



# NIH Public Access

## Author Manuscript

*Neuroimage*. Author manuscript; available in PMC 2015 January 01.

Published in final edited form as:  
*Neuroimage*. 2014 January ; 84: . doi:10.1016/j.neuroimage.2013.08.048.

## Methods to detect, characterize, and remove motion artifact in resting state fMRI

**Jonathan D Power<sup>a</sup>, Anish Mitra<sup>a</sup>, Timothy O Laumann<sup>a</sup>, Abraham Z Snyder<sup>a,b</sup>, Bradley L Schlaggar<sup>a,b,c,d</sup>, and Steven E Petersen<sup>a,b,d,f,g,h</sup>**

Jonathan D Power: powerj@wusm.wustl.edu; Anish Mitra: mitraa@wusm.wustl.edu; Timothy O Laumann: laumannt@wusm.wustl.edu; Abraham Z Snyder: avi@npg.wustl.edu; Bradley L Schlaggar: schlaggarb@neuro.wustl.edu; Steven E Petersen: sep@npg.wustl.edu

<sup>a</sup>Dept. of Neurology, Washington University School of Medicine

<sup>b</sup>Dept. of Radiology, Washington University School of Medicine

<sup>c</sup>Dept. of Pediatrics, Washington University School of Medicine

<sup>d</sup>Dept. of Anatomy & Neurobiology, Washington University School of Medicine

<sup>e</sup>Dept. of Psychiatry, Washington University School of Medicine

<sup>f</sup>Dept. of Psychology, Washington University in Saint Louis

<sup>g</sup>Dept. of Neurosurgery, Washington University School of Medicine

<sup>h</sup>Dept. of Biomedical Engineering, Washington University in Saint Louis

### Abstract

Head motion systematically alters correlations in resting state functional connectivity fMRI (RSFC). In this report we examine impact of motion on signal intensity and RSFC correlations. We find that motion-induced signal changes (1) are often complex and variable waveforms, (2) are often shared across nearly all brain voxels, and (3) often persist more than 10 seconds after motion ceases. These signal changes, both during and after motion, increase observed RSFC correlations in a distance-dependent manner. Motion-related signal changes are not removed by a variety of motion-based regressors, but are effectively reduced by global signal regression. We link several measures of data quality to motion, changes in signal intensity, and changes in RSFC correlations. We demonstrate that improvements in data quality measures during processing may represent cosmetic improvements rather than true correction of the data. We demonstrate a within-subject, censoring-based artifact removal strategy based on volume censoring that reduces group differences due to motion to chance levels. We note conditions under which group-level regressions do and do not correct motion-related effects.

---

© 2013 Elsevier Inc. All rights reserved.

**Corresponding Author:** Jonathan D Power, **Corresponding Author's Contact Information:** Wash Univ Sch Med Dept of Neurol, 660 S Euclid Ave, Box 8111, St. Louis, MO 63110 USA, powerj@wusm.wustl.edu, Phone: +1 314 362 3317, Fax: +1 314 362 2186.  
**Institutional Addresses:**

Washington University School of Medicine in St. Louis, 660 S. Euclid Ave., St. Louis, MO 63110  
Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

### Conflict of Interest Statement:

The authors have no conflicts of interest to report.

## Keywords

Resting state; functional connectivity; MRI; artifact; motion; movement

## Introduction

Head motion correction has become a prominent concern in the field of resting state functional connectivity fMRI (RSFC), especially for investigators studying pediatric, clinical, or elderly populations. The renewed attention to head motion stems in part from the realization that even small amounts of movement can produce spurious but spatially structured patterns in functional connectivity (Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012). The structured artifact arises because motion adds spurious variance to ‘true’ timeseries, and this spurious variance is most similar at nearby voxels. Consequently, correlations between BOLD timeseries are spuriously increased across all voxels, but are most increased between nearby voxels.

At present, all post-hoc subject-level processing strategies that have been examined have incompletely removed motion artifact, as evidenced by residual cross-subject dependence of RSFC measures on summary motion measures, or by distance-dependent changes seen between high- and low-motion scans or subjects (Bright and Murphy, 2013; Mowinckel et al., 2012; Power et al., 2012; Satterthwaite et al., 2013; Satterthwaite et al., 2012; Van Dijk et al., 2012; Yan et al., 2013). This is true for processing strategies that have used large numbers of motion regressors (Satterthwaite et al., 2013; Yan et al., 2013), global signal regression (Power et al., 2012; Satterthwaite et al., 2013; Satterthwaite et al., 2012; Van Dijk et al., 2012; Yan et al., 2013), voxel-specific motion regressors (Satterthwaite et al., 2013; Yan et al., 2013), ICA-based nuisance removal (Mowinckel et al., 2012; Satterthwaite et al., 2012; Tyszka et al., 2013), or extensive modeling of physiological noise (Bright and Murphy, 2013). To eliminate motion-related effects, further corrections at the subject or group level are needed.

Group-level correction has been implemented by regressing a summary quality control (QC) measure for each subject from each set of correlations (or outcomes) across subjects (Satterthwaite et al., 2012; Van Dijk et al., 2012; Yan et al., 2013). This approach effectively suppresses spurious motion-related differences across subjects. However, it only ‘corrects’ the data to the extent that the assumed relationship between a summary QC measure and spurious influence on outcomes exists. If only linear effects exist, then linear regression should completely correct spurious differences. If any non-linear (or unmodeled) effects exist, the correction will be incomplete. Also, since summary QC measures may covary with factors of interest, such as age or diagnosis, group-level regression may remove the very effects that a study seeks to identify. For these reasons it would be desirable to improve subject-level motion correction to the point where group-level regressions are not necessary.

This paper aims to develop methods of motion correction that, applied at the subject level, eliminate the need for group-level corrections. To develop these methods it is necessary to better understand the properties of motion artifact. The paper is therefore composed of 3 parts. Part I aims to create a fuller understanding of motion-related artifact, focusing on the types of signal intensity changes it produces. Part II examines how motion impacts RSFC correlations. Part III takes the lessons of the previous parts and describes processing strategies that generate results in which motion-related influences are no longer detectable.

To orient the reader, we preview several main results here. In Part I, we observe that motion-induced signal changes are highly variable and may persist tens of seconds after a motion.

We demonstrate the modest efficacy of a variety of motion-related regressors but the high efficacy of global signal regression in reducing such intensity changes. In Part II we find that RSFC correlations are systematically impacted by volumes acquired during and up to ~10 seconds after movement. We find that multiple QC measures can identify compromised volumes, but that reductions in outlying QC values over processing are partially cosmetic, i.e., some ‘improved’ volumes continue to harbor motion-related effects. We propose methods to identify values at which a QC measure begins to index spurious changes in RSFC correlations. In Part III we demonstrate that censoring approaches, when applied throughout a processing stream, can reduce spurious motion-related group differences to chance levels. Table 1 outlines several of the chief objectives and findings of the paper.

Several remarks about the scope of this paper will help to frame the analyses to come.

First, this paper is concerned with post-hoc correction of data. Although we are optimistic about improvements in MRI sequences and techniques at data acquisition with regard to motion correction (e.g., (Bright and Murphy, 2013; Kundu et al., 2012)), our focus is on data that have been already acquired.

Second, we have neither external measures of motion nor physiological measures such as respiratory, cardiac, or end-tidal CO<sub>2</sub>. An absence of such measures is characteristic of many existing (and likely future) datasets, including many publicly available datasets, and our data are representative of many datasets that the field would like to utilize. For artifact reduction, we are limited to data-driven methods such as (1) removal of signals derived from matrix decomposition (e.g., ICA), (2) removal of signal variance associated with various brain compartments (e.g., white matter signal), or (3) entirely discarding problematic data (volume censoring). Matrix decomposition techniques are powerful tools to isolate nuisance-related signals but the extent of their success depends upon the correct classification of the resulting components into signals of interest or non-interest. This paper will focus on nuisance regressions (of motion estimates and brain compartment signals) and on the censoring approach.

Third, given the many approaches to processing RSFC data, we have tried to make our results as generally informative as possible. The first section of the paper describes data before artifact reduction and thus provides a picture of the types of problems that motion may create in any dataset. Much of the methodology in this paper is applicable to any BOLD dataset, regardless of how it was acquired or prepared. Several of our analyses use censoring to characterize motion artifact, but other analyses do not. Most analyses are performed with and without global signal regression (GSR), since GSR is both a contested step in processing (Fox et al., 2009; Murphy et al., 2009) and an effective procedure for removing motion artifact (Satterthwaite et al., 2013; Yan et al., 2013). In this manner, our findings should hopefully apply to a wide variety of processing streams.

## Methods

### Subjects

This paper studies 160 healthy subjects in 4 cohorts of 40 subjects: 3 adult cohorts (high, medium, and low motion, binned by mean framewise displacement (FD)) and a child cohort. The datasets were selected from a larger pool for their ability to undergo various volume censoring strategies while still retaining at least 125 volumes of data (~ 5 minutes or more). The cohorts are sex-matched. The high-motion adult cohort and child cohort are matched on all QC measures and are significantly different from the medium- and low-motion adult cohorts on all QC measures (see Figure S1, Table S1).

All subjects were recruited from the Washington University campus and the surrounding community. All subjects were right-handed, native English speakers, reported no history of neurological or psychiatric disease, and were not on psychotropic medications. All subjects (or their guardians) gave informed consent and were compensated for their participation in accord with institutional and national guidelines.

### Data Collection

All subjects were scanned in the same Siemens MAGNETOM Trio 3T scanner with a Siemens 12 channel Head Matrix Coil (Erlangen, Germany). A T1-weighted sagittal MP-RAGE was obtained (TE = 3.06 ms, TR-partition = 2.4 ×, TI = 1000ms, flip angle = 8°, 127 slices with 1×1×1 mm voxels). A T2-weighted turbo spin echo structural image (TE = 84 ms, TR = 6.8 s, 32 slices with 2×1×4 mm voxels) in the same anatomical plane as the BOLD images was also obtained for use in image registration.

RSFC BOLD runs were obtained from subjects visually fixating a white crosshair on a black background. Subjects were asked to stay still, to stay awake, and to watch the crosshair. Functional images were obtained using a Siemens product gradient echo echo-planar sequence. The 160 subjects were pooled from different studies, resulting in slight differences in the parameters for BOLD acquisition, noted below. A representative set of parameters is: TE = 27 ms, flip angle = 90°, 32 contiguous interleaved 4 mm axial slices, with in-plane resolution = 4×4 mm. The TR lengths varied slightly: 158 subjects have TR = 2.5 s, 2 subjects have TR = 2.2 s. Most functional data were acquired in runs of 132 volumes, though some runs were longer or slightly shorter depending on the original study from which a subject was taken. Most subjects contributed 2 or more runs of data. Accordingly, the number of volumes available, typically several hundred, varied across subjects (mean ± s.d: 346 ± 136; range: 184-724; see Table S1).

### fMRI Pre-processing

For the purposes of this paper we distinguish between ‘fMRI pre-processing’ and ‘functional connectivity processing’ (Figure 1). This distinction separates relatively common steps taken by many investigators to preprocess fMRI data for any purpose (e.g., rigid body motion correction) from the highly variable steps that can be taken to prepare data for functional connectivity analyses.

For fMRI pre-processing, functional images underwent (i) slice-time correction, i.e., sinc interpolation to temporally align each slice to the start of each volume, (ii) rigid body realignment to correct for head movement within and across runs, and (iii) within-run intensity normalization, that is, scaling the intensity across all voxels and all magnetization steady-state volumes to achieve a mode value of 1000. In all Figures, BOLD data are presented in a mode 1000 scale (10 units = 1% BOLD). Atlas transformation of the functional data was computed for each individual via the T2-weighted and MP-RAGE scans. Each run was then resampled in atlas space on an isotropic 3 mm grid combining realignment and atlas transformation in a single interpolation (Shulman et al., 2010).

### Functional Connectivity (RSFC) Processing

After fMRI pre-processing, further steps were taken to prepare these data for functional connectivity analysis (Figure 1). Data from various steps in processing will be illustrated in the paper. These steps included (i) demeaning and detrending across each run, (ii) regression of nuisance variables across all runs (various regressors were used and will be described below), (iii) frequency filtering of the data using a zero-phase 2<sup>nd</sup> order Butterworth filter with a pass-band of 0.009 to 0.08 Hz, and (iv) spatial blurring using a Gaussian filter with 6 mm FWHM. The data presented in Parts I and II underwent this procedure. In Part III, this

procedure was performed to yield QC measures, then temporal masks were formed using the QC measures, and then the procedure was re-performed with volume censoring and data replacement by interpolation.

## Regressors

Various combinations of nuisance regressors were used in the multiple regressions and are described for each analysis. Motion estimates ( $R = [X \ Y \ Z \ \text{pitch} \ \text{yaw} \ \text{roll}]$ ) were the detrended realignment estimates from fMRI pre-processing. Their derivatives ( $R'$ ) and squares ( $R^2$ ) were also used as regressors. Our lab has historically used  $R$  and  $R'$  as nuisance regressors ( $[R \ R'] = 12$  motion related regressors). This paper also examines two sets of motion regressors derived by Volterra expansion (Friston et al., 1996):  $[R \ R^2 \ R_{t-1} \ R_{t-1}^2]$ , where  $t$  and  $t-1$  refer to the current and immediately preceding timepoint (24 motion-related regressors) and  $[R \ R^2 \ R_{t-1} \ R_{t-1}^2 \ R_{t-2} \ R_{t-2}^2]$  (36 motion-related regressors). Note that the 24-motion-parameter expansion is the same one used in (Satterthwaite et al., 2013) and (Yan et al., 2013), but that the 36-motion-parameter set in this report is not the same as the 36 parameters used by Satterthwaite and colleagues (their 36 regressors were the 24 motion-related regressors and 12 tissue-based regressors). Tissue-based signals were also used as nuisance regressors and were calculated as the average signal across voxels within a particular spatial mask: an eroded ventricular mask for the CSF signal (CSF or V), an eroded white matter mask for the white matter signal (WM), and the whole-brain mask for the global signal (GS). In all cases, when a tissue-based signal was used as a regressor, its first derivative, computed by backwards differences, was also used.

## Regressions using temporal masks

Temporal masks are incorporated into demeaning and detrending and multiple regressions in Part III. If data were censored during demeaning and detrending or multiple regression, the following procedure was used: (i) ‘bad’ timepoints were censored from the regressors and BOLD data, (ii) the remaining ‘good’ regressors were standardized (zero-mean, unit variance) and detrended, (iii) a least-squares fit of ‘good’ regressors to the ‘good’ data generated beta values, (iv) regressors from all timepoints (‘good’ and ‘bad’) underwent the same transformation that defined the ‘good’ standardized regressors, (v) the ‘good’ betas were applied to regressors from all (‘good’ and ‘bad’) timepoints to generate modeled signal values at all timepoints, and (vi) residuals were calculated for all timepoints as observed minus modeled BOLD values. Thus, only ‘good’ data contributed to betas but residuals were calculated for all timepoints, yielding continuous timeseries. The betas and residuals at ‘good’ timepoints generated by this procedure are theoretically identical to those obtained using ‘spike regressors’ (Lemieux et al., 2007).

## Interpolation using temporal masks

In Part III, potentially compromised data were replaced after the multiple regression but prior to frequency filtering. A least-squares spectral decomposition of the uncensored (‘good’) data was performed and this decomposition was used to reconstitute data at censored (‘bad’) timepoints. To compute the frequency content of uncensored data, we applied a least squares spectral analysis adapted for non-uniformly sampled data, as described in (Mathias et al., 2004), using a method based on the Lomb-Scargle periodogram (Lomb, 1976). A more detailed and formal description is provided in the Supplemental Materials. Thus, the ‘good’ data defined the frequency characteristics of signals that then replace the ‘bad’ data. Figure S2 illustrates this procedure for a synthetic signal. The replaced timepoints almost always have values closer to the signal mean than the original data, and they thus spread less signal into adjacent ‘good’ timepoints during frequency filtering (Carp, 2013). These interpolated timepoints are then re-censored following frequency filtering.

## Frequency filtering

Frequency filtering was performed using a first-order Butterworth FIR filter with passband 0.009 to 0.08 Hz in the forward and reverse directions. After filtering, the first and last 15 TRs of each run were ignored.

## Quality Control (QC) Measures

The QC measures employed here are largely those used in (Power et al., 2012). Figure 2 shows QC measures for 2 subjects. For all realignment-estimate-derived calculations, rotational displacements were converted to translational displacements by projection to the surface of a 50 mm radius sphere.

## RMS motion and RMS d/dt motion

These are the root mean squared values of the detrended realignment estimates and their derivatives across all timepoints.

## Framewise displacement (FD)

This measure indexes the movement of the head from one volume to the next, and is calculated as the sum of the absolute values of the differentiated realignment estimates (by backwards differences) at every timepoint (Power et al., 2012). FD for the first volume of a run is 0 by convention. The purpose of this measure is to index head movement, not to precisely calculate or model it.

## Absolute displacement

(calculated separately for rotation and translation): These measures index the absolute displacement of the head from the origin position at every timepoint. The translational absolute displacement is the sum of the absolute values of the X, Y, and Z estimates for a given volume. The rotational version is the sum of the absolute values of the displacements in pitch, yaw, and roll. The purpose of this measure is to index head absolute position, not to precisely calculate or model it.

## DVARS (DV)

This measure indexes the change in signal intensity from one volume to the next, and is calculated as the root mean square value of the differentiated BOLD timeseries (by backwards differences) within a spatial mask at every timepoint (Smyser et al., 2010). DV for the first volume of a run is set to zero by convention. This paper usually presents DV calculated over the whole-brain mask but it can be calculated over any collection of voxels. Gray matter DV ( $DV_{GM}$ ), which closely parallels whole-brain DV ( $DV_{GS}$ ), is presented at several points in the manuscript where indicated (e.g., after global signal regression has made it pointless to plot mean global signal but still informative to plot mean gray matter signal, we use  $DV_{GM}$  instead of  $DV_{GS}$  so that the same set of voxels is being examined.)

## Standard deviation (SD)

This measure is the standard deviation (across space) of the BOLD signal across all voxels within a spatial mask at every timepoint. As for DV, SD from a whole-brain mask is typically presented unless specifically indicated.

## Spatial Masks

Each subject's MP-RAGE underwent segmentation by FreeSurfer (version 5.0). Segments corresponding to the lateral ventricles, the white matter, the gray matter, and all within-brain

voxels were extracted. 5 versions of the white matter and ventricular masks were formed: uneroded, and erosions 1–4 times. All masks were resliced to 3 mm isotropic dimensions to match the BOLD data. Maximal erosion (4 $\times$ ) masks were preferred; if maximal erosion yielded empty masks, lesser erosions were progressively considered until a mask with qualifying voxels was found. This relaxation occurred infrequently for white matter masks. Erosions of 1 were often used for CSF masks. These masks were used to extract nuisance signals, to calculate DV and SD within the masks, and to define the voxel timeseries displayed in Part I.

### Regions of Interest

264 regions of interest (ROIs) are examined in this paper, identical to those used in (Power et al., 2012). These ROIs represent our current best estimates of the center coordinates of brain areas and nuclei, and derive from the results of many task fMRI meta-analyses and a resting state parcellation strategy (Cohen et al., 2008). ROIs are modeled as 10 mm diameter spheres centered on the coordinates reported in (Power et al., 2011). ROI timecourses are calculated as the average value across voxels within the ROI.

### Correlation Calculations

All correlations are calculated as the Pearson correlation between timeseries. Arithmetic and statistical calculations use Fisher z-transformed values. Reported correlations are Pearson's r.

### Methodological changes from our previous report

Three methodological differences from our previous report on motion artifact (Power et al., 2012) should be noted.

First, the order of processing operations has changed: we now regress nuisance signals, then perform frequency filtering, then perform spatial blurring. Previously we had regressed nuisance signals after frequency filtering. Either order of operations is acceptable, but if regression takes place after frequency filtering, the nuisance regressors should also be filtered so that residuals maintain the proper frequency content (i.e., so that high-frequency regressors do not reintroduce high-frequency content into low-frequency data). In our initial report, although our tissue-based nuisance signals were frequency filtered, our realignment estimates had not been filtered and therefore could have reintroduced high-frequency signal related to motion into our data. Our observations and conclusions are essentially unchanged after altering our procedure.

Second, the slice-timing correction in our initial report was incorrect because we were unaware that on our Siemens MAGNETOM Trio, A Tim System 3T scanner, interleaved acquisitions began with the second slice (but only if the total number of slices is even). Compensation for asynchronous slice timing is now correct. As will be seen, our previous results and conclusions are unchanged after implementing proper slice timing correction.

Third, we now use a different and much larger group of subjects than in our previous report. These subjects, though they clearly exhibit motion artifact, are a selective set of subjects: they were selected out of a larger pool of subjects for the ability of their data to withstand various volume censoring strategies, including very stringent ones, while retaining at least 5 minutes of data. Our previous dataset could not have withstood such measures and was a higher-motion dataset.

## Results

### Part I: Understanding BOLD motion artifact and its relation to QC measures

**The overall processing strategy**—An overview of the processing strategy is shown in Figure 1 (discussed in Methods). This section does not involve scrubbing or iterative processing. This section describes how motion manifests in signal intensity, the extent to which regressions remove motion-related signal, and how QC measures reflect movement-associated signal changes.

In this section, Figures 2–7 contain 14 single-subject illustrations of motion effects in BOLD data. We focus on single-subject data because motion’s effects are highly variable and are therefore incompletely communicated by peri-motion histograms or modeling. Importantly, to impact correlations between timeseries, motion-related signal changes do not need to be highly stereotyped, they only need to be shared across voxels. Highly variable motion-related changes can alter correlations just as well as stereotyped motion-related changes. An understanding of the variety and magnitude of signal changes that motion can produce is quickly developed by studying individual subjects. Important aspects of Figures 2–7 are presented for each subject used in this report in a Supplemental Cohort Illustration. We encourage readers to consult this resource to develop a fuller understanding of motion-related signal changes. The Supplemental Cohort Illustration is available at [www.nil.wustl.edu/labs/petersen/Resources.html](http://www.nil.wustl.edu/labs/petersen/Resources.html).

**Quality control measures**—Figure 2 shows several QC measures in 2 subjects (subjects 1–40 are children, 41–80 are high-motion adults, 81–120 are medium-motion adults, and 121–160 are low-motion adults). Vertical lines denote boundaries between concatenated runs. Spikes in the red FD traces indicate head movement, and elevations in the dotted absolute displacement traces indicate that the head is shifted away from the origin. The subject at left is generally still except for a portion of the 3<sup>rd</sup> run, whereas the subject at right moves more frequently. These subjects are representative of the low- and high-motion adult cohorts (see Figure S1 for mean FD distributions in each cohort).

The bottom two panels show signals and signal-derived QC measures, calculated within a whole brain mask in demeaned and detrended data, prior to nuisance regression. Other than realignment, nothing has been done to counter motion-related artifact. The correspondence of the blue DV trace to FD is evident (DV-FD correlation across subjects (mean  $\pm$  s.d.):  $0.69 \pm 0.38$ ). DV does not so closely reflect absolute head position (DV-abs. trans.:  $0.32 \pm 0.26$ ; DV-abs. rot:  $0.25 \pm 0.23$ ). The bottom panel shows the global signal and the SD trace. The SD trace bears considerable similarity to the FD signal (SD-FD:  $0.52 \pm 0.27$ ), but also reflects absolute head position (black arrows; SD-abs. trans:  $0.65 \pm 0.31$ ; SD-abs. rot:  $0.54 \pm 0.34$ ). The black trace of the global signal shows instances of motion-related signal changes, sometimes manifested as decreases in signal intensity. Throughout the paper, when possible, the same QC measure color code will be used: red for FD, blue for DV, and green for SD.

**Cohort properties; rationale for cohort composition**—Figure S1 and Table S1 show the ages, total quantities of data, summary QC values, and statistical comparisons of the cohorts. Cohorts 1, 2, and 4 (children, high-motion adults, and low-motion adults) are of principal interest. Data from cohort 3, the medium-motion adults, are also shown but will be less emphasized. The 3 principal cohorts do not differ in the amount of data available. The children and high-motion adults do not differ statistically on any summary QC measure, though there are trends for the children to have higher RMS d/dt motion ( $p=0.08$ ) and for the high-motion adults to have higher mean SD values ( $p=0.15$ ). These two high-motion cohorts differ from the low- and medium-motion adult cohorts on every QC measure.

These datasets are formed for particular reasons. First, the children and high-motion adult cohorts are useful datasets for observing effects of motion because motion is so prevalent. Second, if motion related effects are dose-dependent, they should be similar in children and high-motion adults, modest in the medium-motion adult cohort and minor in the low-motion adult cohort. Third, if we assume that above-chance differences between the adult cohorts are due solely to motion, then an effective processing strategy for removing motion-related effects should reduce any differences between these cohorts to chance levels. On this assumption, above-chance differences between the adult cohorts after a motion correction strategy reflect inadequate motion correction and the need for further work. An alternative hypothesis is that there exist endophenotypes detectable with resting state fMRI in people predisposed to move, and that a perfect motion correction strategy would enable detection of true and meaningful residual group differences related to this endophenotype. We cannot exclude this latter hypothesis. But we consider the former scenario much more likely and in Parts II and III we will work toward reducing group differences between the low- and high-motion adults to chance levels. To the extent that this effort is successful, residual differences between children and adults can be interpreted as developmental and not movement-related.

**The nature of motion-related variance**—The next 3 Figures summarize the effects that motion can produce using single-subject data from 12 subjects. The subjects that are presented were selected for their relatively clear illustrations of particular motions or signal changes, but they are representative of the entire dataset (see Supplemental Cohort Illustration). All illustrated data have been demeaned and detrended but have not undergone any further processing. As such, they illustrate the types of spurious variance that investigators seek to remove during functional connectivity processing. The top panels depict QC measures discussed previously. The third panels show voxel timeseries from voxels within the gray matter mask (GM, gray bar). Timeseries for all voxels within the mask are presented (timeseries are displayed in the order of the image but Matlab automatically downsamples the displayed data depending on plot sizing; several hundred voxelwise timeseries are visible). The fourth panel shows voxel timeseries from the eroded white matter (WM, top, white bar) and eroded ventricular masks (bottom, small yellow bar).

In relatively still subjects (Figure 3), the FD traces are flat and there is no indication from any QC measure that the data are disrupted anywhere in the scan. Nevertheless, fluctuations in signal intensity are broadly shared across voxels. These global fluctuations presumably reflect some combination of neural activity and non-motion-related artifact, such as respiratory or cardiac effects or other influences. To the extent that these global fluctuations are represented in white matter and ventricular signals, they presumably do not represent neural activity.

In subjects who move intermittently (Figure 4) and return to their original position, the QC traces are mostly flat except for brief excursions from the baseline. The impact of motion is variable. Some movements (white arrows) produce, across most voxels, an increase then a decrease in signal, lasting many TRs (see the 10 TR scaling bar in the white matter). Other movements (black arrows), appear not to produce any marked effect. Still other movements (yellow arrow), produce increases in signal. Other movements produce predominantly decreases (purple arrow). Not only the sign, but also the duration of intensity changes can vary. In the subject in the bottom left, a large movement (cyan arrow) produces a brief but marked disruption in signal, whereas in the subject in the bottom right, several modest movements produce prolonged disruptions.

Some subjects move and remain displaced from the origin, as shown by sustained elevations in the absolute displacement traces in Figure 5 (dotted dark lines). These movements

produce long elevations or depressions of signal intensity that appear as banding patterns in the voxel intensity traces. The direction of the changes is sometimes seemingly uniform (black arrow), or sometimes different at different voxels (white arrows). These sustained shifts in intensity caused by shifted head position are qualitatively different than the intensity disruptions resulting from head motion.

It is visually evident that signals in the gray matter, white matter, and ventricular compartments often resemble each other during motion. Figure 6 shows, for 160 subjects, the signal correlations between and within different tissue compartments. GM-WM and GM-CSF correlations are typically positive, centered about  $r = 0.3$  and  $r = 0.4$ , respectively. Each of these compartments contributes to the global signal, but the global signal is most strongly correlated with the gray matter signal. The similarity of signal between brain compartments, and of signals within the gray matter to other signals within the gray matter, is higher in subjects with higher FD values (except GS-GM correlations, which are at ceiling). Signal similarity is not as tightly related to RMS motion, which reflects head position in addition to head motion.

### **On the efficacy of motion-related and tissue-based nuisance regressors—**

Motion frequently produces fairly uniform effects across most voxels in the brain. This observation suggests that regressors derived from gray matter or all brain voxels should effectively reduce this variance. Figure 7 demonstrates this principle. Above the black line, the pre-regression data of 3 subjects from Figures 3–5 are reproduced. Below the black line, the gray matter voxel timeseries after several different regressions are displayed. Motion regression alone, even if expanded to 24 or 36 waveforms, does not eliminate the bulk of motion-related artifacts. Some artifacts are visibly improved by additional regressors (black arrows) but others remain largely uncorrected (white arrows). Relatively clean signals are only obtained when the global signal and its derivative are included in nuisance regressions. However, QC traces computed after all nuisance regressions indicate residual artifact during periods of motion. This issue will be addressed in Parts II and III of the paper.

An implication of these results is that the mean WM and CSF regressors are less useful than one would hope since they were included in all regressions. Figure S3 reproduces Figure 7 without any tissue-based regressors. When these Figures are compared, one is struck by the modest changes that WM and CSF regressors produce. These results are consistent with previous analyses wherein WM and CSF signals explained little variance in the gray matter (Jo et al., 2010; Weissenbacher et al., 2009). The modest efficacy of WM and CSF regressors is consistent with the moderate correlation of the WM and CSF regressors to the global signal (Figure 6), which is an effective regressor.

### **Part I discussion: on the nature of motion-related signal changes—**

Our observations so far may be summarized as follows. (1) In still subjects, variance that is not apparently motion-related may be widely shared across voxels, with variable extension to the white matter and CSF compartments. (2) Signal changes coincident with motion vary in magnitude, and may be widely present throughout gray matter, and variably, white matter and CSF voxels. (3) Complex, multi-phasic signal changes often follow head motion. (4) Shifts in head position cause sustained shifts in image intensity. (5) Phasic head motion often produces pronounced effects, typically signal decreases, extending over tens of seconds after motion has ceased (dark vertical bands in Figures 4 and 5).

These signal changes are seen throughout our entire dataset (Supplemental Cohort Illustration). The types and variety of signal changes just described, including the persistent post-motion changes, are also seen in data from the Human Connectome Project (not shown), which is acquired using different sequences on a different scanner and processed

without slice-time correction using different software. We therefore suspect that the signal changes we describe are not unique to our data.

Our findings are partially consistent with the observations of Satterthwaite and colleagues, who reported that motion produces signal decreases throughout the entire brain and that these effects scale with the size of a movement (Satterthwaite et al., 2013). However, these authors also reported that signal depressions were largely complete 2 TRs (6 sec) after a movement. We find, instead, that signal changes may persist for tens of seconds after a movement and are not necessarily signal decreases. The discrepancy between our reports may arise because we are visualizing highly variable signal changes across motions, time, and voxels, whereas Satterthwaite and colleagues were using GLM analyses to identify stereotyped signal changes coincident with motion. We note that to influence correlations, stereotyped signal changes are not necessary, only signal changes that are shared across voxels.

The variety of signal changes underscores the complexity of motion's effects on BOLD signal. Shifts in head position (e.g., slow drifts, or rapid motion without return to the original position) are associated with sustained signal changes that are variable across voxels. Motions themselves (e.g., head nodding, speech, swallowing, etc.) are associated with a variety of transient signal changes. The signal changes at each voxel are partially but not adequately explained by position estimates from the current, preceding, or past 2 TRs (since regressions fail to completely remove motion-related variance). We speculate that the variety of changes at a single voxel arises not only from the variety of specific motions, but also from the voxel's proximity to various tissue types with different signal intensities and from spin history effects. Modeling such interactions would be challenging.

Several implications follow from our observations. First, as we noted in our initial report (Power et al., 2012), motion-induced signal changes are generally not consistent with neural activity related to motion. Signal changes are seen at almost all voxels in the gray matter and the direction and duration of these signal changes varies. Further, the signal changes also occur in white matter and ventricular spaces. Some neural activity must correlate with motion and it is possible that some motion-related activity is represented in the BOLD signal during motion, but such neural activity cannot account for the preponderance of signal changes seen during and after motion. Our perspective therefore differs from that of (Yan et al., 2013).

Second, because prominent signal changes are highly variable and prolonged, it seems unlikely that they can be modeled by retrospective techniques based on relatively simple treatments of realignment estimates, such as regression of current or immediately preceding motion estimates. Prolonged motion-related effects must account in part for the inability of any set of motion regressors thus far tested to completely eliminate motion-related variance (Power et al., 2012; Satterthwaite et al., 2013; Yan et al., 2013). Removal of such prolonged changes will require more expansive modeling than is typically currently performed, such as convolution of some (unknown) function with realignment estimates, or regression of realignment estimates from many preceding timepoints. Signal-based corrections that capture these prolonged effects may prove effective in removing such variance and should receive careful attention.

We have attempted several strategies to create larger and more effective sets of nuisance regressors from the white matter and ventricular voxels, such as subdividing the masks spatially into small cubes and/or using PCA to select several components that represent most of the variance within the mask. These efforts have not yielded encouraging results. The only procedure that produced benefits worth noting was to temporally subdivide these

nuisance signals (such that, for example, a 200-timepoint signal was represented as 4 50-timepoint segments). Several versions of this temporal segmentation were examined, all of which produced modest increases in the ability of timeseries from white matter and ventricles to remove motion-related variance. However, we ultimately abandoned this approach because we were concerned about the possible removal of signal of interest that could occur by chance when using such short timeseries in the regressions.

Part I focused on timeseries. The analyses in Parts II and III focus on correlations between timeseries. Usually the timeseries will have been fully processed, though some analyses will examine timeseries at various steps of functional connectivity processing. In future analyses, the 24-motion-parameter set of regressors will be used. This is because it is likely to be a standard for the field in the short- to medium-term future given its statistical superiority over smaller sets of motion parameters (Satterthwaite et al., 2013; Yan et al., 2013). We choose not to use the 36-motion-parameter set because although it produced modest benefits beyond the 24-motion-parameter set, it was still inadequate for removing motion-related effects, it requires an additional 12 degrees of freedom, and it is useful to for the coming results to be comparable with other recent reports that utilize the 24-motion-parameter set of regressors (Satterthwaite et al., 2013; Yan et al., 2013). Additionally, most future analyses will include data processed with and without GSR given the visible efficacy of this regressor (Figure 7) and the recent reports on its utility for removing motion artifact (Satterthwaite et al., 2013; Yan et al., 2013). Since the field continues to debate the use of GSR (Fox et al., 2009; Murphy et al., 2009; Saad et al., 2012), analyses with and without GSR will provide investigators of all viewpoints with empirical evidence on the efficacy of their preferred approach to artifact removal.

## Part II: Understanding how motion impacts RSFC correlations

It is established that motion produces systematic changes in RSFC correlations (Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012). This section seeks to answer several outstanding questions related to motion and RSFC correlation structure: (1) whether the prolonged signal disruptions after motion systematically impact RSFC correlations; (2) whether absolute head displacement systematically impacts RSFC correlations; (3) whether stricter censoring thresholds produce greater removal of artifactual RSFC structure; (4) whether high values on all QC measures identify similar artifactual properties; (5) whether ‘improvements’ in a volume over processing are complete or partially cosmetic (i.e., does a volume with high DV values before processing but normal DV values after processing now contribute RSFC structure that resembles ‘random’ data or does it still resemble high-DV data); and (6) whether a threshold or inflection point can be found in a QC measure, below which artifactual influences are negligible. The answers to these questions will guide the formation of subject-level censoring-based strategies of artifact correction in Part III.

In our initial report on motion (Power et al., 2012), to characterize the impact of motion on correlations, we compared correlations in full timeseries to correlations in timeseries from which high-motion volumes had been deleted. This censoring procedure was called ‘scrubbing’, a term that refers to the practice of discarding incorrect or untrustworthy pieces of information. In our initial report censoring was only done during correlation calculations, but it can be incorporated into data processing steps such as regressions and temporal filtering (Carp, 2013; Power et al., 2012; Power et al., 2013; Power et al., 2011; Satterthwaite et al., 2013). Certain characterizations of high-motion data (as in Figure 11B) cannot be performed if the data have been censored or replaced. Therefore, in this section we censor only after functional connectivity processing is complete. Iterative processing with censoring incorporated throughout the processing stream is described in Part III.

Several analyses in this section use scrubbing to characterize the impact of particular parts of data on RSFC correlations. These analyses examine the changes in correlation produced in all possible pairwise correlations (34,716) between the 264 ROIs reported in (Power et al., 2011) when particular portions of the data are withheld from correlation calculations.

Across-subject changes in correlation are reported ( $\Delta r$  = mean scrubbed  $r$  – mean unscrubbed  $r$ ). Within-subject analyses ( $\Delta r$  = mean(scrubbed  $r$  – unscrubbed  $r$ ) produced nearly identical results and are not shown. Within-subject analyses that were normalized to the amount of data removed per subject produced similar results and are not shown. In subsequent Figures,  $\Delta r$  will be plotted for each pairwise correlation as a function of the Euclidean distance separating the center coordinates of the ROIs that gave rise to the correlation.

**The temporal extent of motion's impact on RSFC correlations**—The prolonged signal changes seen in Figure 4 suggest that motion may influence RSFC correlations for many TRs after the head has ceased moving. To investigate this issue, 8 ‘target’ temporal masks were formed in each subject. Each target mask identified a particular type of individual volume, such as volumes prior, during, or after motion. Such masks are illustrated for 2 subjects in Figure 8A. The ‘target masks’ bar contains 8 rows, each representing a temporal mask. Rows 1–2 individually identify the 2 TRs prior to motion, row 3 identifies motion ( $FD > 0.2$  mm), and rows 4–8 individually identify the 5 TRs after motion. A set of ‘positive control’ masks was also created, one for each of the target masks. Each ‘positive control’ mask removes, within each subject, identical amounts of data by FD rank (e.g., if a ‘target mask’ removed 8 volumes, the corresponding ‘positive control’ mask removed the 8 volumes with the highest FD values in the scan). A set of ‘random control’ masks removes, for each ‘target mask’ within each subject, identical amounts of data but at random points in the scan. The ‘positive control’ shows what changes in correlation are possible when removing a given amount of data (because sometimes very little data is removed), and the ‘random control’ shows that correlation changes are specific to the type of volume identified by a target mask.

The temporal extent of the influence of motion on RSFC correlations is shown in Figure 8B. Here, for each of the 8 types of mask described above, the  $\Delta r$  for each pairwise relationship is plotted by the distance between that pair of ROIs. Without GSR, motion's influence is evident in the 4 or 5 TRs (10–12.5 seconds) following a motion. With GSR, motion's influence is essentially restricted to the period of movement. These effects were calculated in all subjects with data removed by a given mask ( $N = 150$ ); within-subject analyses yielded virtually identical results (data not shown). An informative feature of these analyses is that no artifact is seen in the TRs prior to motion, indicating that our gentle (and symmetric) frequency filtering did not result in noticeable temporal spread of artifact into adjacent TRs. Figure S4 extends these analyses to the 10 TRs following motion and to timeseries from various steps in functional connectivity processing (pre-regression, post-regression, and final timeseries). Similar effects are seen at all stages of processing. These analyses indicate that, in our data, motion impacts correlations mainly in the 4 TRs (10 s) after a movement. The analyses in Figure S4 also include a mask that identifies volumes temporally distant from motion (at least 10 TRs after  $FD > 0.2$  mm) that have absolute displacements over 0.5 mm. Censoring with this mask reveals no obvious structured influence of absolute head displacement on RSFC correlations, indicating that the sustained signal changes produced by absolute head displacement seen in Figure 5 are largely corrected by existing regression strategies. However, only 20 subjects exhibited isolated absolute head displacement, and we therefore consider this conclusion to be provisional.

**The effect of censoring stringency in different populations**—An important question is the extent to which different populations exhibit motion-related biases in RSFC.

Relatedly, we have posited that above-chance group differences between high-motion and low-motion adults are at least partially and perhaps entirely, due to motion. We are therefore interested in the extent to which different populations display motion-related effects, and the extent to which various processing strategies reduce group differences.

The effects of lenient ( $FD > 0.5$  mm) and stringent ( $FD > 0.2$  mm) FD-based scrubbing are shown in Figure 9. With GSR, children display effects at both thresholds, high-motion adults display effects only at the stringent threshold, and low-motion adults display effects at neither threshold. Importantly, FD-based scrubbing reduces group differences between the high- and low-motion adults. This effect is specific to high-FD volumes: random volume removal produced no such effect. This effect is seen regardless of statistical threshold used to define group differences, and the selective reduction in group differences is seen in all comparisons of all adult cohorts (Figure S5).

Without GSR, similar effects are seen as with GSR, except that effects can now be appreciated in the high-motion adults at lenient thresholds, and in the low-motion adults at stringent thresholds (the red fringes below the black points in both plots). Here, too, FD-targeted scrubbing selectively reduces group differences. Note also the scale of group differences: for a given statistical threshold, using identical temporal masks, the number of significant differences without GSR is nearly 2 orders of magnitude greater. These findings hold regardless of statistical threshold or the adult groups compared (Figure S5). The elimination of group differences by various processing strategies will be revisited in Part III.

**On using DV and SD to censor data**—FD, DV, and SD are all possible QC measures. Thus far, we have characterized data using FD. We now turn to the other QC measures, which differ from FD in important ways. FD is based on realignment estimates and therefore is unaffected by subsequent processing steps. DV and SD, in contrast, derive from BOLD intensity, and may evolve through processing (Figure 10). Before processing, DV strongly resembles FD, but this similarity diminishes with processing (DV-FD pre-regression:  $r = 0.69 \pm 0.34$ ; post-regression:  $r = 0.23 \pm 0.28$ ; post-frequency filtering:  $r = 0.17 \pm 0.27$ ; post-spatial blurring:  $r = 0.18 \pm 0.25$ ). Some peaks in the DV trace are abolished, others are reduced, others remain. The same is true for SD values (SD-FD pre-regression:  $r = 0.52 \pm 0.27$ ; post-regression:  $r = 0.07 \pm 0.29$ ; post-frequency filtering:  $r = 0.07 \pm 0.27$ ; post-spatial blurring:  $r = 0.04 \pm 0.24$ ). Before regression SD traces uniquely display plateaus and scalloping corresponding to absolute head position, but after regression the scalloping and plateaus are largely eliminated, possibly reflecting the efficacy of existing regressions in correcting artifacts attributable to absolute displacement. It may be advantageous to use QC measures that track evolving data quality rather than an unchangeable trace of head motion. For example, a temporal mask made using DV after functional connectivity processing might appropriately retain more (denoised) data than a mask formed using FD (which cannot reflect denoising).

These considerations prompt two questions. First, do volumes with outlying DV and SD values produce changes in correlation similar to what has been seen in FD-based analyses? The right side of Figure 10 answers the first question. Outlying values of either DV or SD, at any stage of processing, when used to censor fully processed timeseries, produce the familiar distance-dependent effects seen with FD-based analyses. This is true regardless of the step in processing at which correlations are calculated (e.g., in pre-regression or pre-frequency-filtering timeseries; data not shown). Thus, the elevations in QC traces identify data with similar characteristics at all times.

Second, the reduction or elimination of some peaks in QC traces prompts the question of whether volumes that ‘improve’ in QC values over processing are truly corrected or whether

such changes are cosmetic. In other words, if a volume begins with an outlying DV value and then acquires a more typical DV value later in processing, does that volume's impact on RSFC correlations now resemble that of a typical volume (from a period of stillness), or does it continue to resemble the impact of an outlying volume (from an 'uncorrected' period of motion)?

Full descriptions of the analyses that answer this question are provided in the Supplemental Materials and Figure S6, but the unfortunate answer is that DV improvements are partially cosmetic, and data that begin with an elevated value that decreases throughout processing many still harbor motion artifact (Figure 11A). Initial QC values before processing are therefore the most reliable indicator of data quality. Because SD reflects absolute head position, which does not clearly impact RSFC correlations (Figure S4), and because its motion-tracking characteristics are otherwise largely represented by FD and DV (Figure 10), we do not pursue further analyses using SD. Thus, the remainder of this paper focuses on using FD and initial DV values to make processing decisions.

**On relating QC values to significant changes in RSFC correlations**—Until more efficacious artifact removal techniques are demonstrated, nuisance regression alone will be inadequate for subject-level correction of motion's effects. A further step that can be taken at the subject level is to entirely discard motion-corrupted volumes. This approach sacrifices data but it is also effective in eliminating motion-related variance (Power et al., 2012; Power et al., 2013; Satterthwaite et al., 2013). An important question is: what data should be censored? We reframe the question as: at what QC values do significant within-subject changes in correlation become evident?

Full descriptions of the approach and analyses developed to answer this question are presented in Supplemental Materials and Figure S7. The top panel of Figure 11B shows the gist of the analyses, wherein a subject's data are ordered by decreasing quality and the changes in correlation observed going from the best to the worst data are plotted. Whether such changes are significant is empirically determined by repeating this procedure with random orderings of the data (irrespective of QC values). The green points are insignificant changes in the QC-ordered data, the red points are significant changes, and the black points are from the random permutations that define empirical significance. When this procedure is repeated for all 160 subjects the resulting empirical ranks can be plotted as a function of QC value, as in the bottom panel of Figure 11B. This plot indicates that significant within-subject changes in correlation are detectable down to FD = 0.15–0.2 mm and are very pronounced at FD = 0.5 mm.

The analyses shown in Figures 11B and S7 indicate that high, outlying values on QC measures (e.g., FD = 0.5 mm, or DV = 20) unquestionably are associated with within-subject elevations in short-distance correlations. Most data have QC values not associated with such artifactual elevations in correlation (see the cumulative distribution curves, the black sigmoid trace in Figure 11B). On the basis of these results, one could argue that very strict thresholds, excluding QC values beyond even a hint of a skew ( $FD > 0.15$  or initial  $DV > 13$ ) are ideal. Most datasets will not tolerate censoring based on such strict thresholds, nor is it obvious that this particular analysis should be the only criterion for setting thresholds. We interpret these results to support the general principle that the stricter one sets thresholds for censoring, the more one can guard against or eliminate motion-related artifact.

**Part II discussion: on motion's influence on RSFC correlations**—The key observations of this section are: (1) effects of motion manifest in RSFC correlations for ~10 seconds after motion; (2) post-motion effects are essentially eliminated when GSR is performed; (3) absolute displacement does not appear to produce systematic changes in

RSFC correlations; (4) stricter censoring thresholds remove larger amounts of distance-dependent artifact; (5) distance-dependent artifact is present with or without GSR; (6) motion scrubbing selectively reduces motion-related group differences in adult cohorts; (7) improved QC values do not guarantee that the data are fully corrected; (8) QC values can be quantitatively linked to the significance of changes in RSFC correlations, but we are unable to point to a single value of FD or initial DV as a definitive threshold beyond which data are compromised – the effect is gradual.

### Part III: Processing data in ways that minimize motion-related influences

In this section, we use the findings of Part II to modify our processing stream to more powerfully suppress motion artifact at the subject level. We assess the ability of this processing stream to eliminate detectable influences of motion and group differences that are attributable to motion.

**Incorporating censoring into an iterative processing strategy**—We now adopt a strategy where the data are processed as in Parts I and II, then QC measures are derived and used to form temporal masks, and then the data are re-processed using the temporal masks to censor data (Figure 1, dotted gray lines). The principal features of the iterative processing strategy are: (1) censoring is incorporated into demeaning and detrending each run, (2) censoring is incorporated into the multiple regressions performed across runs, (3) temporal masks are used to define data that are replaced by interpolation, prior to frequency filtering, and (4) the interpolated data are recensored following frequency filtering.

In this iterative scheme, 4 steps are used to form the temporal masks used to reprocess the data (see Figure S8 for an illustration in a single subject). Step 1: volumes with  $FD > 0.2$  mm or initial DV  $> 20$  were censored. Step 2: uncensored segments of data lasting fewer than 5 contiguous volumes as a result of Step 1 were censored. Step 3: runs with fewer than 50 remaining volumes were entirely eliminated. Step 4: subjects with less than 125 uncensored volumes across runs were eliminated. This procedure eliminated entire runs in 5 children and 2 high-motion adults; 3 children were removed for insufficient remaining volumes across runs. Overall,  $83\% \pm 17\%$  (range 28% – 100%) of the data were retained (Figure S8). To investigate the impact of interpolation, the data were reprocessed both with and without this step. Fully processed timeseries were often similar with and without interpolation (Figure S9). However, in several instances where motion occurred, reprocessing without interpolation yielded anomalous timeseries features (Figure S9). These anomalies occurred because outlying values in the censored timepoints were neither ameliorated by regression (because they did not contribute to regressor fits) nor were they replaced by interpolation, and they could therefore spread large-amplitude artifactual signals into adjacent TRs during frequency filtering. We therefore only focus on data that are reprocessed including interpolation; hereafter ‘reprocessed’ refers to data reprocessed with interpolation.

Other criteria could have been used to form temporal masks and some differences with our previous report should be mentioned. In particular, here we did not augment temporal masks (expanding them forward and backward in time) because we are principally interested in results using GSR, which exhibited no impacts of motion on correlations in TRs before or after motion (Figure 8). Analyses done without GSR may benefit from censoring several TRs after head movement (such augmented masks are examined where appropriate below). The requirements for some number of contiguous volumes (Step 2) and a minimum number of within-run volumes (Step 3) have no particular empirical support. The need for 125 volumes across runs (Step 4) arises from the conventional minimum of ~5 min of data for computing correlations (Van Dijk et al., 2010).

**Evaluating reprocessed data for evidence of motion artifact**—One way of assessing RSFC data for evidence of motion artifact is to correlate a vector of summary QC values (e.g., mean FD or RMS motion) with vectors of outcome measures across subjects. This linear dependence is what is removed by group-level corrections implemented by linearly regressing summary QC measures out of outcome measures across subjects (Satterthwaite et al., 2013; Satterthwaite et al., 2012; Van Dijk et al., 2012; Yan et al., 2013). We refer to these correlations as QC-RSFC correlations.

QC-RSFC correlations across 120 adults are shown for several processing strategies in Figure 12, using mean FD as a summary QC measure. In unscrubbed data prepared with GSR, short-distance correlations are positively correlated with mean FD. With scrubbing, these relationships are reduced, and with reprocessing, they are almost absent. Without GSR, correlations at all distances are positively related to mean FD and these relationships can be reduced but not eliminated by scrubbing and reprocessing. Reprocessing data with temporal masks that also censored the 4 volumes subsequent to  $FD > 0.2$  mm did not alter results (data not shown).

**Between-cohort differences can be reduced to chance levels by scrubbing**—Another way to assess the efficacy of within-subject motion correction is to examine group differences under various processing strategies. We have created 3 adult cohorts that differ by mean FD. Comparisons of all groups yield significant group differences (Figure S5). We now examine the extent to which different processing strategies reduce these group differences, which are in part, if not entirely, due to motion.

The observed numbers of adult group differences under different processing strategies are presented in Figures 13 and S10. Null expectations were established by 10,000 permutations of adult subject identity. In general, scrubbing eliminated many group differences, and reprocessing further reduced group differences. Without GSR, these steps reduced but did not eliminate group differences. With GSR, reprocessing reduced group differences to chance levels at most statistical thresholds (in 8 of 9 analyses, Figure S10).

These results indicate that within-subject corrections may be capable of adequately correcting for motion without a need for group level correction. It is possible that other versions of censoring would outperform the current approach. Indeed, since this is the first attempt we have made at a comprehensive approach to censoring, it would be surprising if the approach could not be improved. We regard these analyses more as proof-of-principle, and an adequate strategy, than as a demonstration of a fully optimized processing strategy.

**Some limitations of group-level correction**—The present work has been motivated in part by a desire to avoid some of the drawbacks of group-level regression. These drawbacks were noted in the Introduction and are now made more explicit.

The chief drawback of group-level regression is that it will tend to remove effects that covary with a factor of interest if the factor of interest covaries with motion (e.g., if a correlation indexes ADHD severity but ADHD severity correlates with motion, group-level motion correction can reduce or eliminate the diagnostic utility of that correlation).

In the current study, 3 adult cohorts have been formed by binning subjects by mean FD. If we were to perform across-group regression of mean FD in these cohorts, regression would tend to remove all group differences, whether “real” (possibly motion-related endophenotypes) or motion-related (since all correlations are influenced by motion, Figure 12). Figure 14A illustrates this effect using modeled data. In all cases group differences are eliminated by across-group regression, though this was appropriate only in the top row.

Therefore, in the current data, across-group regression is of little help in selectively correcting group differences due to motion.

Within-group regression may, in certain circumstances, recover “real” group differences. For instance, in the model data of Figure 14A, “real” differences could be recovered at the intercept by within-group regression. This is possible in the model data because we designed the data and know that its properties allow successful group correction. We examine below whether real data possess similar properties that would enable successful within-group regression.

Within-group regression in the model data works because the within-group rise of RSFC as a function of FD is linear and identical in both groups (the QC-RSFC betas are identical). However, if the data had been constructed such that the QC-RSFC betas differed between groups, another analyst would be unable to determine whether the observed difference in betas were due to 1) non-linearity in QC-RSFC relationships across the QC range, 2) an interaction of some factor of interest with the QC measure (e.g., if ADHD severity modulated correlations but also covaried with FD, QC-RSFC slopes could differ between controls and ADHD patients), or the fact that 3) since FD summarizes several types of motion, cohorts that have similar FD values may differ in the underlying types of motion and therefore exhibit different QC-RSFC relationships. Importantly, even if the cohorts were well matched for FD (unlike the model data or the 3 FD-binned adult cohorts of this report), these concerns would still be relevant: the 2<sup>nd</sup> and 3<sup>rd</sup> concerns could still cause differential betas in FD-matched groups, and the 1<sup>st</sup> concern could undermine the accuracy of the calculated residuals over any FD range if the QC-RSFC relationship is non-linear.

To assess whether within-group regression might be a promising option for correcting motion-related influences in our real data, we examined whether FD-RSFC relationships were linear and similar in each of the adult cohorts. We calculated QC-RSFC betas for all 34,719 correlations across all 120 adults (black line), in each of the 3 40-subject FD-binned adult cohorts used throughout this paper (red lines), and in 3 randomly-formed cohorts of 40 adults that have indistinguishable means and ranges of mean FD (green lines). Figure 14B plots these linear fits for 3 randomly selected correlations and Figure 14C shows the betas found at 100 randomly selected correlations. The fits within subsets of the FD range are considerably more variable than the fits found across the entire FD range. Such variability could arise from noisy estimates due to the truncated FD ranges represented in each adult cohort relative to random subsets of the entire cohort, or it could reflect non-linearity in QC-RSFC relationships.

We have viewed hundreds of these plots and in a non-trivial number of cases it seems that QC-RSFC relationships are not just noisy but are actually non-linear. For example, in the middle plot in Figure 14B, residuals from the ‘all adult’ fit would be skewed negatively at low FD values and then positively in middle FD values. Consistent with this impression, across all 34,719 correlations, the relation between FD and observed RSFC values tends to be strongest across low motion subjects and weaker across medium and high motion subjects. This is reflected in the beta histograms of Figure 14D, where the betas are right-shifted for the low-motion cohort relative to the other cohorts. This result suggests that in many correlations we cannot assume a linear QC-RSFC relation across the range of our QC measure. This finding calls into question the accuracy of the residuals obtained after regression across all subjects (the black or green lines), and it also indicates that in real data, within-group regression will obtain differential QC-RSFC betas in the adult cohorts (red lines) and is therefore unlikely to produce the desired types of corrections illustrated in Figure 14A.

## Discussion

Important points from Parts I and II were discussed at the close of each section. This Discussion therefore focuses on the results of Part III and the implications of the report as a whole.

### On the use of group-level correction

If GSR (or some equivalently effective subject-level correction) is not performed, some other technique is needed to control for motion-related group differences. Group-level regression has been proposed as a corrective measure (Satterthwaite et al., 2012; Van Dijk et al., 2012; Yan et al., 2013). It is important to note that QC-RSFC betas differ at different portions of the mean FD range. In the current data, this nonlinearity would cause mean FD, for a given pairwise correlation, to act differently as a regressor in the different adult cohorts, since the cohorts occupied different portions of the mean FD range. If within-group regression is performed, it is therefore advisable to match cohorts not only on mean QC values, but also on ranges of summary QC values in order to (presumably) obtain homogeneity of regression slopes. Even so, depending on the range of motion, although a linear fit may be obtained, the underlying data may not actually be linearly dependent on motion, and it is also possible that differential betas will still be found even in well-matched groups (concerns 2 and 3 above).

### On the implementation of censoring

Several groups have found a benefit of censoring high-motion data (Power et al., 2012; Satterthwaite et al., 2013; Yan et al., 2013). This benefit accrues because all nuisance regressions thus far examined have failed to adequately model or remove motion-related variance. Until improved motion correction strategies are devised, censoring will be a useful tool for reducing or eliminating motion-related variance in resting state timeseries.

It is important to recognize that some minimum amount of data must remain after censoring. The field accepts ~5 minutes of data as an adequate starting point for RSFC analyses (Van Dijk et al., 2010). Accordingly, we required that 125 volumes (5.2 min) remain after censoring for inclusion in our analyses. Because we started with several hundred volumes of data in most subjects, this requirement was easily met, and we typically had several hundred volumes of data remaining in our subjects even after strict censoring strategies.

As noted in the Introduction, the datasets included in this study were chosen for their ability to withstand such strict censoring strategies. In this paper, few subjects were eliminated in any analysis because of this selection bias. In other completed and ongoing studies we do eliminate subjects: in studies of normal adults, we eliminate perhaps 10% of our subjects; in pediatric and clinical studies, the proportion can exceed 30%. This loss of subjects is regrettable but difficult to avoid if within-subject correction is to be used. We have, over the last two years, begun to collect more resting state data from each subject, in longer runs, so that more subjects qualify for correction by censoring.

We do not advocate relaxing the quantity requirement much below 125 volumes (~5 min) in order to salvage subjects. Removing timepoints increases the sampling error of estimated correlations. In the scrubbing analyses, the addition of white noise by censoring is illustrated by the random  $\Delta r$  analyses that yielded a Gaussian distribution of black points centered on zero.

A related issue is how to deal with degrees of freedom when censoring, since the number of timepoints contributing to outcome measures typically differs across subjects after censoring. One way to address this problem is to reduce all temporal masks to the same

length (e.g., the minimum length found across subjects in a study). We have found that such ‘trimming’ of temporal masks can have an impact when techniques such as multivariate pattern analysis are used to classify individuals that differ in the amount of data available (unpublished observations). In other group-level applications, such as mean seed maps or mean pairwise correlations, especially when calculated over many subjects and in datasets with many hundreds of post-scrubbing volumes in most subjects, such ‘trimming’ appears to make little difference. For example, reanalysis using trimming (to 126 volumes) made no difference in the group difference results of Figure 13 (data not shown). The additional removal of volumes must be balanced against the degrees of freedom available in a dataset and the degrees of freedom required by a processing stream (which is influenced by autocorrelation, nuisance regression, the presence and type of frequency filtering, etc).

Another issue worth mentioning is the interaction of temporal dynamics with censoring. The detection and characterization of dynamic states in resting state MRI is an increasingly examined topic (Chang and Glover, 2010; Smith et al., 2012), and some objections to censoring have been raised on the grounds that censoring might decrease contributions to correlations from certain transient states that are associated with motion. In a sense, this concern is valid: motion could be viewed as a ‘state’ in which short-distance correlations are increased relative to long-distance correlations. Our opinion, guided by data such as that shown in Figures 3–5 and the Supplemental Cohort Illustration, is that a motion-associated state, if it exists, is not a state that can be meaningfully examined with fMRI.

### On anticorrelations observed under different processing strategies

Different processing strategies yield different distributions of correlations, changing the inferences investigators might draw about brain organization. Figure 15A presents the  $264 \times 264$  mean correlation matrices in each adult cohort (rows) under different processing strategies without GSR (columns). Scanning down the first column, it is evident that the low-motion cohort displays more negative correlations than the medium-motion cohort, and similarly that the medium-motion cohort displays more negative correlations than the high-motion cohort.

This pattern is easily appreciated in histograms of cohort correlation values (Figures 15B). Figure 15C shows that these negative correlations prominently involve the posterior midline and angular gyrus, locations associated with the default mode network. Scanning across the first row of Figure 15A, it is evident that more successful motion correction also produces leftward shifts in correlation distributions in the high-motion cohort, a shift shown in Figure 15D. Thus, the presence or absence of negative correlations in a dataset prepared without GSR depends, among other things, on the extent of motion in the data, and the extent to which effects of motion are corrected.

Correlations observed with and without GSR are distinct quantities that require different interpretations (since correlations after GSR are effectively partial correlations, as discussed below). Correlations with and without GSR are therefore not directly comparable in a statistical sense. Nevertheless it is worth empirically noting that the strongest correlations observed without GSR tend to also be the strongest correlations observed with GSR, and that the most negative correlations observed without GSR tend to be the most negative correlations observed with GSR (Figure 15E,F). Thus, an analysis based on matrices to which a high threshold (such as examining the top 5% or 10% of correlations) has been applied will likely yield similar results with or without GSR, whereas an analysis based on the full correlation matrices may yield less convergent results under different processing strategies.

## On the benefits and drawbacks of global signal regression

From a perspective of eliminating artifactual variance, especially motion-related variance, GSR is unquestionably powerful. However, GSR is a contentious step in processing (Anderson et al., 2011; Murphy et al., 2009; Saad et al., 2012). Investigators must therefore weigh the benefits and drawbacks of GSR when deciding how to analyze their data.

Historically, the first objection to GSR was that it induces artifactual anticorrelations (“creates anticorrelations where none exist”) (Murphy et al., 2009). Recently, acquisition of RSFC and ECoG data in several patients has established that anticorrelations in the resting human brain are of neural origin, and that GSR improves the correspondence between the anticorrelations seen with ECoG and those seen with RSFC (GSR also improves the correspondence of positive correlations) (Keller et al., 2013). Further, the present analyses show that RSFC anticorrelations, obtained without GSR, are more evident in subjects with less motion artifact. Thus, the complete or relative absence of anticorrelations may relate to the extent of artifact in a dataset. That the brain exhibits true anticorrelations seems to be well established at this point.

Regardless of the presence of ‘true’ anticorrelations, it is also true that global signal regression imposes a substantial negative bias on computed correlations because correlations must be approximately zero-centered (Fox et al., 2009). When such biases arise by removing shared artifact or truly shared neuronal variance, GSR increases the specificity of correlations in the residual data. As this paper has shown, much globally shared variance is indeed artifactual. Further, Scholvinck and colleagues have found electrophysiological evidence for widely shared neural activity at rest in non-human primates (Schölvinck et al., 2010). To the extent that global signal regression removes artifact and truly shared neural variance, it can be an appropriate and helpful step in processing.

However, it has been pointed out that bias introduced by GSR can cloud interpretation of group differences in RSFC because the global signal is composed of an average of signals throughout the brain. This argument has been made in 2 forms: first, that distortion occurs within a subject or cohort simply by regressing the global signal (Murphy et al., 2009), and second, that differential distortion occurs in different cohorts if the underlying networks are differentially composed (Saad et al., 2012).

These arguments are mathematically sound, but the effect size depends on the dimensionality of the data. That global signal regression ‘distorts’ relationships is very evident in small systems with few independent signals (e.g., a 3 signal system such as that used in (Saad et al., 2012)). However, distortion effects diminish rapidly in systems with increasing numbers of independent signals. Additionally, as (shared) artifactual signal increases relative to real signals, such distorting effects also diminish (Chen et al., 2012).

It is difficult know the number, relative strength, and spatial distribution of the signals present in the brain tissues that comprise the global signal. Studies examining resting state fMRI signal dimensionality suggest that at least a few dozen distinct signals may be present (Cordes and Nandy, 2006), and several groups have reported that resting state signal can be broken into 1–2 dozen major components or communities that correspond to task-associated groups of brain regions (Power et al., 2011; Smith et al., 2009; Yeo et al., 2011). The size of real signals in relation to artifactual signals is not easy to determine and will vary by dataset. Our analyses indicate that artifactual signals can often be quite large relative to real signals (Figures 3–6 and Supplemental Cohort Illustration). Simulations that capture the dimensionality of actual data and the magnitude of artifactual signal in relation to real signal will best inform the debate over the use of GSR.

In our studies of motion effects in resting state data, two empirical observations are noteworthy in relation to the concern raised by Saad and colleagues (that differential composition of global signal may induce spurious group differences). The first is that this concern can be addressed empirically: one can compare global signal composition (between groups, subjects, scans, etc.) by examining the spatial beta maps arising from global signal regression (Power et al., 2013) (or, similarly, maps of correlations of the global signal to voxel timeseries, as in (Gotts et al., 2012)). When interpreting the results of this approach, it is important to remember that the global signal is a combination of artifactual and neural sources, and that the balance of these sources will impact observed differences in global signal composition. For example, when the global signal composition in children and adults are compared using t-tests of global signal regression beta maps without censoring, prominent and significant group differences are found (Figure 1C of (Power et al., 2013)). However, when even lenient censoring ( $FD > 0.5$  mm) is used to remove motion-contaminated data from global signal regression beta calculation, such group differences are markedly reduced (Figure 1D of (Power et al., 2013)). These results indicate that pre-censoring differences in global signal composition do not necessarily reflect differences in the underlying distribution of neural signals, but instead may at least partially be due to different amounts of motion artifact in the data contributing to the global signal.

The second empirical observation is that, on the argument that differential global signal composition and regression induces spurious post-GSR group differences, one might expect that adult group differences observed with GSR would exceed group differences observed without GSR. Instead, the opposite effect is observed: with or without censoring, processing with GSR reduces the number of observed group differences seen between adult cohorts by almost 2 orders of magnitude compared to processing without GSR (Figures 9 and 13). These results suggest that removal of spurious differences related to motion artifact far outweighs spurious post-GSR distortions related to differential network composition. This result is consistent with the recent report by Tyszka and colleagues, wherein control versus autistic group differences were much smaller than motion-related effects within the same subjects (Tyszka et al., 2013).

Investigators considering global signal regression therefore face a choice. This paper has shown empirically that GSR is highly effective in removing artifactual variance in RSFC data (Figure 7), that adequate subject-level motion artifact correction can be implemented with GSR in combination with censoring (Figures 12), and that motion-related group differences are 1–2 orders of magnitude less with GSR than without GSR (Figure 13). Recently published evidence shows that GSR increases the resemblance of RSFC correlations to electrophysiological measures (Keller et al., 2013). Balanced against these benefits are the possible distortions within networks and across groups highlighted by Murphy et al and Saad et al. Our discussion confirms these concerns but emphasizes that the relevance of these concerns to actual data is incompletely established. On the whole, and certainly until other adequate methods of artifact removal are developed, we find the empirical benefits to outweigh the theoretical costs of GSR.

### On detecting the influence of motion

This paper has used several approaches to quantify the influence of motion on BOLD signal and RSFC correlations. We wish to outline conceptually what, in our opinion, does and does not establish that a dataset is free of artifactual influences related to motion.

Scrubbing analyses are useful because they can be applied at the individual or the group level, and because they effectively characterize motion-contaminated data. However, a censoring analysis that finds no distance-dependent artifact does not necessarily establish an absence of distance-dependent artifact. Suppose, for instance, that motion with  $FD > 0.2$  mm

corrupts data with a distance-dependent artifact. If data with  $FD > 0.5$  mm are replaced (by interpolation, by a mean value, etc.) and then a scrubbing analysis is performed with a threshold of  $FD > 0.5$  mm, no distance-dependence will be found. This result does not actually demonstrate that the data are free of artifact. If the scrubbing threshold were lowered to  $FD > 0.4$  or  $FD > 0.2$  mm, distance-dependent artifact would be found. If the threshold were excessively lowered into the range of normal data (e.g.,  $FD > 0.1$  mm), data without motion artifact would also be censored, diluting the characterization of the artifact-laden data. The central point is that outlying data (to be censored) must be sensibly defined for the scrubbing analyses to be useful.

Because censoring analyses are threshold-dependent and can be performed in ways that miss motion artifact, examination of QC-RSFC correlations is an important part of evaluating a dataset for influences of motion. These analyses can only be employed across multiple scans or subjects, but their virtue is that no binarizing threshold is needed. For instance, in the example above, the scrubbing analysis using a threshold of  $FD > 0.5$  mm might show no distance-dependence, but QC-RSFC correlations examined in the same dataset would reveal a distance-dependent effect. Even if a scrubbing threshold were sensibly defined (e.g.,  $FD > 0.2$  mm in the present data), QC-RSFC correlations might reveal influences of smaller movements that are not practically addressable or detectable by censoring analyses. It is for this reason that we focus on both scrubbing and QC-RSFC correlation analyses in the present manuscript.

One other important consideration relates to the selection of QC measures for defining scrubbing or QC-RSFC correlation analyses. DV, since it is based on BOLD signal intensity, will differ across datasets and processing strategies, and can be influenced by blurring kernel size, frequency filter characteristics, sequence characteristics, etc. DV values may therefore not be comparable across datasets. FD measures (and other measures of absolute or relative displacement), on the face of it, seem like they should be comparable across datasets. However, our experience with Human Connectome Project (HCP) data (see (Smith et al., 2013) for a description of the data) has altered this assumption. In contrast to the present data (acquired with TRs of 2.2–2.5 seconds), the HCP data are acquired with TRs of a few hundred milliseconds. This shorter TR has the effect of dividing large movements into several smaller movements, while simultaneously sampling effects like head bobbing due to cardiac pulsations or respiration much more frequently. The net effect is that, in the HCP data, DV traces exhibit signal-to-noise ratios that are useful for identifying outlying datapoints, while FD traces are noisier and less useful in identifying outlying timepoints (unpublished observations). We raise these points to emphasize that choosing an appropriate QC measure and utilizing it effectively requires attention to an individual dataset; simply taking a threshold or a QC measure from the literature without assessing its characteristics in the data at hand may result in uninformative analyses.

## Overall Summary

This paper aimed to expand knowledge about motion artifact in several ways. It illustrated the diversity of signal changes that motion produces and pointed out some reasons why current subject-level regressions inadequately capture motion-related variance. It placed temporal limits on the ability of motion to impact RSFC correlations. It showed that data improvements are partially cosmetic in terms of QC values and proposed methods to link QC measures to significant changes in RSFC correlations. It proposed a within-subject correction strategy that greatly reduces motion-related variance, and showed some conditions under which group-level correction is or is not necessary, or effective. Our current practices and recommendations are outlined in Table 2.

## Limitations and Future Directions

The options for processing resting state data are numerous. We have partially explored some parameter spaces corresponding to a subset of possibilities for data processing. Other possibilities remain to be explored. A further limitation of this work is that only data acquired using a single sequence were examined. Other sequences may display different characteristics with regard to the magnitude, variety, and duration of motion-related effects. An additional limitation is that only a single set of criteria were examined for forming temporal masks for reprocessing. It is possible that simple changes to our criteria might yield easily obtainable improvements in control of motion artifact. Another limitation is that only ‘functional connectivity processing’ was examined. It is possible that modifications earlier in the processing stream would be more desirable, such as performing interpolation procedures prior to slice-timing correction.

Post-hoc artifact reduction in resting state BOLD data is challenging. With advances such as short-TE or multi-band sequences (Bright and Murphy, 2013; Feinberg et al., 2010), future datasets will hopefully contain within themselves powerful means of artifact avoidance and removal. Newer ICA-based techniques may prove successful in identifying and removing artifactual influences. Novel treatments of realignment estimates may also prove more useful than current realignment-based corrections. We believe, however, that effective means for retrospective correction of motion-related artifact are already available.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was funded by NIH F30 MH940322 (J.D.P.), NIH R21NS061144 (S.E.P.), a McDonnell Foundation Collaborative Action Award (S.E.P.), Simons Foundation Award 95177 (S.E.P.), NIH 5R01 HD057076-03-S1 (B.L.S.), NIH R01HD057076 (B.L.S.), and NIH U54MH091657 (David Van Essen), NIH P30 NS048056 (A.Z.S) and P30 NS006833 (A.Z.S.). This project was supported by the Intellectual and Developmental Disabilities Research Center at Washington university (NIH/NICHD P30 HD062171). Data were acquired with the support of NIH K12 EY16336 (John Pruett). We thank Jessica Church, Joe Dubis, Eric Feczko, Katie Ihnen, Maital Neta, and Alecia Vogel for data contribution.

## Bibliography

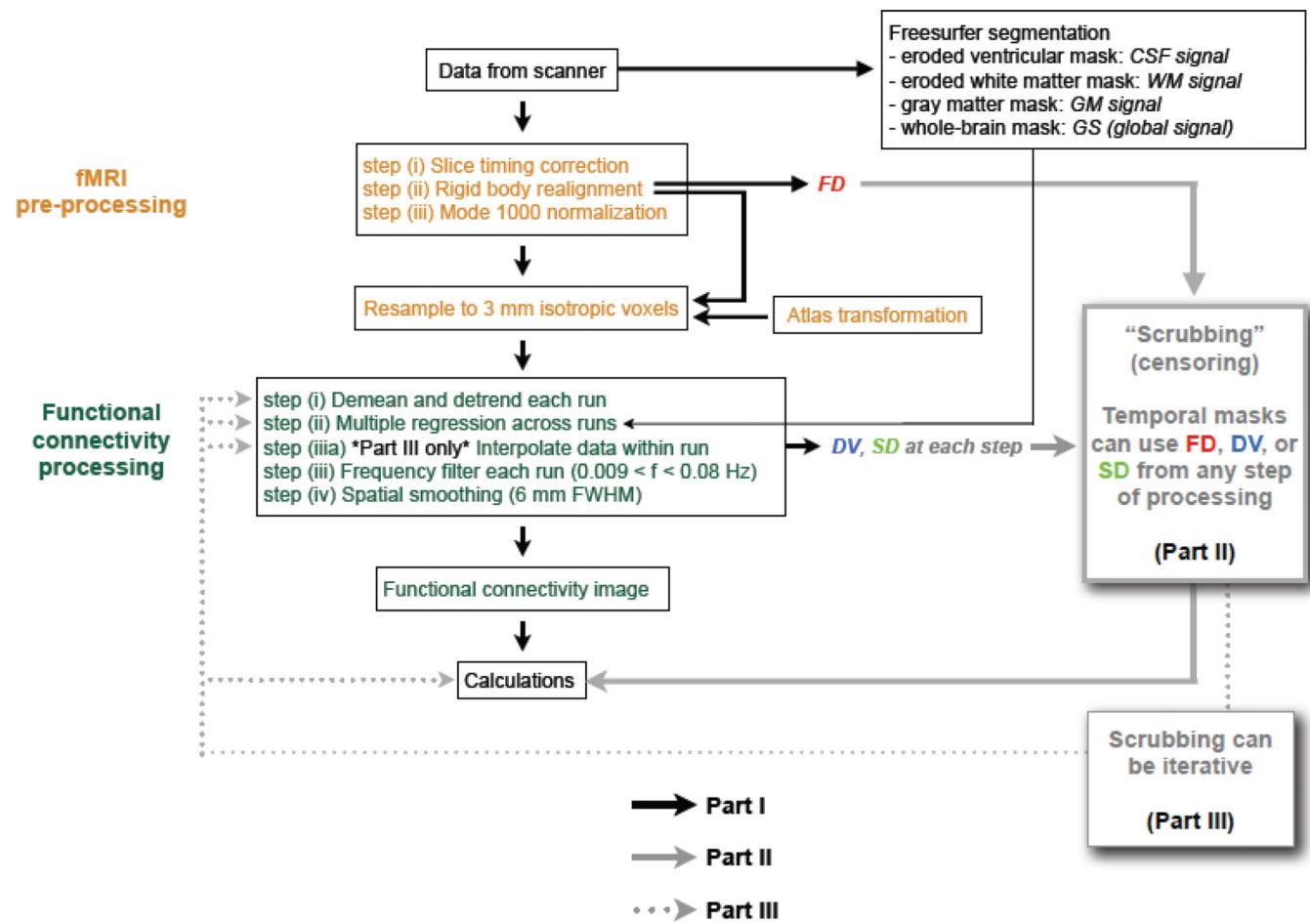
- Anderson JS, Druzgal TJ, Lopez-Larson M, Jeong E-K, Desai K, Yurgelun-Todd D. Network anticorrelations, global regression, and phase-shifted soft tissue correction. *Human brain mapping*. 2011; 32:919–934. [PubMed: 20533557]
- Bright MG, Murphy K. Removing motion and physiological artifacts from intrinsic BOLD fluctuations using short echo data. *NeuroImage*. 2013; 64:526–537. [PubMed: 23006803]
- Carp J. Optimizing the order of operations for movement scrubbing: Comment on Power et al. *NeuroImage*. 2013; 76:436–438. [PubMed: 22227884]
- Chang C, Glover GH. Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *NeuroImage*. 2010; 50:81–98. [PubMed: 20006716]
- Chen G, Chen G, Xie C, Ward BD, Li W, Antuono P, Li S-J. A method to determine the necessity for global signal regression in resting-state fMRI studies. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*. 2012; 68:1828–1835. [PubMed: 22334332]
- Cohen AL, Fair DA, Dosenbach NUF, Miezin FM, Dierker D, Van Essen DC, Schlaggar BL, Petersen SE. Defining functional areas in individual human brains using resting functional connectivity MRI. *NeuroImage*. 2008; 41:45–57. [PubMed: 18367410]
- Cordes D, Nandy RR. Estimation of the intrinsic dimensionality of fMRI data. *NeuroImage*. 2006; 29:145–154. [PubMed: 16202626]

- Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Gunther M, Glasser MF, Miller KL, Ugurbil K, Yacoub E. Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. *PLoS one*. 2010; 5:e15710. [PubMed: 21187930]
- Fox MD, Zhang D, Snyder AZ, Raichle ME. The global signal and observed anticorrelated resting state brain networks. *Journal of Neurophysiology*. 2009; 101:3270–3283. [PubMed: 19339462]
- Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R. Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*. 1996; 35:346–355. [PubMed: 8699946]
- Gotts SJ, Simmons WK, Milbury LA, Wallace GL, Cox RW, Martin A. Fractionation of social brain circuits in autism spectrum disorders. *Brain: a journal of neurology*. 2012; 135:2711–2725. [PubMed: 22791801]
- Jo HJ, Saad ZS, Simmons WK, Milbury LA, Cox RW. Mapping sources of correlation in resting state fMRI, with artifact detection and removal. *NeuroImage*. 2010; 52:571–582. [PubMed: 20420926]
- Keller CJ, Bickel S, Honey CJ, Groppe DM, Entz L, Craddock RC, Lado FA, Kelly C, Milham M, Mehta AD. Neurophysiological Investigation of Spontaneous Correlated and Anticorrelated Fluctuations of the BOLD Signal. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*. 2013; 33:6333–6342. [PubMed: 23575832]
- Kundu P, Inati SJ, Evans JW, Luh W-M, Bandettini PA. Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage*. 2012; 60:1759–1770. [PubMed: 22209809]
- Lemieux L, Salek-Haddadi A, Lund TE, Laufs H, Carmichael D. Modelling large motion events in fMRI studies of patients with epilepsy. *Magnetic resonance imaging*. 2007; 25:894–901. [PubMed: 17490845]
- Lomb NR. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*. 1976; 39:447–462.
- Mathias A, Grond F, Guardans R, Seese D, Canela M, Diebner H. Algorithms for Spectral Analysis of Irregularly Sampled Time Series. *Journal of Statistical Software*. 2004; 11:1–27.
- Mowinckel AM, Espeseth T, Westlye LT. Network-specific effects of age and in-scanner subject motion: a resting-state fMRI study of 238 healthy adults. *NeuroImage*. 2012; 63:1364–1373. [PubMed: 22992492]
- Murphy K, Birn RM, Handwerker DA, Jones TB, Bandettini PA. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *NeuroImage*. 2009; 44:893–905. [PubMed: 18976716]
- Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*. 2012; 59:2142–2154. [PubMed: 22019881]
- Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Steps toward optimizing motion artifact removal in functional connectivity MRI; a reply to Carp. *NeuroImage*. 2013; 76:439–441. [PubMed: 22440651]
- Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE. Functional network organization of the human brain. *Neuron*. 2011; 72:665–678. [PubMed: 22099467]
- Saad ZS, Gotts SJ, Murphy K, Chen G, Jo HJ, Martin A, Cox RW. Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. *Brain connectivity*. 2012; 2:25–32. [PubMed: 22432927]
- Satterthwaite TD, Elliott MA, Gerraty RT, Ruparel K, Loughead J, Calkins ME, Eickhoff SB, Hakonarson H, Gur RC, Gur RE, Wolf DH. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*. 2013; 64:240–256. [PubMed: 22926292]
- Satterthwaite TD, Wolf DH, Loughead J, Ruparel K, Elliott MA, Hakonarson H, Gur RC, Gur RE. Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. *NeuroImage*. 2012; 60:623–632. [PubMed: 22233733]

- Schölvinck ML, Maier A, Ye FQ, Duyn JH, Leopold DA. Neural basis of global resting-state fMRI activity. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:10238–10243. [PubMed: 20439733]
- Shulman GL, Pope DLW, Astafiev SV, McAvoy MP, Snyder AZ, Corbetta M. Right hemisphere dominance during spatial selective attention and target detection occurs outside the dorsal frontoparietal network. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*. 2010; 30:3640–3651. [PubMed: 20219998]
- Smith SM, Beckmann CF, Andersson J, Auerbach EJ, Bijsterbosch J, Douaud G, Duff E, Feinberg DA, Griffanti L, Harms MP, Kelly M, Laumann T, Miller KL, Moeller S, Petersen S, Power J, Salimi-Khorshidi G, Snyder AZ, Vu AT, Woolrich MW, Xu J, Yacoub E, U urbil K, Van Essen DC, Glasser MF, Consortium WU-MH. Resting-state fMRI in the Human Connectome Project. *NeuroImage*. 2013; 80:144–168. [PubMed: 23702415]
- Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR, Beckmann CF. Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:13040–13045. [PubMed: 19620724]
- Smith SM, Miller KL, Moeller S, Xu J, Auerbach EJ, Woolrich MW, Beckmann CF, Jenkinson M, Andersson J, Glasser MF, Van Essen DC, Feinberg DA, Yacoub ES, Ugurbil K. Temporally-independent functional modes of spontaneous brain activity. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109:3131–3136. [PubMed: 22323591]
- Smyser CD, Inder TE, Shimony JS, Hill JE, Degnan AJ, Snyder AZ, Neil JJ. Longitudinal analysis of neural network development in preterm infants. *Cerebral Cortex* (New York, N.Y.: 1991). 2010; 20:2852–2862.
- Tyszka JM, Kennedy DP, Paul LK, Adolphs R. Largely Typical Patterns of Resting-State Functional Connectivity in High-Functioning Adults with Autism. *Cerebral Cortex* (New York, N.Y.: 1991). 2013
- Van Dijk KRA, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL. Intrinsic Functional Connectivity As a Tool For Human Connectomics: Theory, Properties, and Optimization. *J Neurophysiol*. 2010; 103:297–321. [PubMed: 19889849]
- Van Dijk KRA, Sabuncu MR, Buckner RL. The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage*. 2012; 59:431–438. [PubMed: 21810475]
- Weissenbacher A, Kasess C, Gerstl F, Lanzenberger R, Moser E, Windischberger C. Correlations and anticorrelations in resting-state functional connectivity MRI: a quantitative comparison of preprocessing strategies. *NeuroImage*. 2009; 47:1408–1416. [PubMed: 19442749]
- Yan C-G, Cheung B, Kelly C, Colcombe S, Craddock RC, Di Martino A, Li Q, Zuo X-N, Castellanos FX, Milham MP. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *NeuroImage*. 2013; 76C:183–201. [PubMed: 23499792]
- Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, Fischl B, Liu H, Buckner RL. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*. 2011; 106:1125–1165. [PubMed: 21653723]

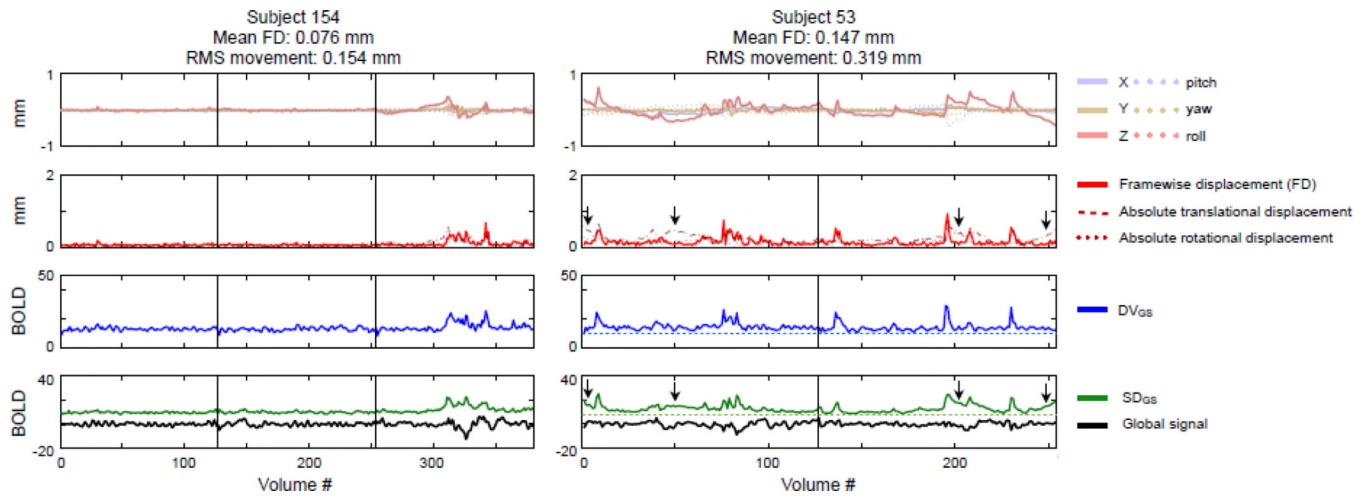
**Highlights**

- Motion-related signal changes are varied and can persist >10 seconds after motion ceases
- Such signal changes are often shared across almost all brain voxels
- Within-subject correction strategies can eliminate motion-related group differences
- Examines the linearity of motion's influence on resting state correlations



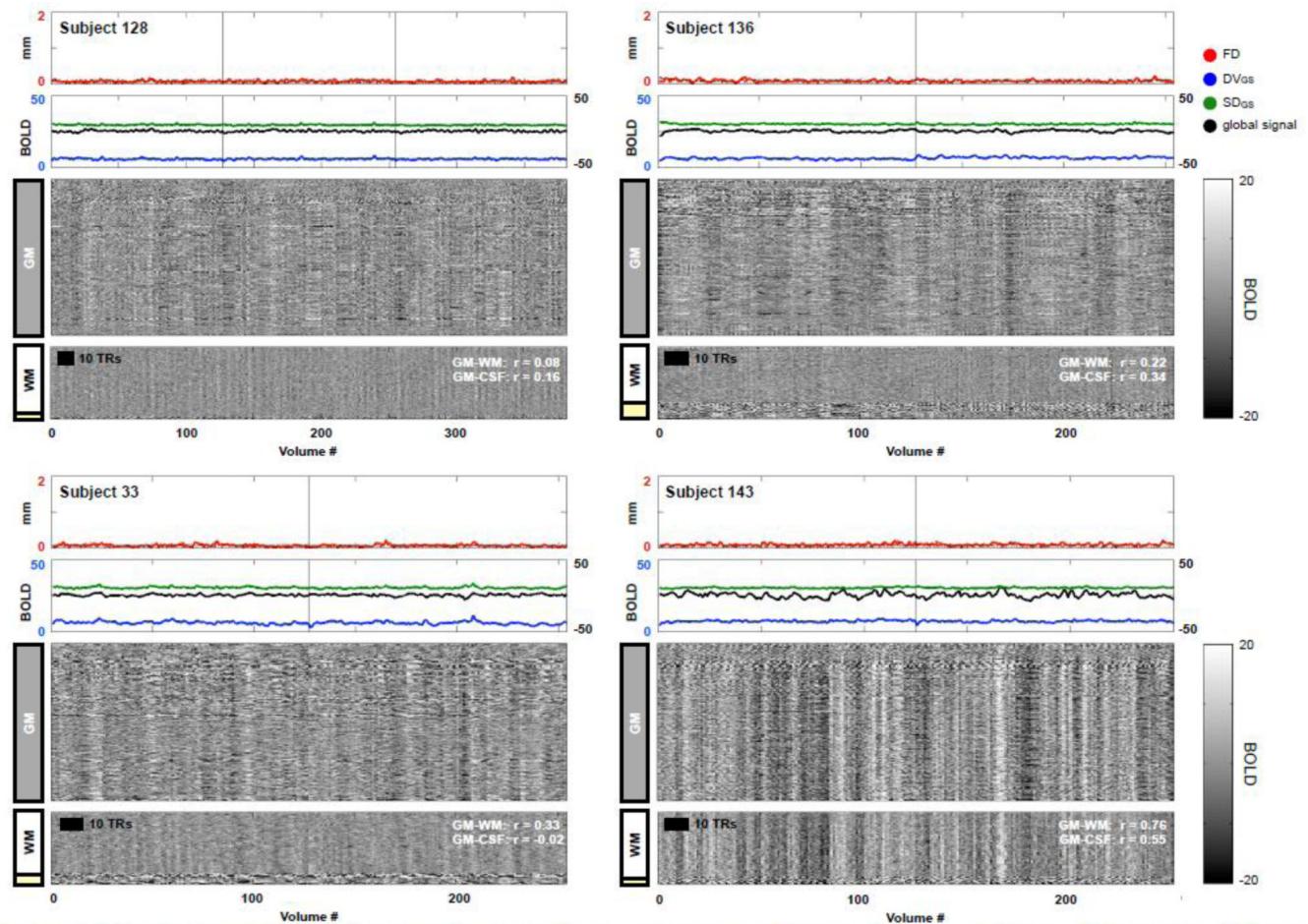
**Figure 1. Outline of data processing and scrubbing strategies**

The column of boxes in the middle depict the general BOLD processing strategy. Part I of the paper only uses the flow of the middle column (no scrubbing). The thick solid gray arrows depict scrubbing as implemented in Part II, in which censoring is only performed after the data are fully processed. The finer dotted gray arrows depict iterative processing as implemented in Part III, in which censoring is incorporated into data processing steps.



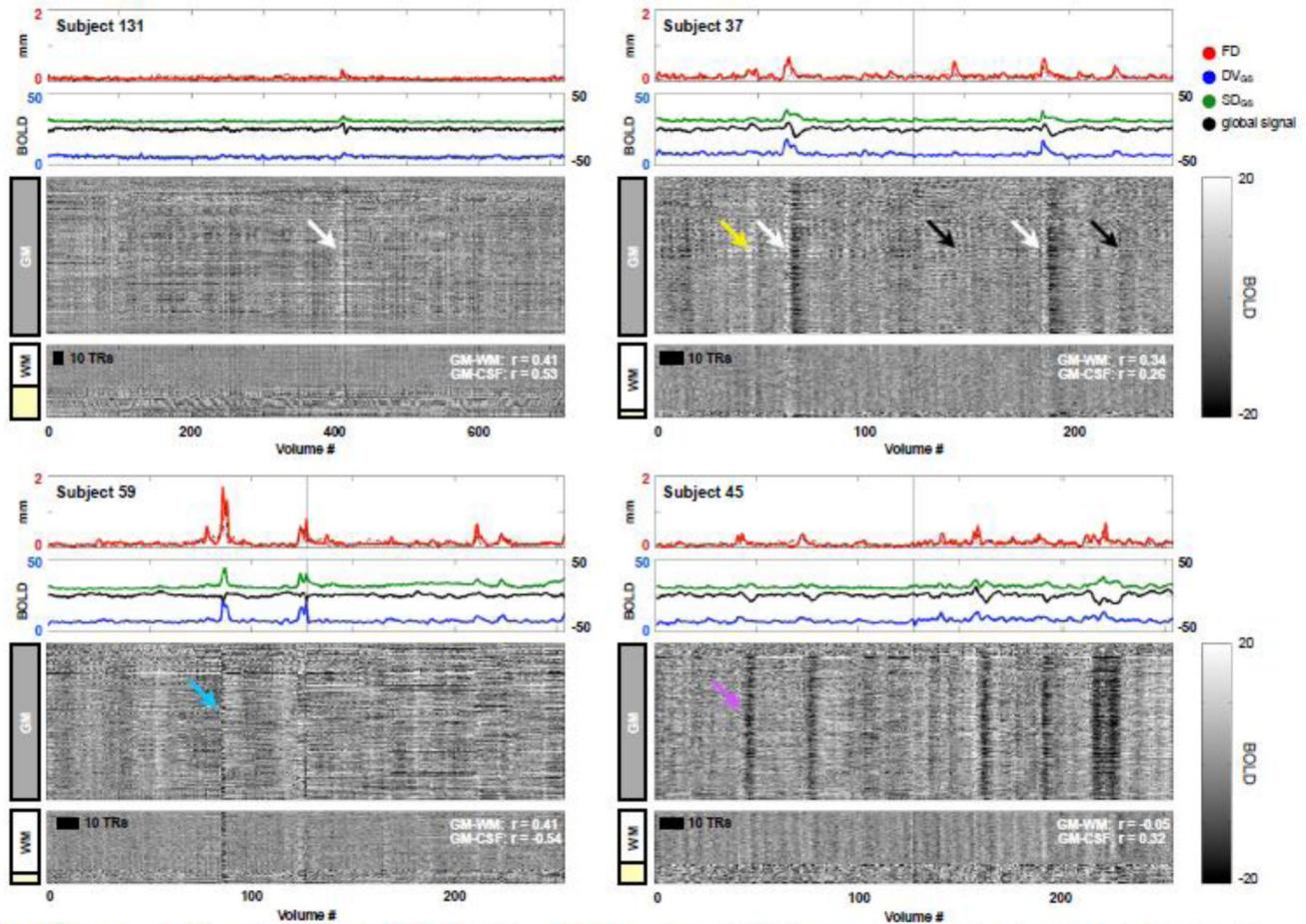
**Figure 2. Quality control measures**

At left and right are data from subjects in the low- and high-motion adult cohorts, respectively. The BOLD data have been demeaned and detrended but have not otherwise undergone functional connectivity processing. Vertical lines denote run borders. At top, the 6 rigid body realignment parameters are shown. Immediately below in red, the framewise displacement (FD) trace is shown, indexing how much the head moves from volume to volume. To give a sense of absolute rotational and translational head displacement, two traces representing the summed absolute values of the translational and rotational realignment parameters are shown. In the third panel the blue DV trace, calculated over a whole brain mask (the mask used to define the global signal, hence the GS subscript) shows the volumetric root mean squared value of the differentiated BOLD timeseries, indexing how much timeseries across the brain change from volume to volume. At bottom, the volumetric mean BOLD signal across the brain (the global signal) is shown with its volumetric standard deviation (SD). FD, DV, and SD measures are elevated during periods of motion. In addition to transient displacements, SD also tracks absolute head displacement (black arrows). RMS movement denotes root mean square realignment estimates. The dotted blue and green lines in the right panel are there simply to provide a reference from which to see elevations in the traces.



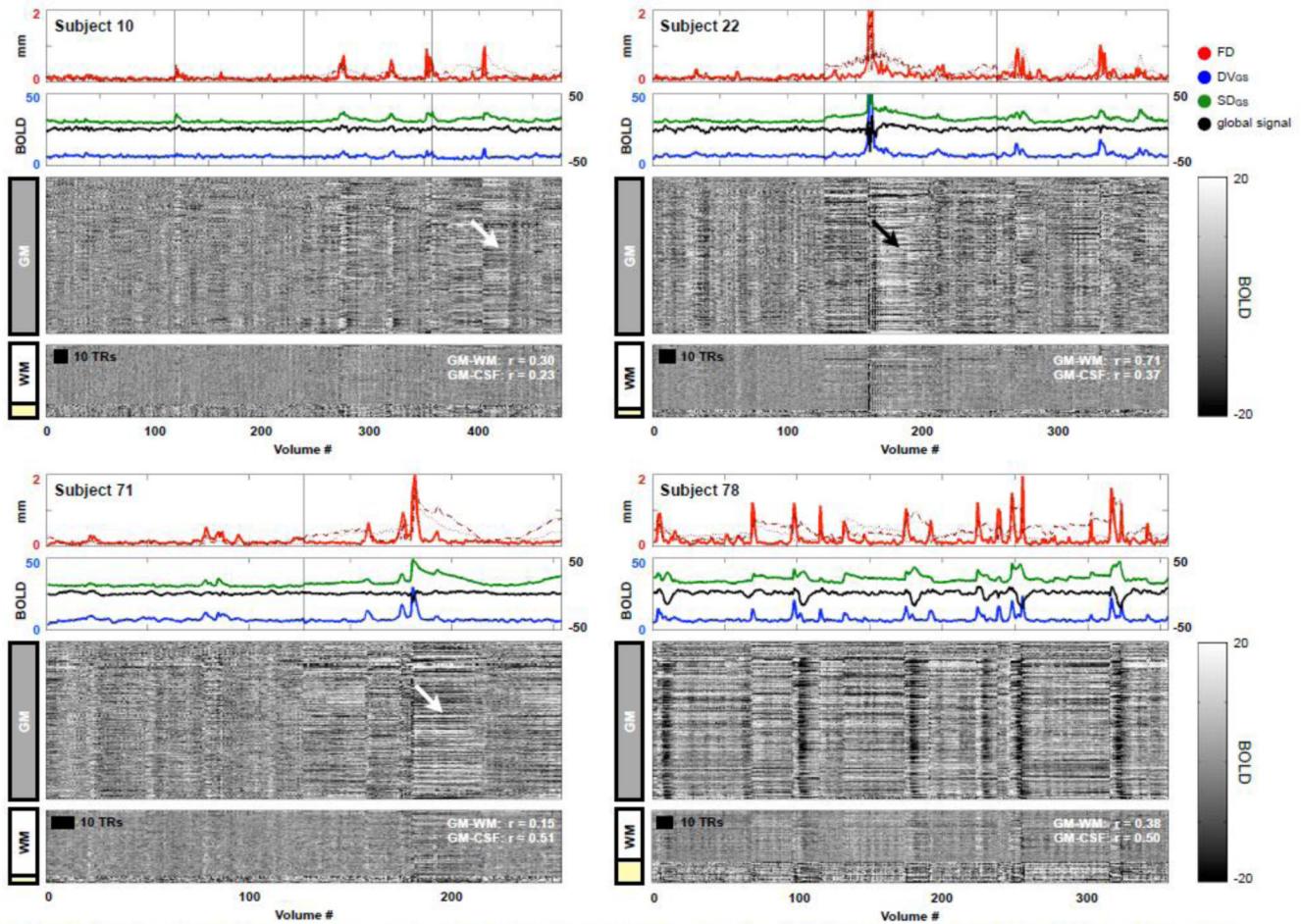
**Figure 3. QC measures and timeseries in low-motion subjects**

The upper traces in each plot are as in Figure 2. The timecourses of voxels in gray matter, white matter, and CSF are shown as intensity plots (at left, the gray bar denotes gray matter, the white bar denotes white matter, and the yellow bar denotes CSF). In the white matter plot, the black bar indicates 10 TRs of data, and the white text indicates the Pearson correlations between the mean gray matter (GM), mean white matter (WM), and mean ventricular (CSF) signals. No motion-associated variance is evident, though there are systematic fluctuations, of varying intensity in various subjects, that presumably reflect neural activity and non-motion-related noise. These data are demeaned and detrended only, as in Figure 2.



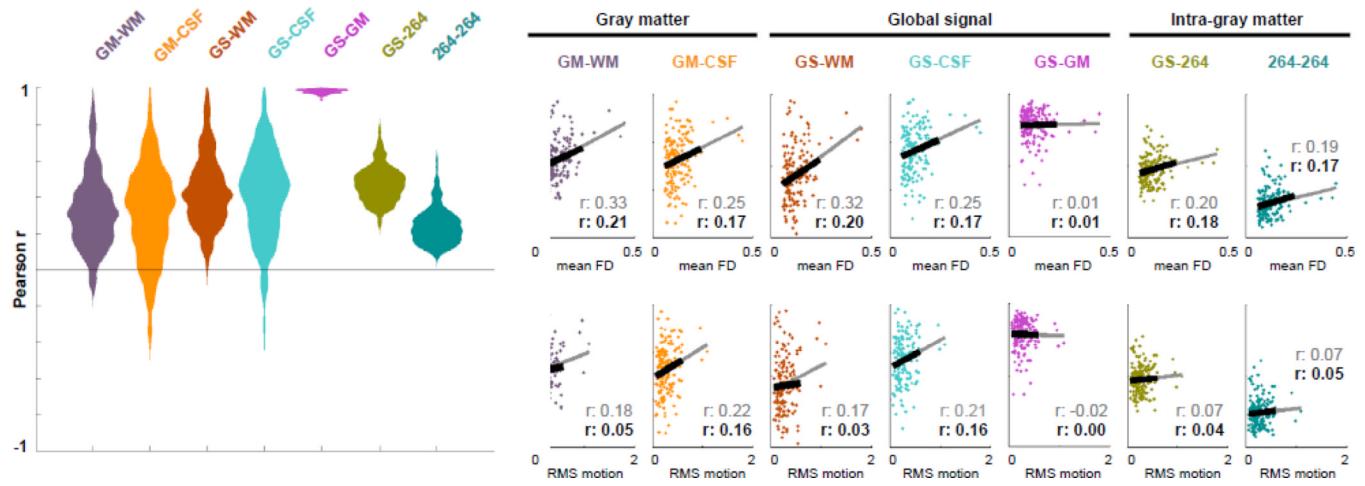
**Figure 4. QC measures and timeseries in subjects with intermittent movements**

These subjects exhibit head movements in which the head departs from and returns to positions near the origin (the dotted absolute displacement traces in the upper panels are not elevated). Motion-related signal changes can be brief or long. They can be decreases, increases, or complex waveforms. They are often but not always similar across voxels. Motion-related variance is variably reflected in white matter or CSF voxels.



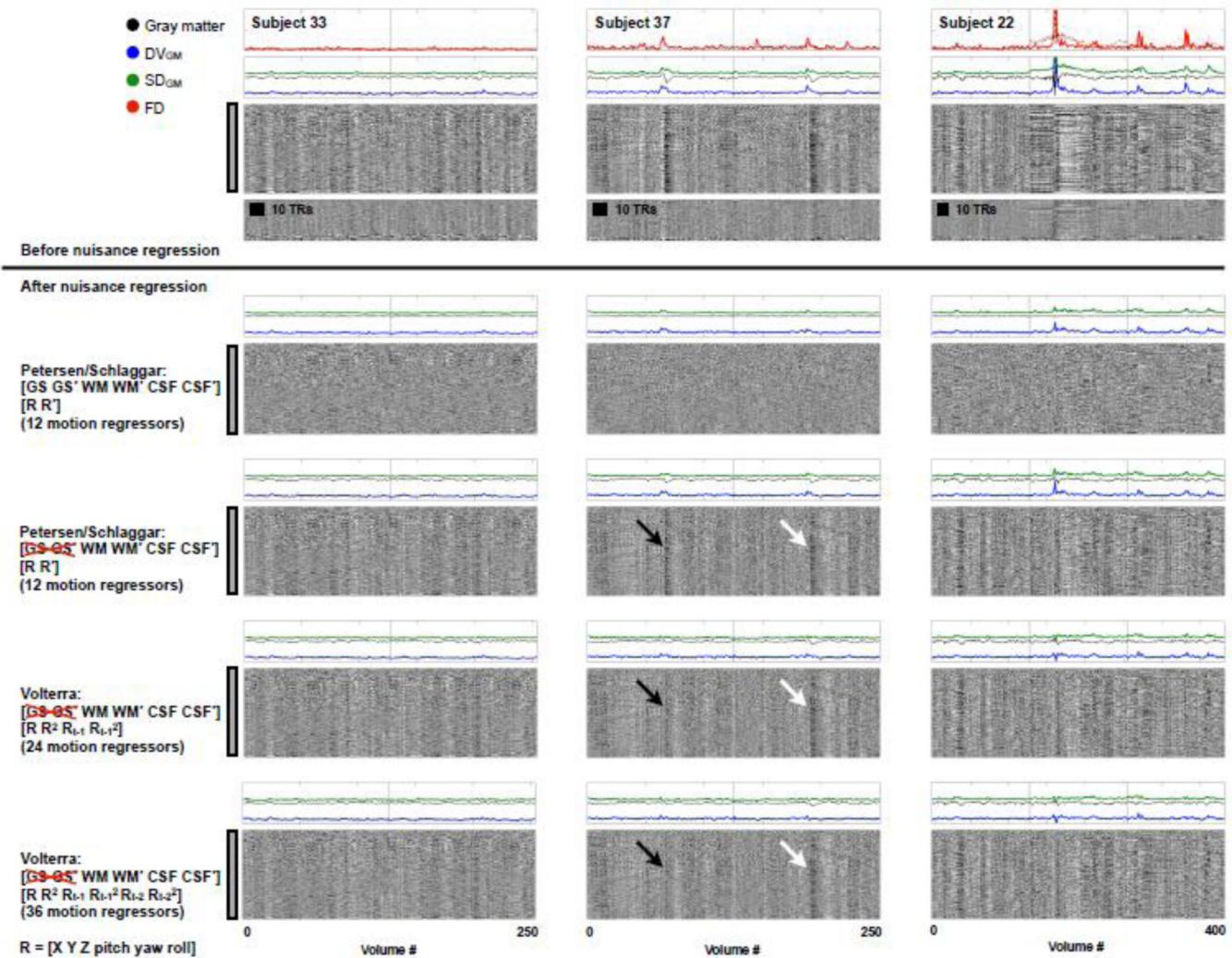
**Figure 5. QC measures and timeseries in subjects with shifted head position**

These subjects exhibit movements that displace the head from the origin over prolonged epochs (the dotted red traces in the top panels). SD traces reflect this absolute displacement. Timeseries reflect this displacement and are often elevated or depressed for long periods by shifted head position.



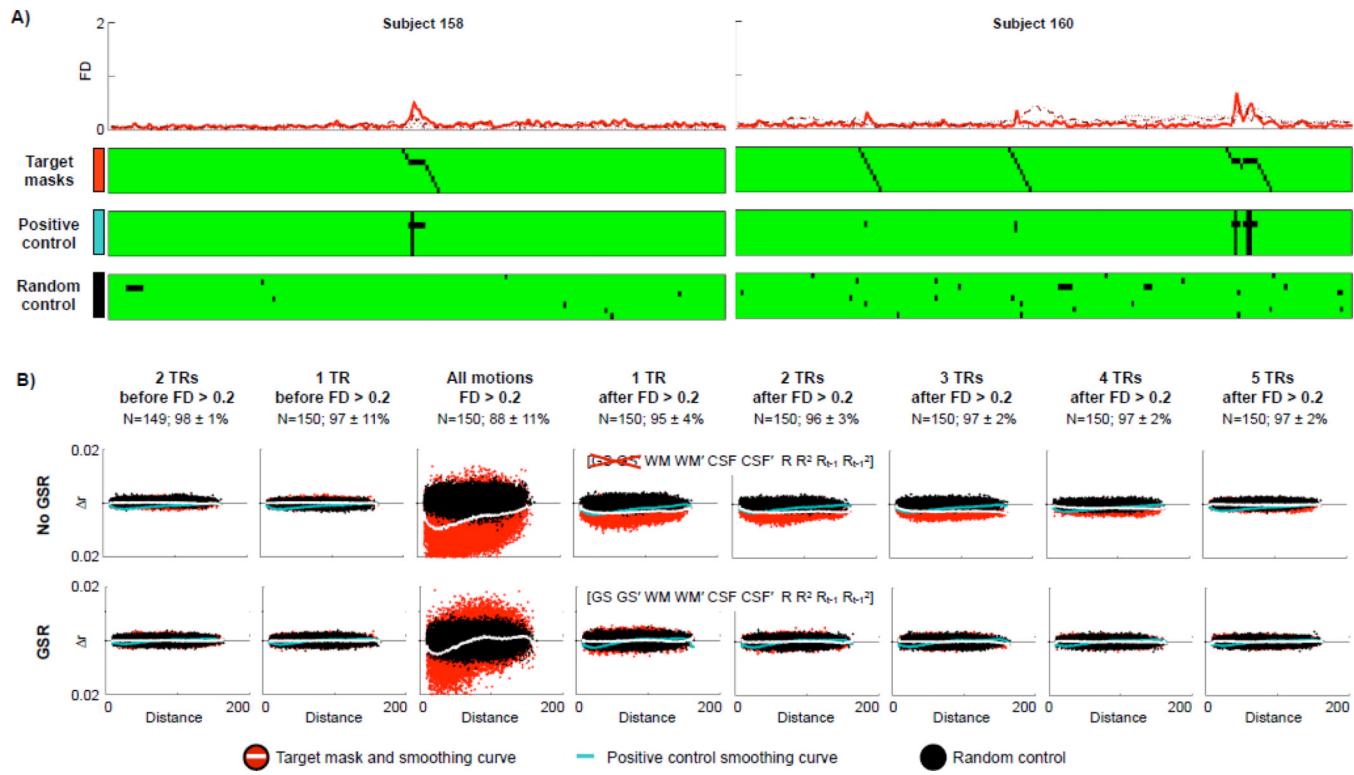
**Figure 6. Pre-regression relationships between mean compartment signals, the global signal, and ROI timecourses**

For all 160 subjects, the indicated correlations between the gray matter (GM), the white matter (WM), the ventricular (CSF), and whole-brain (GS) signals were calculated. Additionally, for 264 regions of interest, within-subject averages for the correlations of the 264 timeseries with the global signal were calculated, as well as the mean correlation over all possible pairwise correlations between the 264 ROIs. All timeseries are from demeaned and detrended data, as in Figures 3–5. The values in each subject are plotted as a function of mean FD value and RMS motion. Linear fits including all subjects (gray) or excluding outliers (black) are shown. Signal similarity, generally, is higher in subjects with more movement. Mean FD, generally, is a better predictor of signal similarity than RMS motion. Gray matter signal is highly correlated with the global signal.



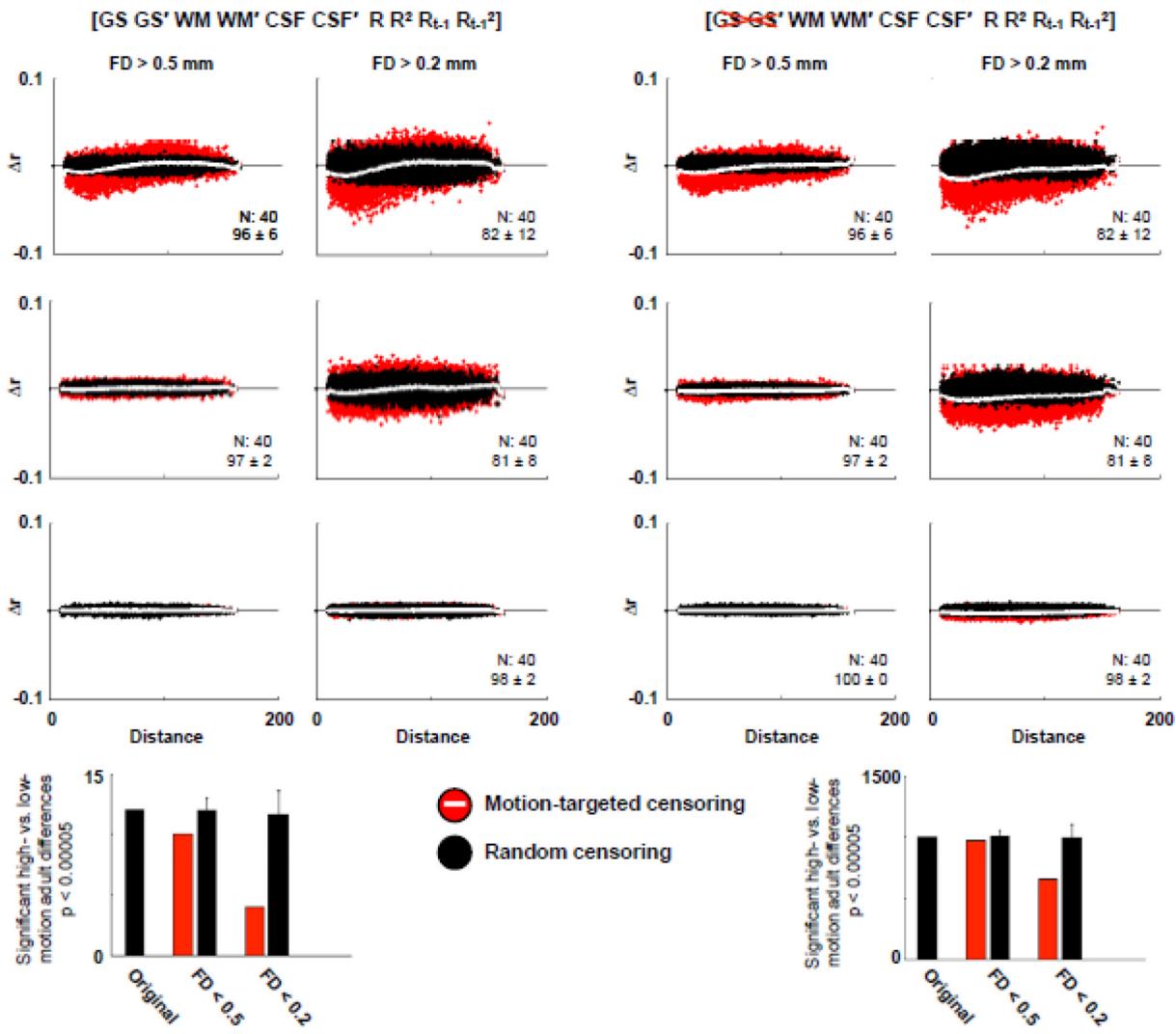
**Figure 7. Common regressions only partially remove motion-related variance**

At top, the data from 3 subjects of Figures 3–5 are re-presented, now with traces (DV, SD, mean signal) representing values derived from gray matter voxels only (instead of whole-brain values). Below the horizontal line, the data after 4 different regression strategies are shown. The top panels represent the 18-parameter regression historically used in the Petersen/Schlaggar lab (12 motion-related, 6 signal-related). The next panels are the same regressors without the global signal and its derivative. The next rows replace the 12 motion-related parameters with 24- and 36-parameter Volterra expansions of realignment estimates. Regardless of the regression strategy, the signal-derived QC measures (DV and SD) indicate artifacts in the post-regression data at periods of motion. Global signal regression visibly removes much of the motion-related signal in addition to non-motion-related signal shared across voxels. Larger numbers of motion-related regressors capture more, but not all, motion-related variance. All scales are identical to those of Figures 3–5. Similar results are seen in Figure S3, where the same analyses are repeated with no tissue-based signals (no GS, WM, or CSF signals).



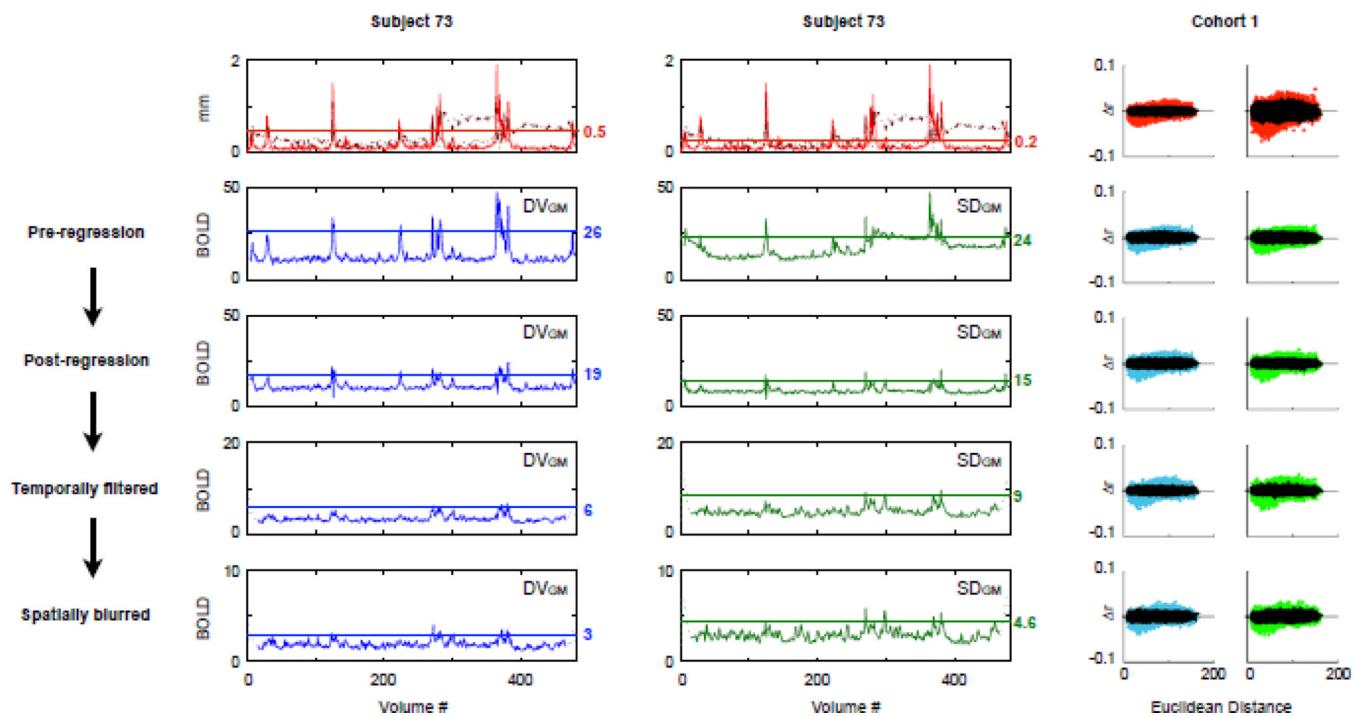
**Figure 8. The temporal limits of motion's influence on RSFC correlations**

This analysis reveals the impact on RSFC correlations of volumes acquired before, during, and after head motion. (A) Illustrations of temporal masks in two subjects. (B) For completely processed data prepared without GSR (top) and with GSR (bottom), the effects of each mask in (A) are shown.  $\Delta r$  is calculated across all subject impacted by a particular mask. The number of subjects impacted by a mask and the mean and standard deviation of remaining data are shown for each analysis (e.g., the N=150 for the 3rd mask means that 150/160 subjects had some volumes with FD > 0.2 mm, and the 10/160 who did not were not included in  $\Delta r$  calculations). TRs prior to motion were examined because frequency filtering can spread artifact backward and forward in time, and TRs subsequent to motion were examined especially due to the prolonged signal changes seen in Figure 4.



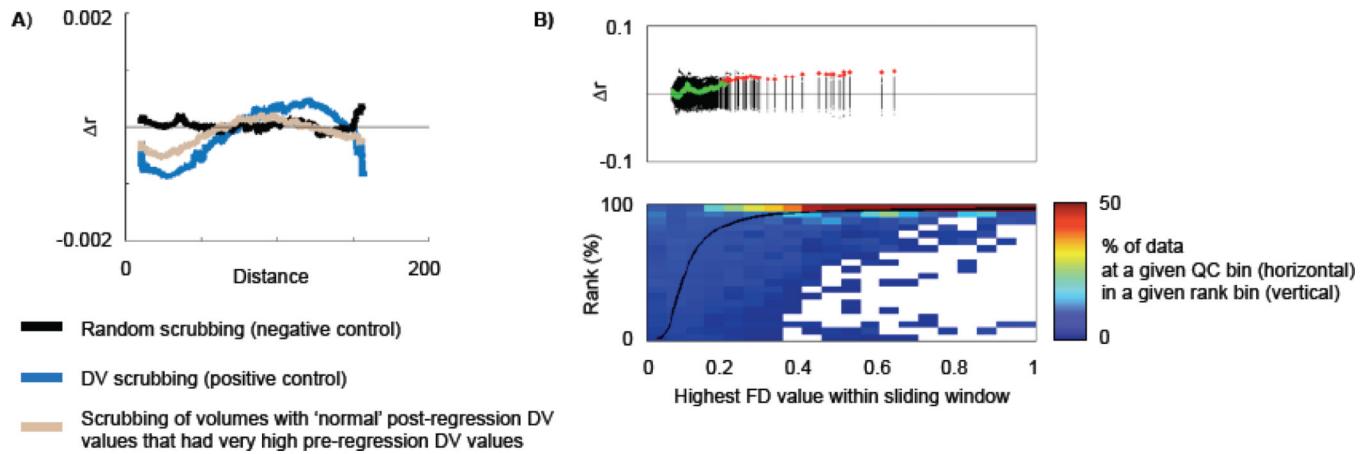
**Figure 9. Motion scrubbing selectively decreases group differences**

These plots show, for analyses with GSR (left) and without GSR (right), the distance-dependent changes in correlation produced by FD-targeted scrubbing in various cohorts at various thresholds (a lenient threshold of  $FD > 0.5$  mm and a strict threshold of  $FD > 0.2$  mm). N indicates the number of subjects in the analysis, and the numbers below indicate the mean and standard deviation of the percentage of data remaining after scrubbing. At bottom, the number of significant differences between low-motion and high-motion adult cohorts are shown, out of  $\sim 35,000$  pairwise correlations, as determined by a two-sample t-test. The error bars on the random bars are the standard deviations across 30 repetitions of random censoring. Comparisons of all adult cohorts at other statistical thresholds yield similar patterns and are shown in Figure S5. In these analyses, unlike other Figures, mean  $\Delta r$  is calculated across all subjects in a cohort, regardless of whether any volumes were censored, to illustrate the types of ‘bottom line’ changes in RSFC that would actually be seen in cohorts upon scrubbing and entire dataset.

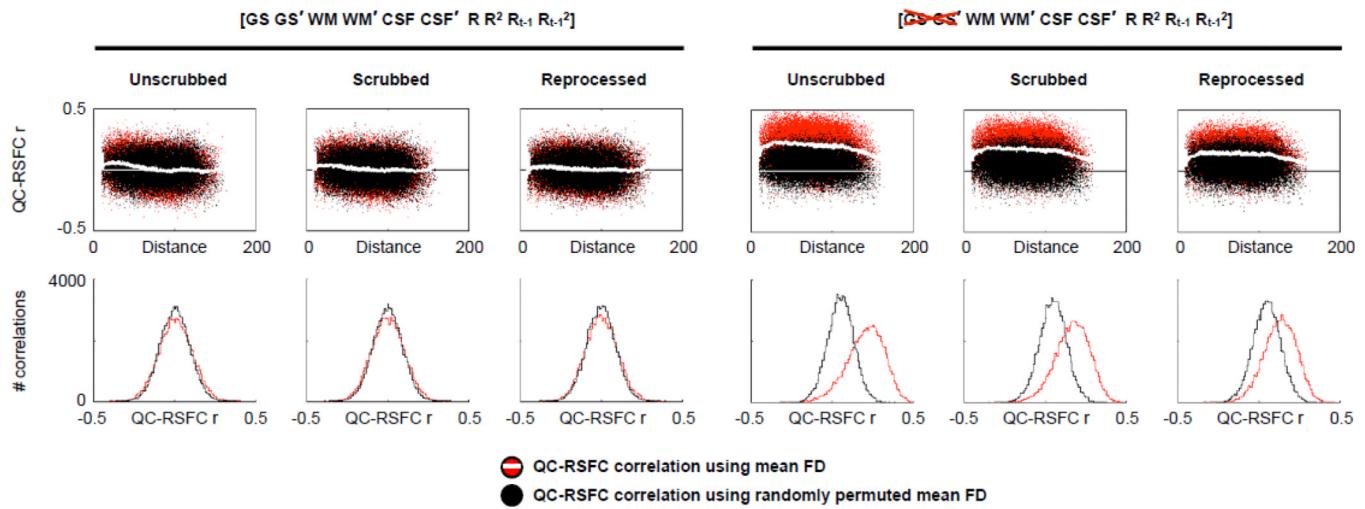


**Figure 10. An illustration of the evolution of signal-based QC measures through functional connectivity processing**

At left, for a single subject, FD traces are shown at top, and DV and SD traces are shown at different steps of functional connectivity processing. DV and SD traces evolve throughout processing. For DV and SD, the horizontal lines represent, across all subjects, a threshold 2 standard deviations above the median value (the number beside the plot). At right, in Cohort 1, the across-subject  $\Delta r$  ( $N=40$ ) produced by censoring volumes above the thresholds displayed at left. QC values from any stage of processing produce temporal masks with similar effects. These results are obtained by censoring fully processed timeseries; similar effects are seen when timeseries from any stage of processing are censored (data not shown).

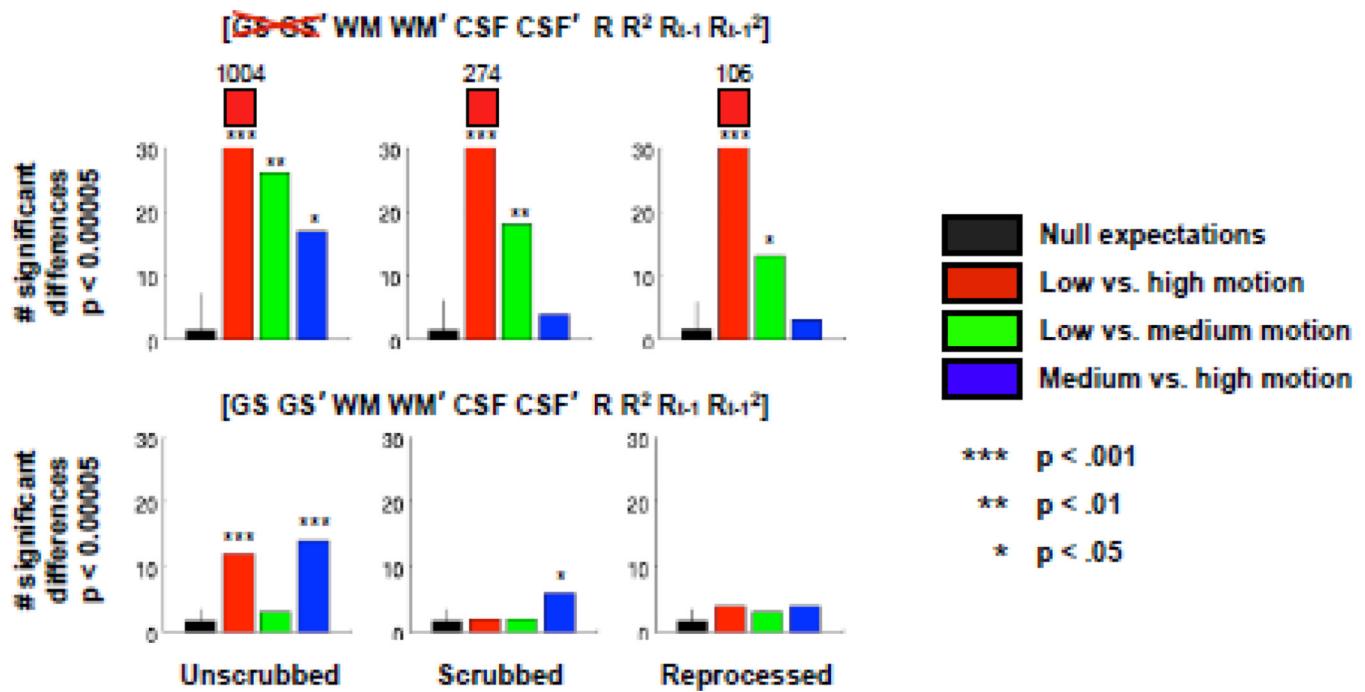
**Figure 11.**

**A)** Scrubbing of volumes that begin with high outlying DV values before functional connectivity processing but which then exhibit DV values within 1 s.d. of the median DV value following nuisance regression. The timeseries are post-regression, pre-frequency-filtering (other stages showing similar effects are shown in Figure S9). **B)** Top, for a single subject, the within-subject changes in mean short-distance correlation seen in a 50-volume sliding window are shown as a function of the highest FD value found within each window. The black points establish random expectations, and are produced by random orderings of the data. Bottom, a heat map showing the across-subject distribution of empirical ranks within binned QC ranges. Rank bins are 0–100% in 5% bins. With 20 rank bins, 5% of the data should fall in each rank bin by chance. The black sigmoidal trace is the cumulative distribution of datapoints across subjects.



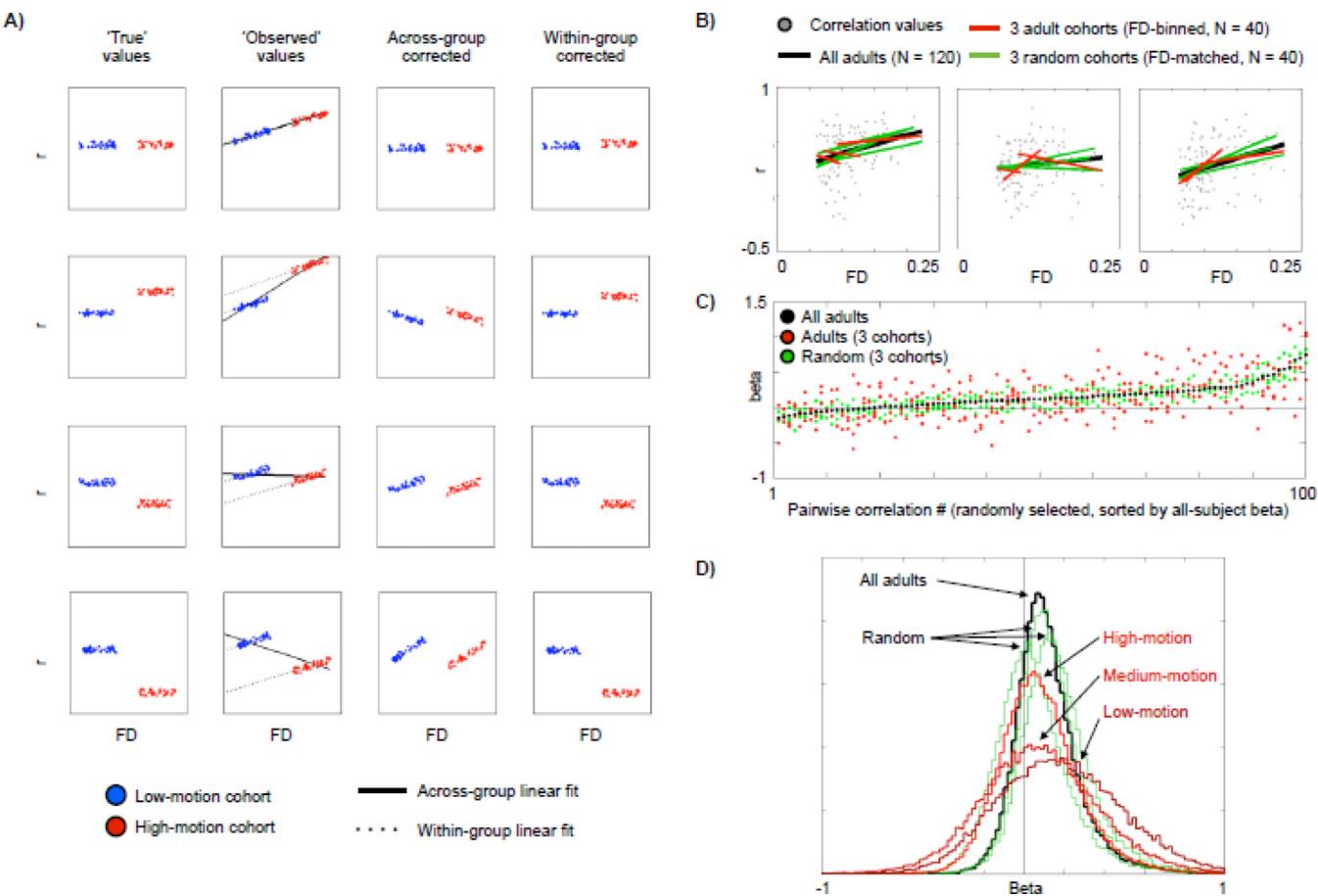
**Figure 12. Scrubbing and reprocessing reduce QC-RSFC correlations**

For all 120 adults, mean FD was correlated across subjects with each pairwise correlation under several processing regimes. The histograms plot the observed and random QC-RSFC relationships observed under each processing stream.



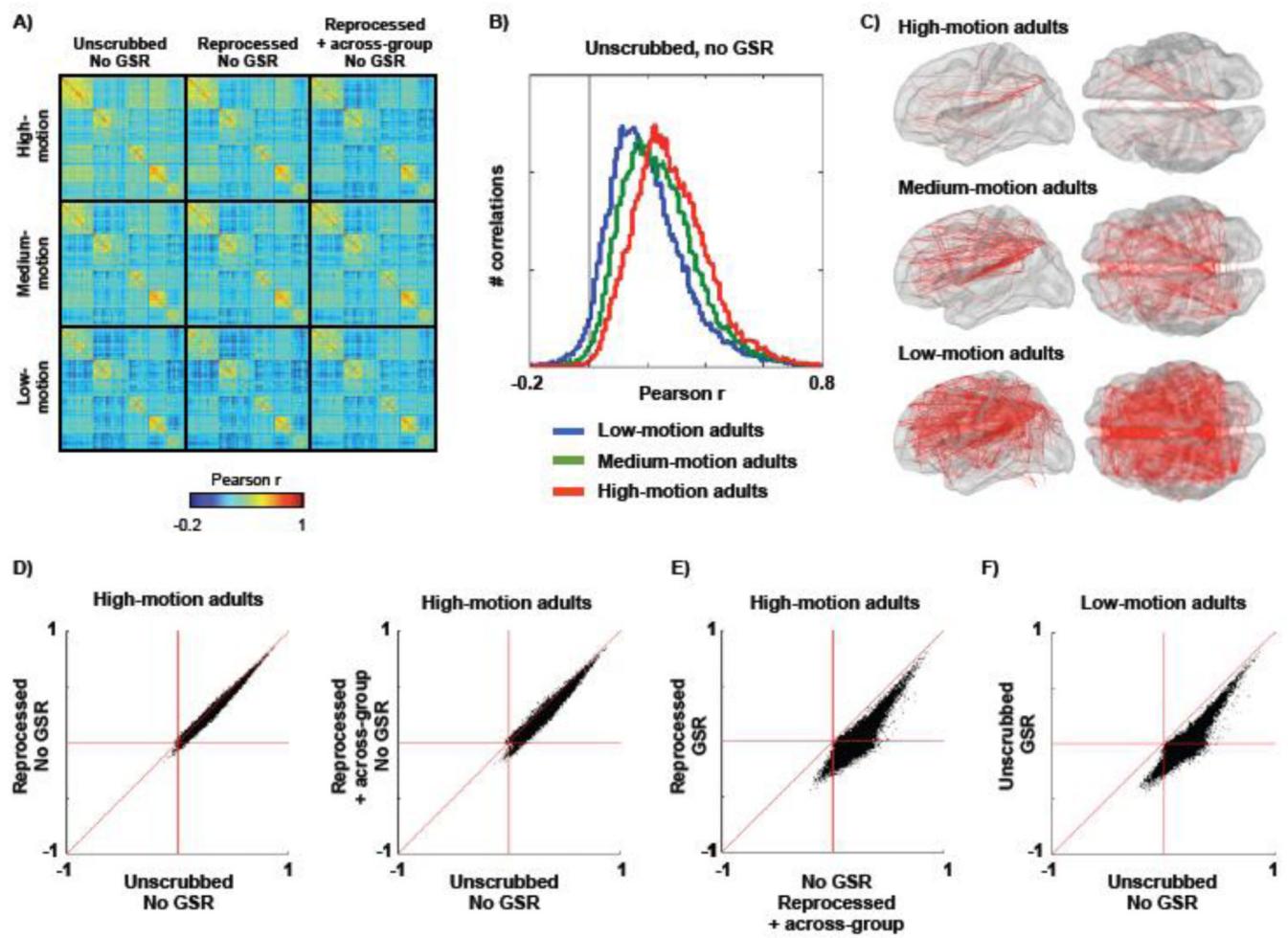
**Figure 13. Adult group differences under different processing strategies**

This figure shows the number of group differences expected by chance (black) and the number of observed differences seen between the adult cohorts under different processing strategies. The number of significant differences is defined by  $p > 0.00005$  in 2-sample t-test, as in Figure 9. Identical analyses with stricter and more lenient statistical thresholds (those of Figure S5) yield similar results and are shown in Figure S10. Permutation tests (10,000 fold) among the 120 adults established null expectations and the significance level of the observed group differences. The numbers above the red bars are the number of differences, which were much greater than the other group differences.



**Figure 14. Some limitations of group-level regression**

**A)** A diagram of how across-group and within-group regression may and may not correct artifactual increases in correlations. In these simulations, a ‘true’ value is specified for each cohort, as is an FD distribution (low-motion and high-motion). ‘Observed values’ are computed by adding a constant beta multiplied by the subject’s FD value to the ‘true’ value. ‘Across-group corrected’ values reflect residuals after a linear fit is made across both cohorts, and ‘within-group corrected’ values reflect residuals (intercepts retained) after a linear fit is made within each cohort. The ‘within-group’ correction works because the rise in  $r$  per unit FD is linear and identical in both groups. **B)** For 3 randomly selected pairwise relationships, the RSFC correlations (without GSR) of 120 adult subjects are plotted as a function of each subject’s mean FD. Lines show linear QC-RSFC fits on different subsets of the data, including all subjects (black), the actual FD-binned cohorts of the paper (red), and randomly-formed 40-subject cohorts that have indistinguishable mean FD distributions. **C)** The beta values at 100 randomly selected pairwise correlations are shown below. **D)** Across all pairwise correlations, the distribution of beta values in the different cohorts of (B, C). The betas in the low-motion cohort are higher and span a broader range of values than the betas found in the other cohorts. These data were prepared using regressors [WM WM' CSF CSF' R R<sup>2</sup> R<sub>t-1</sub> R<sub>t-1</sub><sup>2</sup>].



**Figure 15. Discussion points**

**A)** The 264–264 average correlation matrix in the adult cohorts under different processing streams. **B)** Histograms of the correlation values found without GSR in the adult cohorts. **C)** Red vectors show where negative correlations are located in (B). **D)** Plots of RSFC correlation values under different processing strategies. Points under the line have higher values under the processing indicated on the X axis. **E)** Comparison of optimal within-subject processing vs. optimal processing without GSR in high-motion adults. **F)** Comparison of processing with and without GSR in low-motion adults.

**Table 1**

Section	Objective and Figure	Finding
<b>I. Understand motion's effects on BOLD signal</b>	<ul style="list-style-type: none"> <li>See timeseries from major brain compartments (gray matter (GM), white matter (WM), and ventricles (CSF)) in still subjects (Fig. 3)</li> </ul>	<ul style="list-style-type: none"> <li>Motion-unrelated signals changes are shared across gray matter voxels and in white matter and ventricles</li> </ul>
	<ul style="list-style-type: none"> <li>See timeseries during movement (Fig. 4)</li> </ul>	<ul style="list-style-type: none"> <li>Motion-induced signal changes: variable waveforms often similar across voxels present in white matter and ventricles can persist &gt;10 seconds after motion ends</li> </ul>
	<ul style="list-style-type: none"> <li>See timeseries during head displacement (Fig. 5)</li> </ul>	<ul style="list-style-type: none"> <li>Displacement causes prolonged shifts in signal intensity</li> </ul>
	<ul style="list-style-type: none"> <li>Quantify timeseries similarity in GM, WM, and CSF</li> <li>Quantify how subject motion impacts similarity (Fig. 6)</li> </ul>	<ul style="list-style-type: none"> <li>Compartment signals correlate positively</li> <li>Compartment correlations increase with subject motion</li> </ul>
	<ul style="list-style-type: none"> <li>Visualize the efficacy of nuisance regressors, focusing on the global signal, WM and CSF signals, and several versions of motion-related regressors (Fig. 7)</li> </ul>	<ul style="list-style-type: none"> <li>Global signal markedly reduces motion-related variance</li> <li>WM and CSF signals modestly remove such variance</li> <li>Motion-related regressors alone are modestly effective Diminishing returns are seen with more regressors</li> </ul>
<b>II. Understand motion's effects on correlations</b>	<ul style="list-style-type: none"> <li>Evaluate distance-dependent artifact found in volumes before, during, and after motion (Fig. 8)</li> </ul>	<ul style="list-style-type: none"> <li>Distance-dependent artifact found with and without GSR</li> <li>Post-motion artifact found only without GSR</li> </ul>
	<ul style="list-style-type: none"> <li>Examine lenient and strict censoring criteria</li> <li>Determine whether motion censoring selectively removes motion-related group differences (Fig. 9)</li> </ul>	<ul style="list-style-type: none"> <li>As censoring becomes stricter, more artifact is removed.</li> <li>Motion-related group differences are selectively reduced by censoring motion-contaminated data.</li> </ul>
	<ul style="list-style-type: none"> <li>Show QC traces over stages of processing (Fig. 10)</li> </ul>	<ul style="list-style-type: none"> <li>QC traces evolve to possibly reflect data improvement</li> <li>“Bad” QC values denote motion artifact at all stages</li> </ul>
	<ul style="list-style-type: none"> <li>Determine whether data with initially “bad” QC values that then become “good” still harbor artifact (Fig. 11)</li> </ul>	<ul style="list-style-type: none"> <li>“Improvement” in QC values at later stages of processing is partially cosmetic; the data still harbor motion artifact.</li> </ul>
	<ul style="list-style-type: none"> <li>Determine QC thresholds below which data are “good” and above which data are “bad” (Fig. 11)</li> </ul>	<ul style="list-style-type: none"> <li>FD &gt; 0.5 mm: marked correlation changes observed</li> <li>FD = 0.15–0.2 mm: significant changes begin to be seen</li> </ul>
<b>III. Methods to suppress motion artifact</b>	<ul style="list-style-type: none"> <li>Quantify cross-subject correlation between summary QC measures and RSFC correlations (a la Satterthwaite et al.) under different processing streams (Fig. 12)</li> </ul>	<ul style="list-style-type: none"> <li>Without GSR, QC-RSFC correlations are positive</li> <li>With GSR, QC-RSFC correlations are zero-centered and overlap random expectations except at short distances</li> <li>Censoring reduces QC-RSFC correlations</li> <li>Interpolation further reduces QC-RSFC correlations</li> </ul>

Section	Objective and Figure	Finding
		<ul style="list-style-type: none"><li>With censoring and interpolation Without GSR: QC-RSFC correlations remain positive With GSR: QC-RSFC correlations are virtually eliminated</li></ul>
	<ul style="list-style-type: none"><li>Quantify the number of motion-attributable group-level differences under different processing streams (Fig. 13)</li></ul>	<ul style="list-style-type: none"><li>GSR, in combination with censoring and interpolation, reduces motion-related group differences to chance levels</li><li>Without GSR, motion-related group differences remain</li></ul>
	<ul style="list-style-type: none"><li>Outline theoretical issues with group-level correction</li><li>Assess the issues empirically (Fig. 14)</li></ul>	<ul style="list-style-type: none"><li>All group-level corrections May miss true group differences obscured by motion May remove true differences that covary with regressors</li><li>QC-RSFC correlations are often non-linear</li></ul>

**Table 2**

	<b>Why we do it</b>	<b>Possible drawbacks</b>
<b>Censoring</b>	<ul style="list-style-type: none"> <li>• Eliminates the influence of corrupted data</li> <li>• Conceptually and empirically most effective when implemented throughout a processing stream</li> <li>• Reduces dependence of correlations on motion</li> </ul>	<ul style="list-style-type: none"> <li>• Loss of data and possibly subjects</li> <li>• Unequal degrees of freedom across subjects</li> </ul> <p style="text-align: center;">But can trim to equal size</p>
<b>Interpolation</b>	<ul style="list-style-type: none"> <li>• Reduces amplitude of artifactual signal spread into adjacent TRs during frequency filtering</li> </ul>	<ul style="list-style-type: none"> <li>• Replacement data has synthetic characteristics</li> </ul> <p style="text-align: center;">Better than original characteristics</p> <p style="text-align: center;">But should probably not be treated as if it were original data (which is why we re-censor it)</p>
<b>Regression: Global signal</b>	<ul style="list-style-type: none"> <li>• Strongly reduces dependence of correlations on motion           <ul style="list-style-type: none"> <li>Most effective at medium to long distances</li> <li>Positive relationships remain at short distances, which can be suppressed by censoring and interpolation</li> </ul> </li> <li>• Eliminates post-motion influences on correlations           <ul style="list-style-type: none"> <li>Otherwise ~ 10 sec post-motion influences</li> </ul> </li> <li>• Reduces shared non-motion artifact           <ul style="list-style-type: none"> <li>Otherwise need some other type of artifact removal</li> </ul> </li> <li>• Increases RSFC correspondence to ECoG</li> </ul>	<ul style="list-style-type: none"> <li>• Removal of shared neural signal</li> <li>• If composed of few signals           <ul style="list-style-type: none"> <li>Distorted correlation structure</li> <li>Possible misattributed group differences</li> </ul> </li> </ul>
<b>Regression: white matter, CSF</b>	<ul style="list-style-type: none"> <li>• Modestly helpful at reducing artifact</li> </ul>	
<b>Regression: motion estimates</b>	<ul style="list-style-type: none"> <li>• Modestly helpful at reducint artifact</li> <li>• 24-parameter Volterra expansion increasingly used           <ul style="list-style-type: none"> <li>Superior to our old 12-parameter regression</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Even large expansions (36 parameters) are insufficient to remove motion artifact</li> <li>• 12 degrees of freedom lost per order of expansion</li> </ul>