLabels -i <label set> -o <output>
```

Next, warp the depth map into the same coordinate space as the minimally pre-processed image and as the final processed image, using nearest-neighbor interpolation.

- 30 Prepare a subject-level voxelwise summary plot for the minimally pre-processed data and for the denoised data⁵⁷:

```
{XCPEDIR}/utils/voxts.R -i <img1>,<img2> -r <roi> -o <output> -t
FD:<FD>:<FD threshold>,DV:<DVARS>:<DVARS threshold>
```

Use the depth map from Step 29 as the <roi>. Provide the minimally pre-processed image as <img1> and the final, denoised image as <img2>. Provide the path to the FD time series as <FD> and the path to the DVARS time series as <DVARS>. If the data were censored based on FD or DVARS score, then provide the censoring thresholds as <FD threshold> and the DVARS threshold as <DVARS threshold>. Otherwise, select reasonable thresholds for flagging outliers. Label names can be found at {XCPEDIR}/atlas/segmentation3 for 3-tissue-class segmentations and {XCPEDIR}/atlas/segmentation6 for 6-tissue-class segmentations. Example output is presented in Fig. 2.

- 31 Compute the DVARS score of the denoised dataset as in Step 7. Indicate that the input data are de-meaned, using the -d 1 argument, and provide the skull-stripped mean image from Step 10 as the -b argument:

```
{XCPEDIR}/utils/dvars -i <input> -o <output root> -b <mean image> -d 1
```

- 32 Compute FD–DVARS correlations¹⁹ using the XCP utility featureCorrelation.R:

```
{XCPEDIR}/utils/featureCorrelation.R -i <path to framewise displacement time series>,<path to standardised DVARS time series> >>
<output correlation>
```

Perform the calculation for both the minimally pre-processed DVARS score from Step 7 and the denoised DVARS score from Step 31.

- 33 *Estimate the loss of tDOFs*. This loss can be approximated as the total number of terms in the confound model (columns in the model matrix from Step 23), including spike regressors (if censoring was performed) and independent components identified as noise (if ICA-AROMA was run).

? TROUBLESHOOTING

Performance assessment: sample level ● Timing 20 min

▲ **CRITICAL** XCP's *fcon* and *qcfc* modules are used in this section to implement performance assessment.

- 34 Prepare a sample matrix with $P + 2$ columns, where P is the number of unique identifiers (e.g., subject identifier, scan identifier, and session identifier) for each subject. Place each subject in a separate row, and place each identifier in a separate column. Each column containing identifiers should have a header beginning with the string id, and the remaining two columns should have the headers *motion* and *connectivity*. In each subject's *motion* column, enter the subject's mean FD. This can be obtained from a file with the suffix *rel_mean.rms* produced in Step 2. In each subject's *connectivity* column, enter the path to the subject's connectivity matrix, the final output produced in Step 26. Save the matrix in .csv format. An example sample matrix is presented in Box 4.
- 35 *Compute the distribution of residual QC–FC correlations*^{22,28,56}. This step can be facilitated using the XCP utility *qcfc.R* (usage examples in Box 4). The utility script computes, for each edge in the connectome, the partial correlation of motion with the strength of that edge, after controlling for the effects of any user-provided covariates. Each row in the sample matrix from Step 34 represents an observation for the correlation.

▲ **CRITICAL STEP** The number of connections significantly related to motion at any significance threshold is sensitive to the total sample size. Accordingly, it is important to ensure adequate statistical power for detection of motion-related variance. When evaluating the comparative efficacy of a novel denoising procedure, it is especially critical to use a sample that is sufficiently large to detect residual artifact. Refer to previous benchmarking studies to guide sample selection^{22,23}.

- 36 Compute the distance dependence of QC–FC correlations^{10,11,17}. QC–FC distance dependence can be computed using the XCP utility *qcfcDistanceDependence*. This step entails (i) computing

a pairwise matrix of distances between each pair of parcels and (ii) computing the correlation between distance and residual QC-FC. `qcfcDistanceDependence` performs both of these steps, using the warped parcellation from Step 28 and the QC-FC values from Step 35:

```
${XCPEDIR}/utils/qcfcDistanceDependence -a <parcellation> -q <QC-FC values> -d <output distance matrix> -o <output distance-dependence> -f <output figure>
```

Troubleshooting

Troubleshooting advice can be found in Table 3.

Table 3 | Troubleshooting table

Step	Problem	Possible reason	Solution
8	Very few frames survive censoring	(i) Subject data are highly contaminated by noise; (ii) The censoring criterion is too stringent	(i) Exclude the subject from further analysis; (ii) Use a more lenient censoring threshold (refer to CRITICAL STEP note in Step 8 for threshold selection details)
16	Tissue time series are not produced	The tissue compartment mask is empty	Increase the retention criterion for mask erosion
22	Confound regression fails	The confound model is not well formulated	Evaluate the confound model for missing values and collinearity; reduce or reevaluate the model as necessary
26	Missing values (NA (not available) or NaN (not a number)) in the connectome	(i) Poor registration quality; (ii) Parcels are of variable size	(i) Ensure that the sequence of transforms applied to the parcellation is correct, and then re-run registration steps as necessary, potentially using a different cost function; (ii) Select a different parcellation, or exclude smaller parcels from the connectome across the entire sample
33	The QC-FC distribution is very broad	Insufficient statistical power	Increase the sample size, or use subject-level diagnostics only
	No edges have a significant relationship with motion	(i) Insufficient statistical power (ii) Effective denoising model	(i) Increase the sample size, or use a less stringent correction for multiple comparisons; (ii) Ensure that statistical power is sufficient; otherwise, there is nothing to troubleshoot
	Excessive residual QC-FC relationship	(i) Confound regression failed; (ii) High-motion subjects are driving the correlations; (iii) The confound model performs poorly in the dataset under analysis	(i) See ? TROUBLESHOOTING for Step 22; (ii) Apply subject-level exclusions based on motion; (iii) Use a different confound model, or benchmark performance of alternative models on a subsample of the data to select a model

Timing

Step 1, preparing a subject list: 1–10 min

Steps 2–10, minimal pre-processing: 5–15 min per subject, depending on the number of samples and the spatial resolution of the input time series; timing is potentially significantly longer for high-resolution multiband time series of long duration

Steps 11–13, co-registration: 20–45 min per subject

Steps 14–20, confound modeling: 1–5 min per subject; 20–30 min or longer if ICA-AROMA is included; timing scales with the number of frames and the spatial resolution of the input time series

Steps 21–27, confound regression: 5–10 min per subject; 60 min per subject if censoring is included

Steps 28–33, subject-level performance assessment: 5–10 min per subject

Steps 34–36, sample-level performance assessment: 20 min

Anticipated results

The expected product of the protocol is a set of denoised functional time series and connectomes, one for each input to the processing stream. Although some motion artifact will probably be present even

Box 4 | Example subject list, model specification, and output files for qcfc.R

The sample-level performance assessment script `qcfc.R` has a number of options for model, input, and output specification:

```
${XCPEDIR}/utils/qcfc.R -c <cohort> -o <output root> [-s <multiple comparisons correction> -t <significance threshold> -n <confound> -y <conformula>]
```

Optional arguments are denoted in square brackets ([]). The primary input to `qcfc.R` (<cohort>) should be the subject list created in Step 34, an example of which is provided here:

```
id0,id1,motion,connectivity
ACC,001,0.0391,processedData/ACC_001_fc/connectome.txt
ACC,002,0.0455,processedData/ACC_002_fc/connectome.txt
ACC,003,0.0367,processedData/ACC_003_fc/connectome.txt
DSQ,001,0.1532,processedData/DSQ_001_fc/connectome.txt
CAT,001,0.0811,processedData/CAT_001_fc/connectome.txt
```

The type of correction for multiple comparisons can be specified as `fdr` (default), `bonferroni`, or `none`. The maximal *P*-value threshold for significance can also be specified by the user; in the absence of user input, a default value of 0.05 will be used.

The values of any model covariates (such as age and sex) should be included in another file containing the same subject identifiers as the sample matrix. The file containing model covariates should be provided as the <confound> argument, and the formula for the model should be provided as the <conformula> argument, with any categorical variables specified as factors (see the example below). (If the user wishes to obtain only the direct correlation between motion and functional connectivity, then no formula or covariates file is necessary.) For example, to control for the participants' age and sex when computing motion effects, prepare a second file (<confound>) containing the same identifiers as the first, with additional columns for each of the covariates to be considered. In the example below, age is defined in months and sex is coded as a binary variable:

```
id0,id1,age,sex
ACC,001,217,0
ACC,002,238,1
ACC,003,238,1
DSQ,001,154,0
CAT,001,176,1
```

If this file is saved as `sample-covariates.csv`, then call `qcfc.R` as above, with the additional argument <confound> set to `sample-covariates.csv` and the additional argument <conformula> set to:

```
'age+factor(sex)'
```

Note that the categorical variable `sex` is specified as a `factor`. If the dataset contains repeated measures (e.g., multiple scans from the same subject), then the subject identifier can be included in the model specification `conformula` as a random intercept:

```
'age+factor(sex)+(1|id0)'
```

When the output path is specified as `-o <output root>`, as in the example call above, the outputs of `qcfc.R` include the following:

Output path	Output description
<out>.txt	A matrix containing the QC-FC correlation for each edge in the input matrix
<out>_thr.txt	The above matrix, thresholded to include only significant edges, can be used to plot glass brain visualizations of significant edges, for instance, using BrainNetViewer
<out>_absMedCor.txt	The absolute median QC-FC correlation over all edges
<out>_nSigEdges.txt	The number and percentage of edges with significant QC-FC correlations
<out>_pctSigEdges.txt	
<out>.svg	A visualization of the QC-FC distribution (Fig. 3)

in images processed through a high-performance denoising stream, the residual artifact will be markedly mitigated in comparison with the artifact present in a time series that has been processed minimally or by using a less effective stream (see Fig. 3 for an illustration). If GSR is included in the denoising model, then denoised connectomes can be expected to feature a considerable fraction of negatively weighted connections, along with a relatively symmetric and approximately zero-centered distribution of connection weights. The diagnostics produced in the performance assessment stages of the protocol can be used to determine whether the final result of denoising is appropriate. Example results from subject-level and group-level diagnostic steps are summarized in Figs. 2 and 3, respectively. Although the details are beyond the scope of this protocol, it should be noted that XCP uses these denoised data as part of additional modules for generation of seed-based connectivity maps¹, network partitions^{83,84}, regional homogeneity maps⁸⁵, and amplitude of low-frequency fluctuation maps⁸⁶.

Code availability

A code for all stages of the protocol is freely and publicly available online under the MIT license (<https://github.com/PennBBL/xcEngine>, <https://doi.org/10.5281/zenodo.1326936>).

References

1. Biswal, B., Yetkin, F. Z., Haughton, V. M. & Hyde, J. S. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* **34**, 537–541 (1995).
2. Biswal, B. B. et al. Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. USA* **107**, 4734–4739 (2010).
3. Power, J. D. et al. Functional network organization of the human brain. *Neuron* **72**, 665–678 (2011).
4. Yeo, B. T. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
5. Raichle, M. E. et al. A default mode of brain function. *Proc. Natl. Acad. Sci. USA* **98**, 676–682 (2001).
6. Sharma, A. et al. Common dimensional reward deficits across mood and psychotic disorders: a connectome-wide association study. *Am. J. Psychiatry* **174**, 657–666 (2017).
7. Gu, S. et al. Emergence of system roles in normative neurodevelopment. *Proc. Natl. Acad. Sci. USA* **112**, 13681–13686 (2015).
8. Dosenbach, N. U. F. et al. Prediction of individual brain maturity using fMRI. *Science* **329**, 1358–1361 (2011).
9. Fair, D. A. et al. The maturing architecture of the brain's default network. *Proc. Natl. Acad. Sci. USA* **105**, 4028–4032 (2008).
10. Satterthwaite, T. D. et al. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. *Neuroimage* **60**, 623–632 (2012).
11. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142–2154 (2012).
12. Van Dijk, K. R. A., Sabuncu, M. R. & Buckner, R. L. The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage* **59**, 431–438 (2012).
13. Power, J. D., Plitt, M., Laumann, T. O. & Martin, A. Sources and implications of whole-brain fMRI signals in humans. *Neuroimage* **146**, 609–625 (2016).
14. Fair, D. A. et al. Development of distinct control networks through segregation and integration. *Proc. Natl. Acad. Sci. USA* **104**, 13507–13512 (2007).
15. Friston, K., Williams, S., Howard, R., Frackowiak, R. S. J. & Turner, R. Movement-related effects in {fMRI} time-series. *Magn. Reson. Med.* **35**, 346–355 (1996).
16. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90–101 (2007).
17. Satterthwaite, T. D. et al. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* **64**, 240–256 (2013).
18. Power, J. D. et al. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84**, 320–341 (2014).
19. Muschelli, J. et al. Reduction of motion-related artifacts in resting state fMRI using aCompCor. *Neuroimage* **96**, 22–35 (2014).
20. Pruim, R. H. R. et al. ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* **112**, 267–277 (2015).
21. Salimi-Khorshidi, G. et al. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* **90**, 449–468 (2014).
22. Ciric, R. et al. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage* **154**, 174–187 (2017).
23. Parkes, L., Fulcher, B. D., Yucel, M. & Fornito, A. An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *Neuroimage* **171**, 414–436 (2018).
24. Burgess, G. C. et al. Evaluation of denoising strategies to address motion-correlated magnetic resonance imaging data from the Human Connectome Project. *Brain Connect.* **6**, 414–436 (2016).
25. Yan, C. G. et al. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *Neuroimage* **76**, 183–201 (2013).
26. Macey, P. M., Macey, K. E., Kumar, R. & Harper, R. M. A method for removal of global effects from fMRI time series. *Neuroimage* **22**, 360–366 (2004).
27. Fox, M. D., Zhang, D., Snyder, A. Z. & Raichle, M. E. The global signal and observed anticorrelated resting state brain networks. *J. Neurophysiol.* **101**, 3270–3283 (2009).
28. 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).
29. Greene, D. J. et al. Behavioral interventions for reducing head motion during MRI scans in children. *Neuroimage* **171**, 234–245 (2018).
30. Faraji-Dana, Z., Tam, F., Chen, J. J. & Graham, S. J. A robust method for suppressing motion-induced coil sensitivity variations during prospective correction of head motion in fMRI. *Magn. Reson. Imaging* **34**, 1206–1219 (2016).

31. Bright, M. G. & Murphy, K. Removing motion and physiological artifacts from intrinsic BOLD fluctuations using short echo data. *Neuroimage* **64**, 526–537 (2013).
32. Kundu, P. et al. Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proc. Natl. Acad. Sci. USA* **110**, 16187–16192 (2013).
33. Dosenbach, N. U. F. et al. Real-time motion analytics during brain MRI improve data quality and reduce costs. *Neuroimage* **161**, 80–93 (2017).
34. Power, J. D. et al. Ridding fMRI data of motion-related influences: removal of signals with distinct spatial and physical bases in multiecho data. *Proc. Natl. Acad. Sci. USA* **115**, E2105–E2114 (2018).
35. Satterthwaite, T. D. et al. Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth. *Neuroimage* **83**, 45–57 (2013).
36. Zeng, L.-L. et al. Neurobiological basis of head motion in brain imaging. *Proc. Natl. Acad. Sci. USA* **111**, 6058–6062 (2014).
37. Van Essen, D. C. et al. The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62–79 (2013).
38. Satterthwaite, T. D. et al. Linked sex differences in cognition and functional connectivity in youth. *Cereb. Cortex* **25**, 2383–2394 (2015).
39. Xia, C. H. et al. Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat. Commun.* **9**, 3003 (2018).
40. Murphy, K. & Fox, M. D. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *Neuroimage* **154**, 169–173 (2017).
41. Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B. & Bandettini, P. A. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *Neuroimage* **44**, 893–905 (2009).
42. Chai, X. J., Castañán, A. N., Öngür, D. & Whitfield-Gabrieli, S. Anticorrelations in resting state networks without global signal regression. *Neuroimage* **59**, 1420–1428 (2012).
43. Saad, Z. S. et al. Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. *Brain Connect.* **2**, 25–32 (2012).
44. Yan, C.-G., Craddock, R. C., He, Y. & Milham, M. P. Addressing head motion dependencies for small-world topologies in functional connectomics. *Front. Hum. Neurosci.* **7**, 910 (2013).
45. Power, J. D., Laumann, T. O., Plitt, M., Martin, A. & Petersen, S. E. On global fMRI signals and simulations. *Trends Cogn. Sci.* **21**, 911–913 (2017).
46. Yang, G. J. et al. Altered global brain signal in schizophrenia. *Proc. Natl. Acad. Sci. USA* **111**, 7438–7443 (2014).
47. Hahamy, A. et al. Save the global: global signal connectivity as a tool for studying clinical populations with functional magnetic resonance imaging. *Brain Connect.* **4**, 395–403 (2014).
48. Carone, D. et al. Impact of automated ICA-based denoising of fMRI data in acute stroke patients. *NeuroImage Clin.* **16**, 23–31 (2017).
49. Tremblay, M., Tam, F. & Graham, S. J. Retrospective coregistration of functional magnetic resonance imaging data using external monitoring. *Magn. Reson. Med.* **53**, 141–149 (2005).
50. Cox, R. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
51. Avants, B. B. et al. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* **54**, 2033–2044 (2011).
52. Smith, S. M. et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, 208–219 (2004).
53. Lemieux, L., Salek-Haddadi, A., Lund, T. E., Laufs, H. & Carmichael, D. Modelling large motion events in fMRI studies of patients with epilepsy. *Magn. Reson. Imaging* **25**, 894–901 (2007).
54. Klein, A. et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* **46**, 786–802 (2009).
55. Tustison, N. J. et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *Neuroimage* **99**, 166–179 (2014).
56. Satterthwaite, T. D. et al. Motion artifact in studies of functional connectivity: characteristics and mitigation strategies. *Hum. Brain Mapp.* <https://doi.org/10.1002/hbm.23665> (2017).
57. Power, J. D. A simple but useful way to assess fMRI scan qualities. *Neuroimage* **154**, 150–158 (2017).
58. Zuo, X.-N. et al. An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci. Data* **1**, 140049 (2014).
59. Carp, J. Optimizing the order of operations for movement scrubbing: comment on Power et al. *Neuroimage* **76**, 436–438 (2013).
60. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Steps toward optimizing motion artifact removal in functional connectivity MRI; a reply to Carp. *Neuroimage* **76**, 439–441 (2013).
61. Sikka, S. et al. Towards automated analysis of connectomes: the configurable pipeline for the analysis of connectomes (C-PAC). Abstract presented at Neuroinformatics 2013; 27–29 August 2013; Karolinska Institute, Stockholm. <https://doi.org/10.3389/conf.fninf.2014.08.00117> (2012).
62. Shehzad, Z. et al. The preprocessed connectomes project quality assessment protocol—a resource for measuring the quality of MRI data. Abstract presented at Neuroinformatics 2015; 20–22 August 2015; Cairns, Australia. <https://doi.org/10.3389/conf.fnins.2015.91.00047> (2015).

63. Whitfield-Gabrieli, S. & Nieto-Castanon, A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect.* **2**, 125–141 (2012).
64. Zuo, X. N. et al. Toward reliable characterization of functional homogeneity in the human brain: pre-processing, scan duration, imaging resolution and computational space. *Neuroimage* **65**, 374–386 (2013).
65. Esteban, O. et al. FMRIPrep: a robust preprocessing pipeline for functional MRI. Preprint at <https://www.biorxiv.org/content/early/2018/07/24/306951> (2018).
66. Rosen, A. et al. Quantitative assessment of structural image quality. *Neuroimage* **169**, 407–418 (2018).
67. Gordon, E. M. et al. Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex* **26**, 288–303 (2016).
68. Cammoun, L. et al. Mapping the human connectome at multiple scales with diffusion spectrum MRI. *J. Neurosci. Methods* **203**, 386–397 (2012).
69. Honnorat, N. et al. GraSP: Geodesic Graph-based Segmentation with Shape Priors for the functional parcellation of the cortex. *Neuroimage* **106**, 207–221 (2015).
70. Schaefer, A. et al. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* 1–20. <https://doi.org/10.1093/cercor/bhx179> (2017).
71. Gorgolewski, K. J. et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* **3**, 1–9 (2016).
72. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
73. Power, J. D., Plitt, M., Kundu, P., Bandettini, P. A. & Martin, A. Temporal interpolation alters motion in fMRI scans: magnitudes and consequences for artifact detection. *PLoS ONE* **12**, 1–20 (2017).
74. Nichols, T. E. Notes on creating a standardized version of DVARS. Preprint at <https://arxiv.org/abs/1704.01469> (2017).
75. Afyouni, S. & Nichols, T. E. Insight and inference for DVARS. *Neuroimage* **172**, 291–312 (2018).
76. Patel, A. X. et al. A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *Neuroimage* **95**, 287–304 (2014).
77. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72 (2009).
78. Jo, H. J. et al. Effective preprocessing procedures virtually eliminate distance-dependent motion artifacts in resting state FMRI. *J. Appl. Math.* <https://doi.org/10.1155/2013/935154> (2013).
79. Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S. & Petersen, S. E. Intrinsic and task-evoked network architectures of the human brain. *Neuron* **83**, 238–251 (2014).
80. Lomb, N. R. Least-squares frequency analysis of unequally spaced data. *Astrophys. Space Sci.* **39**, 447–448 (1976).
81. Hallquist, M. N., Hwang, K. & Luna, B. The nuisance of nuisance regression: spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures functional connectivity. *Neuroimage* **82**, 208–225 (2013).
82. Smith, S. M. & Brady, J. M. SUSAN—a new approach to low level image processing. *Int. J. Comput. Vis.* **23**, 45–78 (1997).
83. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P1008 (2008).
84. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99**, 7821–7826 (2002).
85. Zang, Y., Jiang, T., Lu, Y., He, Y. & Tian, L. Regional homogeneity approach to fMRI data analysis. *Neuroimage* **22**, 394–400 (2004).
86. Yu-Feng, Z. et al. Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain Dev.* **29**, 83–91 (2007).
87. Satterthwaite, T. D. et al. Neuroimaging of the Philadelphia Neurodevelopmental Cohort. *Neuroimage* **86**, 544–553 (2014).
88. Pruim, R. H. R., Mennes, M., Buitelaar, J. K. & Beckmann, C. F. Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI. *Neuroimage* **112**, 278–287 (2015).

Acknowledgements

This work was supported by grants from the National Institutes of Health: R01MH107703 (T.D.S.), R01MH112847 (T.D.S.), R21MH106799 (D.S.B. and T.D.S.), R01EB022573 (C.D.), and R01MH101111 (D.H.W.); and the Lifespan Brain Institute at Penn/CHOP. D.S.B. acknowledges support from the John D. and Catherine T. MacArthur Foundation, the Alfred P. Sloan Foundation, the Army Research Laboratory through contract number W911NF1020022, the Army Research Office through contract numbers W911NF1410679 and W911NF1610474, and the National Institute of Health (grants R01DC00920911, R01HD086888, R01MH107235, R01MH109520, and R01NS099348). The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

Author contributions

T.D.S., D.H.W., and R.C. developed and designed the protocol with numerous contributions from the functional neuroimaging community. R.C. developed the software implementation of the protocol. R.C., A.F.G.R., M.C., and A.A. developed the associated software libraries. M.C. and A.A. built Docker and Singularity containers for the associated software libraries. G.E. reviewed and tested the code and implemented the software on the Image Processing Portal. G.E., P.A.C., D.S.B., and C.D. provided consultation and guidance for

methodological implementation and interpretation of results. R.C. and T.D.S. wrote the manuscript. R.C. prepared the figures and tables. All authors reviewed and revised the manuscript.

Data availability

Sample data and results for testing the denoising and benchmarking protocol are available publicly at FigShare under the MIT license (<https://doi.org/10.6084/m9.figshare.6225968.v1>).

Competing interests

The authors declare no competing interests.

Additional information

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to T.D.S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published online: 16 November 2018

Related links

Key references using this protocol

Satterthwaite, T. D. et al. *Neuroimage* **64**, 240–256 (2013): <https://doi.org/10.1016/j.neuroimage.2012.08.052>
 Ciric, R. et al. *Neuroimage* **154**, 174–187 (2017): <https://doi.org/10.1016/j.neuroimage.2017.03.020>
 Satterthwaite, T. D. et al. *Neuroimage* **83**, 45–57 (2013): <https://doi.org/10.1016/j.neuroimage.2013.06.045>