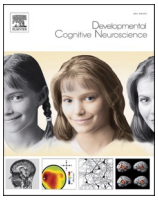
Developmental Cognitive Neuroscience 54 (2022) 101087

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

journal homepage: www.elsevier.com/locate/dcn



Benchmarking common preprocessing strategies in early childhood functional connectivity and intersubject correlation fMRI

Kirk Graff [a,b,c,d,*], Ryann Tansey [a,b,c,d], Amanda Ip [a,b,c,d], Christiane Rohr [a,b,c,d], Dennis Dimond [a,b,c,d], Deborah Dewey [a,f,g], Signe Bray [a,b,c,e]

[a] Child and Adolescent Imaging Research Program, University of Calgary, Calgary, AB, Canada
[b] Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB, Canada
[c] Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada
[d] Department of Neuroscience, University of Calgary, Calgary, AB, Canada
[e] Department of Radiology, University of Calgary, Calgary, AB, Canada
[f] Department of Pediatrics, University of Calgary, Calgary, AB, Canada
[g] Community Health Science, University of Calgary, Calgary, AB, Canada

ARTICLE INFO

Keywords:
FMRI
Functional connectivity
Motion
Connectome fingerprinting
Early childhood
Preprocessing

ABSTRACT

Preprocessing choices present a particular challenge for researchers working with functional magnetic resonance imaging (fMRI) data from young children. Steps which have been shown to be important for mitigating head motion, such as censoring and global signal regression (GSR), remain controversial, and benchmarking studies comparing preprocessing pipelines have been conducted using resting data from older participants who tend to move less than young children. Here, we conducted benchmarking of fMRI preprocessing steps in a population with high head-motion, children aged 4–8 years, leveraging a unique longitudinal, passive viewing fMRI dataset. We systematically investigated combinations of global signal regression (GSR), volume censoring, and ICA-AROMA. Pipelines were compared using previously established metrics of noise removal as well as metrics sensitive to recovery of individual differences (i.e., connectome fingerprinting), and stimulus-evoked responses (i.e., intersubject correlations; ISC). We found that: 1) the most efficacious pipeline for both noise removal and information recovery included censoring, GSR, bandpass filtering, and head motion parameter (HMP) regression, 2) ICA-AROMA performed similarly to HMP regression and did not obviate the need for censoring, 3) GSR had a minimal impact on connectome fingerprinting but improved ISC, and 4) the strictest censoring approaches reduced motion correlated edges but negatively impacted identifiability.

1. Introduction

Functional connectivity magnetic resonance imaging (FC-MRI) has become a popular tool for investigating functional brain development (Grayson and Fair, 2017; Ball et al., 2014; Marek et al., 2015) and brain-behavior associations in children (Rohr et al., 2017; Vanderwal et al., 2021; Fair et al., 2013). While FC-MRI identifies consistent FC patterns across individuals (Fox et al., 2005; Damoiseaux et al., 2006), it is highly sensitive to artifacts from physiological sources, such as heart rate and respiration, and head motion (Power et al., 2012; Satterthwaite et al., 2012; Van Dijk et al., 2012). This presents a particular challenge in fMRI studies that include young children as they have increased head motion in the scanner (Greene et al., 2018; Dosenbach et al., 2017) leading to systematic artifacts (Power et al., 2012; Fair et al., 2013). For this reason, developmental fMRI studies have struggled to collect usable resting state data from young children, and consequently have often excluded participants younger than seven (Vanderwal et al., 2019), with at least one large study revising their study protocol to omit the resting scan from children younger than six years of age due to difficulty in obtaining high-quality data (Alexander et al., 2017).

To reduce head motion and increase compliance, early childhood FC-MRI studies are increasingly conducted using passive viewing tasks such as movies (Rohr et al., 2017; Vanderwal et al., 2019; Moraczewski et al., 2018; Alexander et al., 2017; Reynolds et al., 2020). Movie watching significantly reduces head motion in participants younger than 10 years of age, with Greene et al. (2018) reporting a more than 70% reduction in

* Correspondence to: Alberta Children's Hospital Research Institute, B4-290, 28 Oki Drive NW, Calgary, AB T3B 6A8, Canada.
E-mail address: kirk.graff@ucalgary.ca (K. Graff).

https://doi.org/10.1016/j.dcn.2022.101087
Received 21 May 2021; Received in revised form 14 February 2022; Accepted 17 February 2022
Available online 18 February 2022

mean framewise displacement (FD). Importantly, FC networks derived from passive viewing paradigms have been shown to be globally similar to those derived from resting-state data (Bray et al., 2015; Greene et al., 2018; Vanderwal et al., 2019). Despite these improvements, as in all FC-MRI studies, head motion noise remains a major concern, and noise mitigating preprocessing steps are essential to prepare data for analysis. Such steps involve tradeoffs, however, and remain widely debated (Murphy and Fox, 2017; Satterthwaite et al., 2019); there is currently no gold-standard preprocessing pipeline, for early childhood studies or otherwise.

In response to the proliferation of preprocessing approaches (Carp, 2012), several studies have compared the effectiveness of different pipelines (Churchill et al., 2017; Ciric et al., 2017; Parkes et al., 2018; Taymourtash et al., 2020; Kassinopoulos and Mitsis, 2021), but no study has specifically considered the impact of preprocessing steps on data from young children, which is important for several reasons. First, children move more than adults, even when implementing strategies such as passive viewing (Greene et al., 2018; Dosenbach et al., 2017), meaning that child datasets may be particularly sensitive to the impact of preprocessing choices. Further, smaller head size could alter the impact of rotational motion, and because GSR shows distance-dependent effects, this step warrants investigation in young children. Children also have faster respiratory and heart rates than adults (Fleming et al., 2011), which may alter the temporal characteristics of physiological noise. Notably, physiological noise is not accounted for with quality control – functional connectivity (QC-FC), a widely used benchmarking metric that is specific to head motion (Satterthwaite et al., 2013; Power et al., 2012). Impacts of physiological noise may be better assessed using metrics sensitive to information recovery.

Common motion-correction strategies include regression of motion estimates, independent component analysis (ICA)-based approaches, regression of the global signal, censoring volumes of high FD, and temporal filtering. Regressing out motion estimates remains one of the most common denoising approaches (Satterthwaite et al., 2019). However, based on QC-FC benchmarks, Ciric et al. (2017) suggest that this approach is insufficient on its own and leads to concerns about losses in degrees of freedom. An alternative approach is to remove effects of head motion through ICA-based approaches, such as ICA-AROMA (Pruim et al., 2015), which decomposes data into components that reflect either brain activity or structured noise. ICA-AROMA automatically classifies components as noise using both temporal features (high frequency content, correlation with realignment parameters) and spatial features (near CSF or the edge of the brain) (Pruim et al., 2015). These structured noise components can then be regressed from the data (Thomas et al., 2002).

Global signal regression (GSR) is an often used but widely debated preprocessing step (Murphy and Fox, 2017; Chai et al., 2012; Gotts et al., 2013). GSR is a simple and arguably effective (Ciric et al., 2017; Parkes et al., 2018) denoising technique, improving the specificity of positive correlations and showing results that are more consistent with anatomical connectivity (Fox et al., 2009). However, the use of GSR tends to increase the apparent strength of short-range connections while decreasing the apparent strength of long-range connections (Saad et al., 2012; Ciric et al., 2017; Parkes et al., 2018), and creates anti-correlations which may not exist by effectively centering the connectome (Aguirre et al., 1998). The global signal has been shown to temporally resemble established networks and significantly associates with behavioral data (Li et al., 2019), suggesting GSR should be used with caution as it may remove signal of interest in addition to noise.

To mitigate the effect of specific motion contaminated volumes, it has become common to either remove, or interpolate over, specific time points by censoring or 'scrubbing' (Power et al., 2012). Despite efficacy in removing noise (Ciric et al., 2017) there are concerns with censoring, such as disrupting temporal autocorrelations, and leaving variable amounts of scan data between participants. Further, even if the number of censored volumes is matched across individuals, not all volumes are equally rich in information (Power et al., 2015). Censoring strategies vary widely across early childhood studies (Vanderwal et al., 2021; Miranda-Dominguez et al., 2018; Rohr et al., 2019; Fair et al., 2013) and it is unclear which level of censoring optimizes the recovery of individual-specific FC information.

A further step to consider is temporal filtering, as fMRI signals at both very low frequencies (under 0.01 Hz) and high frequencies (above 0.1 Hz) are variably filtered out to remove noise (Satterthwaite et al., 2013). It has been suggested that filtering above 0.1 Hz may also be removing connectivity information (Niazy et al., 2011), or artificially increasing correlations by introducing sample dependence (Davey et al., 2013), leading to concerns regarding the appropriateness of a bandpass filter compared to a highpass filter (Satterthwaite et al., 2019).

Ciric et al. (2017) speculated that aggressive preprocessing choices could be removing both signal and noise, improving outcomes on metrics sensitive to noise removed, such as QC-FC, but in the process reducing sensitivity to individual differences. Intra-class correlation (ICC; Shrout and Fleiss, 1979) is a commonly used measure of test-retest reliability, and has been used as a preprocessing benchmark of signal retention after preprocessing (Parkes et al., 2018; Kassinopoulos and Mitsis, 2021). However, both Parkes et al. (2018) and Kassinopoulos and Mitsis (2021) counter intuitively found that pipelines with poor denoising had higher ICC, suggesting higher reliability with less denoising. An alternative metric that provides information about individual-specific information remaining or recovered following preprocessing is functional connectome fingerprinting (Finn et al., 2015; subsequently referred to simply as "fingerprinting"), which aims to match scans from the same participant based on connectome similarity. Analysis of connectome individuality through fingerprinting has become popular in recent years, though how preprocessing affects FC information necessary for individual identifiability is not yet well understood.

As there is no ground-truth in resting FC-MRI analyses, it has been suggested that preprocessing benchmark studies should ideally consider task-evoked effects to ensure that while noise is mitigated signal variation of interest is preserved (Bijsterbosch et al., 2021). Here we leverage the passive viewing task to assess how preprocessing choices impact functional responses to the video using temporal intersubject correlation (ISC; Hasson et al., 2004). ISC is calculated as the temporal signal correlation between a given brain region across individuals during passive viewing of the same video stimulus.

The aim of the present study is to extend previous pipeline benchmarking studies (Ciric et al., 2017; Parkes et al., 2018) to passive viewing fMRI data from young children, while extending benchmarks beyond metrics sensitive to head motion, namely with the inclusion of ICC, fingerprinting, and ISC. As preprocessing steps likely have a larger impact in noisier samples, we also investigate whether preprocessing strategies are particularly efficacious depending on quantity of head motion by dividing scans into lower- and higher- motion groups. Our findings can support researchers conducting fMRI in early childhood samples to consider the tradeoffs and effectiveness of common preprocessing steps.

2. Methods

2.1. Participants

Data were collected as part of a longitudinal study of early childhood brain development (Dimond et al., 2020a, 2020b; Rohr et al., 2019, 2017). Participants were recruited from the local community through advertisements and existing databases. All procedures were approved by the University of Calgary Conjoint Health Research Ethics Board. Parents provided informed consent and children provided assent to participate. Participants were children between 4 and 7 years of age at baseline without any major health concerns. Children were excluded if

they had full-scale IQ more than 2 standard deviations below the standardized population mean of 100 on the Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition, a history of neurodevelopmental or psychiatric disorders, or any neurological diagnoses. At the time the analyses for this study were initiated, 168 participants had completed a baseline scan and of these 59 (15 male) had completed a 12-month follow-up scan. From the sample of participants with both baseline and follow-up data, participants were included if after volume-wise censoring of the fMRI data (described in more detail below), both scans retained at least 11 min of uncensored data. 56 of the 59 children (14 male) reached this threshold, for a total of 112 scans used in the present study.

2.2. Data collection

MRI data were acquired on a 3 T GE MR750w MRI (Waukesha, WI) scanner using a 32-channel head coil, at the Alberta Children's Hospital. fMRI was acquired during a passive-viewing task, where participants watched clips from a children's television show (Elmo's World) for 1100 s. Prior to scans, children underwent a practice scan in an MRI simulator during which they watched the same video and practiced staying still. fMRI scans were collected using a gradient-echo EPI sequence (TR = 2.5 s, TE = 30 ms, FA = 70°, voxel size 3.5 ×3.5 ×3.5 mm^3). An anatomical scan was acquired using a T1w 3D BRAVO sequence (TR = 6.764 ms, TE = 2.908 ms, FA = 10°, voxel size 0.8 ×0.8 ×0.8 mm^3).

2.3. Higher- vs. lower-motion subgroups

Preprocessing steps may impact data differently depending on the amount of contamination from motion or other sources. Low motion samples may see only modest improvements with more aggressive strategies such as censoring or GSR, whereas metrics derived from high motion samples will likely change more if, for example, a greater number of volumes are censored. Therefore, we consider here how preprocessing may differentially impact lower- and higher- motion scans. Towards this goal, after preprocessing pipelines were first compared using the whole sample of 112 scans, the 112 scans were median-split into two groups of 56 scans, based on average FD. The preprocessing pipelines were then separately compared for the higher- and lower-motion subgroups, using the metrics described below.

2.4. First-stage preprocessing steps common across pipelines

The following first-stage preprocessing steps were run on all scans, ahead of the specific pipeline variations tested here (described below). All preprocessing was carried out with custom Python scripts integrating Nipype functionality (version 1.1.5; Gorgolewski et al., 2011) using FSL version 6.0.0 (Smith et al., 2004), ANTs version 3.0.0.0 (Avants et al., 2011), and AFNI version 18.3.03 (Cox, 1996). Scans from the same individual were preprocessed separately. Structural (T1w) images were preprocessed using ANTs, including bias field correction, brain extraction, and tissue segmentation. For generation of WM and CSF time courses for regression, tissue masks were eroded using AFNI (CSF eroded twice, WM eroded 7 times).

Basic preprocessing for EPI data was as follows, largely following the procedure described by Ciric et al. (2018): a) FSL *MCFLIRT* to estimate head motion parameters (HMPs) and FD (Jenkinson et al., 2002). b) FSL *slicetimer* for slice time correction. c) FSL *MCFLIRT* for rigid body realignment. We followed the recommendation in Power et al. (2017) to estimate HMP on raw data but carry out slice time correction prior to rigid body realignment, necessitating two FSL *MCFLIRT* steps. d) FSL *BET* to skull-strip EPI images (Smith, 2002). e) ANTs *Registration* (Avants et al., 2011) to generate a transformation matrix to warp the EPI image to a study-specific EPI template. This template was produced based on the procedure described by Huang et al. (2010). Specifically, a 3D EPI reference image was taken from each fMRI scan, chosen as a volume of low motion approximately in the middle of the scan. These references were warped to MNI space, then averaged together and smoothed to create the final study-specific template. f) FSL *FLIRT* boundary-based registration (Jenkinson et al., 2002) was used to generate a transformation matrix to warp the EPI image to the T1w image, then the inverse transformation matrix was used to warp tissue segmentations to functional image space. All preprocessing and confound mitigation steps were carried out in native space. g) A linear regression to remove the mean, linear trends, and quadratic trends from each voxel was conducted. For pipelines that include censoring, time points marked for censoring were excluded from the model to calculate linear and quadratic trends. In pipelines that did not include censoring, all time points were included.

2.5. Preprocessing pipelines

After first-stage preprocessing, pipelines varied systematically in whether they used GSR, censoring, and ICA-AROMA, as shown in Table 1. For comparison, we also tested two minimal pipelines: M1, which included first-stage preprocessing and temporal filtering, but no nuisance regression, and M2, which included first-stage preprocessing, temporal filtering, and regression of the WM and the CSF time series.

2.6. Additional preprocessing

2.6.1. ICA-AROMA

For pipelines I1 through I4, ICA-AROMA was applied immediately following the first-stage preprocessing steps described above. Given that this was a relatively high motion sample, we used ICA-AROMA's aggressive denoising feature, where all variance associated with noise components are removed (Pruim et al., 2015), but otherwise default options. The ICA-AROMA code was modified to warp to MNI space via the ANTs transformation matrix rather than a FNIRT transformation.

2.6.2. Temporal filtering

For all pipelines, data were temporal bandpass filtered (0.01 – 0.08 Hz) using a fast Fourier transform. For pipelines I1 through 14, this occurred following ICA-AROMA. For all other pipelines, this occurred immediately after first-stage preprocessing. In follow-up comparisons we tested the effect of a highpass filter; see Section *2.9.1. Filtering.* To avoid reintroducing artifacts, HMPs were also filtered, and WM, CSF, and the global signal were calculated following temporal filtering (Lindquist et al., 2019).

2.6.3. Nuisance regression

Following filtering, nuisance regression was applied as part of all pipelines, with the exception of M1, which did not undergo further preprocessing. Excepting M1, all pipelines included WM + CSF regression. In pipelines R1 through R4 the six HMPs were regressed. Pipelines

Table 1
Preprocessing pipelines tested to compare effects of ICA-AROMA, global signal regression and censoring.

Pipeline	Primary motion artifact removal	WM + CSF regression	Global signal regression (GSR)	Censoring
M1	None			
M2		Yes		
R1	Regress head	Yes		
R2	motion parameters	Yes	Yes	
R3	(HMPs)	Yes		Yes
R4		Yes	Yes	Yes
I1	ICA-AROMA	Yes		
I2		Yes	Yes	
I3		Yes		Yes
I4		Yes	Yes	Yes

R2, R4, I2 and I4 included GSR. For all nuisance parameters, both linear and quadratic terms, along with the first temporal derivative of those terms were regressed (i.e., 4 regressors per parameter). Censoring was carried out as part of the regression step for pipelines R3, R4, I3 and I4. As a default, we censored volumes above a FD threshold of 0.25 mm (based on FSL *MCFLIRT*, i.e., $FD_{Jenkinson}$ as described in Ciric et al., 2018), and censored only the identified frames. Varied censoring schemes were tested separately as described below. Pipelines are summarized in Table 1.

2.7. Connectome generation

Following preprocessing, each scan was registered to the study specific template using the ANTs transformation matrix. Each voxel was then assigned to one of 325 nodes (regions) within the MIST 325 parcellation (Urchs et al., 2019). The mean time course was calculated for each node by averaging the time courses for all voxels within the node. The Pearson correlation between each pair of nodes was then calculated, generating 52,650 edges. Correlation values were Fisher z- transformed to better approximate a normal distribution.

2.8. Pipeline comparison metrics

2.8.1. Motion correlated edges (QC-FC)

QC-FC was calculated using the approach described in Satterthwaite et al. (2013). Specifically, for each edge the correlation was calculated between average FD (as estimated from the first FSL *MCFLIRT* step) and edge strength, across all 112 scans. Due to the concern that preprocessing steps such as GSR affect edges differently depending on the inter-node distance, we also assessed distance-dependent effects by plotting the edge strength-motion correlation vs. Euclidian inter-node distance for each edge (Satterthwaite et al., 2013). Inter-node distance was calculated as the Euclidian distance between the center of mass of the two nodes linked by a given edge, as defined by the MIST parcellation. Three QC-FC metrics were extracted: (1) the percentage of edges correlated with head motion at a p-value < 0.05 uncorrected; (2) the mean of the absolute value of the correlation between head motion and edge strength; absolute value was used to assess the magnitude of the motion-effect, rather than the direction; and (3) QC-FC distance dependence as the correlation between edge strength-motion correlation and Euclidian edge length. In theory, more effective preprocessing pipelines will lead to fewer edges significantly associating with head motion, and ideally this relationship would not depend on distance, such that across these three metrics, smaller values are considered better.

2.8.2. Fingerprinting

Pipelines that aggressively remove noise may also have the undesirable side-effect of removing signal of interest (Ciric et al., 2017). We used functional connectome fingerprinting to determine whether information unique to an individual is reduced or amplified by noise-mitigation steps. Following the approach used by Finn et al. (2015) and related work (Byrge and Kennedy, 2019; Kaufmann et al., 2017; Miranda-Dominguez et al., 2018), connectomes were vectorized, then correlated for each pair of scans. The fingerprinting match rate was calculated by determining how often a scan correlated most strongly with the second scan from the same individual, and dividing by the total number of scans (112). We expressed this as a percentage. For each scan we also calculated 1) group similarity, i.e., the mean correlation to a scan from another individual, 2) stability, i.e., the correlation between scans from the same person, and 3) individualization, the difference between stability and the highest correlation to a scan from another individual (nearest miss). Negative individualization scores reflect scans that did not successfully match. To assess whether pipelines affected associations with head motion, for each pipeline we regressed stability and individualization against head motion, calculating R^2, the slope, and the intercept. For each individual, we used the higher of their scans' two motion values for this analysis.

2.8.3. Intra-class correlation (ICC)

For each pipeline, we calculated the intra-class correlation (ICC; Shrout and Fleiss, 1979) of each edge as a measure of test-retest reliability. ICC was calculated in two ways: 1) a between session ICC, where we used the two scans from each individual to calculate the within-subject mean square, and 2) a within-session (split-half) ICC for each participant's first session, where each scan's time series was split in two in order to calculate the within-subject mean square. For pipelines that included censoring, split-half time series were generated based on volumes retained after censoring.

2.8.4. Intersubject correlation (ISC)

We used intersubject correlation (ISC; Hasson et al., 2004) to assess whether stimulus-evoked activity is amplified or suppressed by noise-mitigation. Preprocessing pipelines that fail to adequately remove noise, or remove task-evoked signal, will have lower ISC values. ISC values were calculated as the temporal correlation between pairs of individuals for each node, and averaged across pairs. For pipelines that included censoring, we did not include time points that were censored for either scan in a given pair. ISC calculated on passive viewing fMRI are typically highest in visual and auditory regions (Hasson et al., 2004; Kauppi et al., 2010). For pipeline comparisons, we identified the 10 nodes with the highest average ISC across all pipelines and averaged these generate a single ISC score for each scan for each pipeline. In addition to group-level ISC, we calculated the "intra-subject correlation", or the time series correlation between scans from the same participant.

As noted above, scans were median-split into higher- and lower-motion subgroups. For ISC, when both of a participant's scans were in the same subgroup, we only retained the more representative scan, i.e., the lower motion scan in the lower-motion subgroup, or the higher motion scan in the higher-motion subgroup.

2.8.5. Intra-scan inter-pipeline correlation

Preprocessing choices may have relatively large or relatively small effects on connectomes, and these may influence downstream analyses and convergence across studies. To assess the impact of preprocessing choices on FC estimates, for each scan we calculated the correlation between edge strengths across pipelines. These were Fisher z-transformed, averaged across the 112 scans, then converted back to correlations for ease of comparison.

2.9. Follow-up comparisons

2.9.1. Filtering

In addition to the highpass filter applied as part of nearly all fMRI analyses, the use of a lowpass filter has become common as a noise mitigation step in FC-FMRI due to frequencies above approximately 0.1 Hz being more highly associated with noise than signal of interest (Satterthwaite et al., 2019). However, this remains controversial, as connectivity information at higher frequencies will be lost (Niazy et al., 2011). We therefore repeated benchmark comparisons using a highpass ($>$0.01 Hz) rather than a bandpass filter. All other preprocessing steps were identical.

2.9.2. Varying censoring parameters

We conducted an additional comparison focusing specifically on the effect of different censoring strategies. There is variation in thresholds used in the literature, with older studies generally using more lenient censoring relative to more recent work (Satterthwaite et al., 2019). There is also variation in the literature as to whether censoring a single volume per motion artifact (Satterthwaite et al., 2013) or censoring multiple volumes per motion artifact (Power et al., 2012) is preferred. Thresholding decisions are challenging because there is limited guidance in early childhood samples.

We used pipeline R4 above, which performed well on all metrics, and modified the approach to censoring using 11 additional pipelines that varied in the censoring threshold and the volumes censored per motion artifact (Table 2). In pipelines that censored 3 volumes per motion artifact, we censored volumes immediately before and after any volume or series of volumes flagged for censoring. Similarly, for pipelines that censored 4 volumes per motion artifact, 1 vol prior and 2 volumes after the motion-contaminated volume(s) were censored. All 112 scans were included for these comparisons, even if stricter censoring caused scans to fall below the threshold of 11 min of uncensored data.

2.10. Statistical Comparisons

To statistically assess the impact of specific preprocessing steps, we compared pairs of pipelines that differed in their use of GSR, censoring, ICA-AROMA, or temporal filtering. For most analyses, pipelines were compared using paired samples t-tests with uncorrected p-values and Cohen's d (mean difference ÷ standard deviation of the difference) reported. For correlation-based metrics – namely QC-FC distance dependence, individualization by motion, and stability by motion – we used Steiger's test to assess whether correlations were statistically different. We note that because QC-FC and ICC analyses are edgewise comparisons (n = 52,650), while fingerprinting and ISC analyses are scanwise comparisons (n = 112), p-values are not directly comparable across metrics.

3. Results

3.1. Participant characteristics

The mean age of participants at baseline was 5.47 years old (standard deviation: 0.76 years), and 6.52 for the second scan. Time from initial to follow up scan ranged from 0.88 to 1.19 years (mean: = 1.05 years, SD = 0.069 years). The average motion of our sample was high, with a mean average FD of 0.126 mm across all 112 scans (median average FD 0.095 mm). For comparison, in a study of adolescents aged 8–23, the whole sample had a mean average FD of 0.062 mm, with lower- and higher-motion subgroups having a mean average FD of 0.029 mm and 0.097 mm respectively (Satterthwaite et al., 2013).

The correlation in average FD between scans from the same participant was r = 0.18 (p = 0.19). While other studies found comparatively large within-subject FD correlation (Zeng et al., 2014), we attribute a smaller correlation to the long period of time between scans in a developing sample, where motion may be less trait-like. We used a linear mixed model to assess the correlation between age and mean FD and found a non-significant negative relationship (p = 0.186; Supplemental Fig. 1).

For analyses in which scans were divided into two groups based on median head motion, the 56 low motion scans had a mean average FD of 0.062 mm (range: 0.035–0.094 mm, std = 0.016 mm). The 56 high motion scans had a mean average FD of 0.191 mm (range: 0.097–0.506 mm, std = 0.088 mm). Mean age at scan was similar across groups (lower-motion: 6.01 years, std = 0.920 years; higher-motion: 5.98 years, std = 0.935 years). For the higher- and lower-motion group ISC analysis, scans were divided into two groups based on median head motion, such that only one scan per individual was included in each group, resulting in a total of 40 scans per group. The lower-motion scans had a mean average FD of 0.059 mm (range: 0.035–0.092 mm, std = 0.016 mm). The higher-motion scans had a mean average FD of 0.202 mm (range: 0.097–0.506 mm, std = 0.096 mm). Mean age at scan remained similar across groups (lower-motion: 6.13 years, std = 0.888 years; higher-motion: 6.03 years, std = 0.934 years).

Table 2
Preprocessing pipelines tested to compare effects of censoring at different thresholds and volumes censored per motion artifact. Note that the pipeline with 1 vol censored at a threshold of 0.25 mm is equivalent to pipeline R4 described in Table 1.

Volumes censored per artifact	Censoring threshold (mm FD)	Mean and std of number of volumes censored
1	0.3	31.8 ± 35.0
	0.25	39.3 ± 41.1
	0.2	51.0 ± 49.8
	0.15	72.2 ± 62.2
3	0.3	57.6 ± 59.4
	0.25	69.7 ± 67.6
	0.2	88.6 ± 77.7
	0.15	121.9 ± 91.0
4	0.3	67.6 ± 68.2
	0.25	81.3 ± 76.6
	0.2	102.5 ± 86.5
	0.15	139.3 ± 98.5

3.2. QC-FC

Fig. 1a shows the percentage of edges significantly correlated with motion, ranging from 19% to 79%. In the minimally processed M1 pipeline, the vast majority of edges were correlated with motion (79%). Regressing only WM and CSF (M2) reduced this to 44%. Both GSR and censoring further reduced the number of edges correlated with motion, and pipelines that paired these two steps had the smallest number of motion-associated edges (e.g., R1: 44% vs. R4: 19%). In general, ICA-AROMA pipelines fared better than comparable pipelines that regressed HMP, though the pipeline with the fewest QC-FC associated edges (R4 – regress HMP+GSR+censoring) did not use ICA-AROMA.

Fig. 1b (Supplemental Table 1) shows the mean absolute correlation with motion for all edges. Here the same trends are observed, with a large motion effect in minimally processed data (M1) greatly reduced with the inclusion of regressing WM and CSF (M1: z = 0.375 vs. M2: z = 0.189, p ≈ 0, d = 0.88). Again, GSR and censoring further reduced the effect of motion when paired (e.g., R1: z = 0.189 vs. R4: z = 0.115). Intermediate improvements, with smaller effect sizes, were seen when using only one of GSR or censoring (Supplemental Table 1).

Fig. 1c shows an example from pipeline R4 of edge correlation with motion vs. edge distance with a linear fit. Fig. 1d (Supplemental Table 2) shows the QC-FC distance dependence for each pipeline. All differences in correlation between pipelines were significant ($p \approx 10^{-49}$ or smaller). Pipelines that include GSR were associated with a more negative correlation (e.g., R1: r = −0.305 vs. R2: r = −0.421), suggesting that GSR affects shorter edges differently than longer edges, with shorter edges having more remaining motion influence and longer edges more likely to be negatively correlated with motion. However, censoring reduced this correlation, partially compensating for the effect of GSR.

When scans were split into lower- and higher- motion subgroups, the higher-motion scans fared worse on all metrics (Fig. 2a vs 2b, 2c vs 2d, 2e vs 2f). Both the lower- and higher-motion groups had fewer edges significantly correlated with motion than the entire group (Fig. 2a and b vs 1a), and differences between pipelines on mean absolute correlation with motion showed smaller effect sizes (Supplemental Tables 3 and 4 vs Supplemental Table 1). We attribute these changes to a smaller sample size and a narrower range of motion values. In the lower- motion group, with the exception of the poor-performing minimal pipelines, choice of pipeline had a minimal effect on the percent of edges correlated with motion (ranging from 6.1% to 8.3% in pipelines R1-R4, I1-I4; Fig. 2a). We saw a similar lack of impact on the mean absolute correlation with motion, ranging from z = 0.11 to z = 0.12 in pipelines R1-R4 and I1-I4 (Fig. 2c), and effect sizes comparing pipelines smaller than 0.15 (Supplemental Table 3). Small improvements were seen with censoring, and pipelines that included GSR fared slightly worse. Most of the effects noted in the entire sample (Fig. 1) were only seen in the higher-motion group (Fig. 2b and d, Supplemental Table 4), where censoring and GSR used together improved the percentage of edges significantly correlated with motion (e.g., R1: 22% vs. R4: 11%) and the mean absolute correlation with motion (e.g., R1: z = 0.17 vs. R4: z = 0.13, though effect

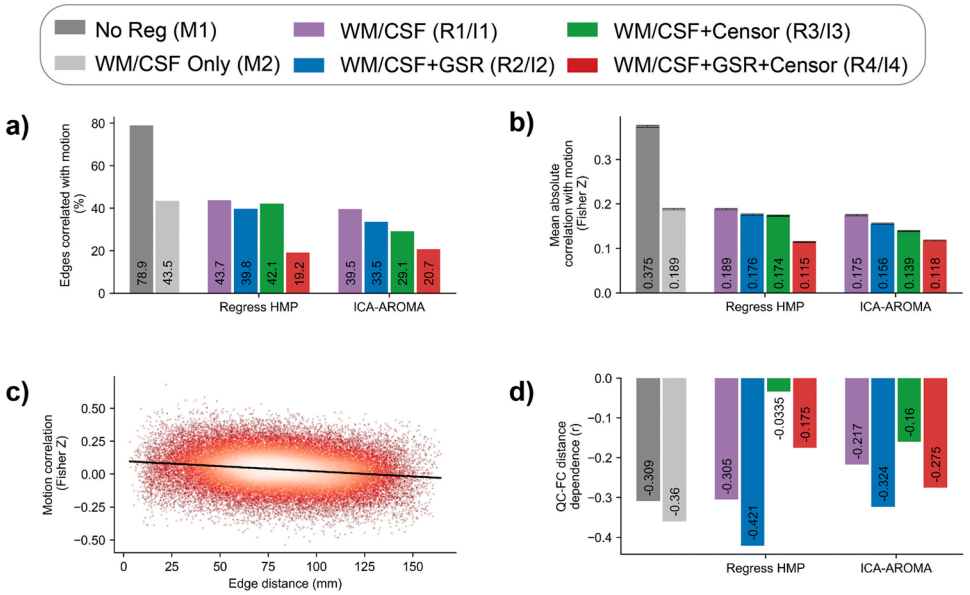


Fig. 1. Quality control-functional connectivity (QC-FC) across pipelines. a) Percentage of edges with a significant correlation between edge value and head motion across all 112 scans (uncorrected $p < 0.05$). b) Mean and 99% confidence interval of the absolute correlation between edge strength and subject motion across all 112 scans. c) Example QC-FC distance dependence plot, from pipeline R4 (regress HMP + WM/CSF + GSR + censor). Each point is an edge in the connectome, plotted based on the length between its nodes (edge distance) and correlation between edge strength and subject motion. Overlapping points are represented by brighter colors. d) QC-FC distance dependence for each pipeline. This is the correlation between edge length and the association between edge strength and subject motion.

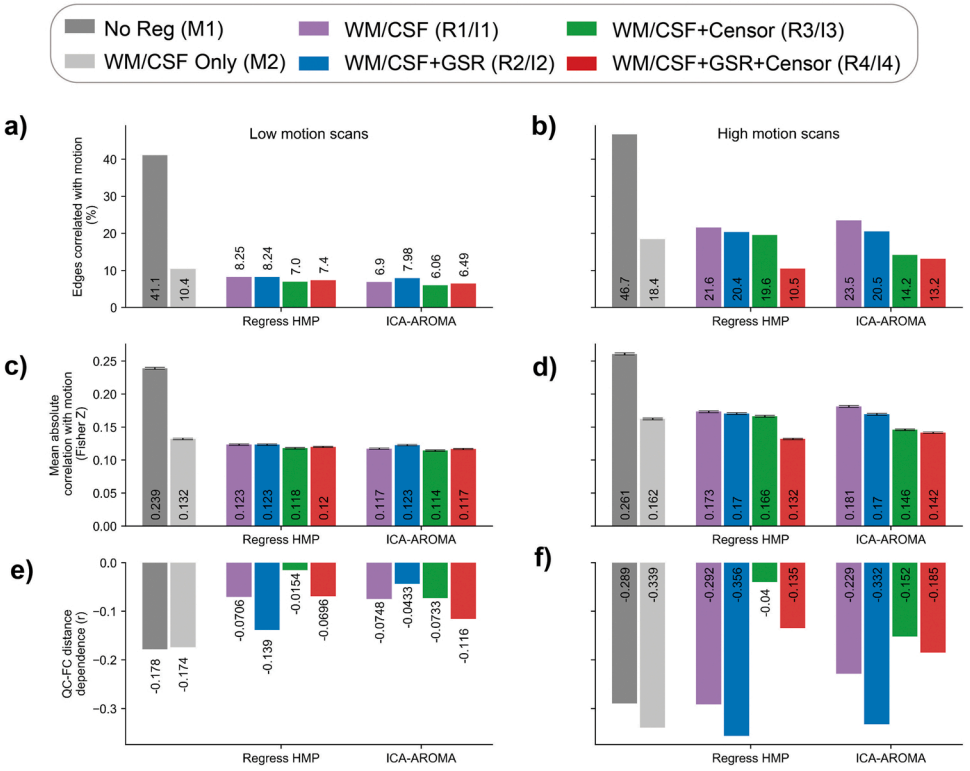


Fig. 2. Quality control-functional connectivity (QC-FC) across pipelines, separately for lower or higher motion scans. Higher and lower motion scans were identified as above or below median average framewise displacement. a, b) Percentage of edges that have a significant correlation between edge strength and head motion (uncorrected $p < 0.05$). c, d) Mean and 99% confidence interval of absolute correlation between edge strength and subject motion. e, f) QC-FC distance dependence. The correlation between edge distance and association between edge strength and head motion.

sizes remained small). For inter-pipeline distance dependence, trends were similar to the entire sample in lower- and higher-motion groups (2e and 2 f vs 1d, Supplemental Tables 5 and 6), with GSR making the association stronger, but censoring mitigating this effect.

3.3. Fingerprinting

Fingerprinting match-rate was assessed for each pipeline (Fig. 3a). Overall, except for the two minimally processed pipelines (M1 and M2), both with under 50% success, the match rate was high for all pipelines tested, ranging from 71% to 90% (chance <1%). GSR had a minimal effect on the overall match rate (e.g., R1: 75.0% vs. R2: 76.8%), but pipelines that included censoring were more successful, especially when regressing HMP rather than using ICA-AROMA (e.g., R3: 88.4% and R4: 90.2%).

While group similarity varied minimally in terms of absolute numbers across pipelines (Fig. 3b), effect sizes comparing pipelines were large, suggesting consistent effects across participants (e.g., R1: $z = 0.565$ vs. R2: $z = 0.572$, $p \approx 10^{-31}$, $d = 1.57$; Supplemental Table 7). The two pipelines with the highest stability were R3 ($z = 0.81$) and R4 ($z = 0.79$; Fig. 3c), though interestingly pipeline I4 (ICA-AROMA+GSR+censoring) had a comparably high stability ($z = 0.79$), but a

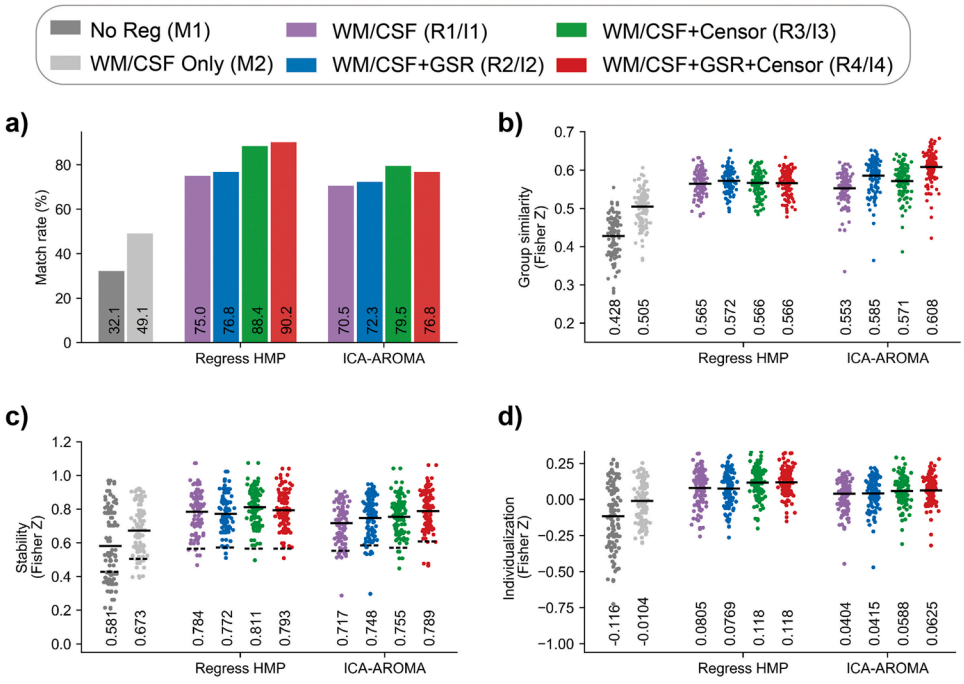


Fig. 3. Functional connectome fingerprinting across pipelines. a) Match rate across pipelines. A scan matched if its highest correlation was to the other scan from the same individual. b) Group similarity across pipelines. Each dot represents an individual scan. Group similarity was assessed as the average correlation to scans from other participants. Lines represent mean values. c) Stability across pipelines. Each dot represents one individual. Stability was assessed as the correlation between scans from the same individual. Solid lines represent mean values, dashed lines represent the average group similarity (from 3b) for comparison. d) Individualization across pipelines. Each dot represents one scan. Individualization was assessed as the difference between stability and the highest correlation to a scan from another participant. Any point below 0 fails to successfully match. Lines represent mean values.

smaller gap between stability and group similarity. Effect sizes between pipelines were large for stability (e.g., R1 vs. R3: $d = 1.41$, $p \approx 10^{-14}$; Supplemental Table 8).

Fig. 3d (Supplemental Table 9) shows individualization (i.e., the fingerprinting margin) for each of the 112 scans. As expected, pipelines with a higher match-rate (Fig. 3a) had a higher average individualization, which ranged from $z = 0.040$ to $z = 0.081$ for less successful pipelines (R1, R2, I1 through I4), while reaching $z = 0.12$ for R3 and R4 ($d > 1.6$ when comparing other pipelines to R3 or R4). There was no significant difference between R3 and R4. The similar distributions and large effect sizes between pipelines suggests that in more successful pipelines, scans that fail to match nonetheless see greater recovery of individual information via a reduced margin of failure.

Fig. 4 (Supplemental Table 10) shows individualization plotted against head motion; a higher r^2 value indicates greater sensitivity to motion noise. Censoring was the most impactful option on reducing this metric (e.g., R1: $r^2 = 0.40$ vs. R3: $r^2 = 0.27$, $p = 6.9 \times 10^{-3}$), suggesting greater impact of censoring on higher motion scans. The participant with the most motion (mean FD of 0.51 mm for one of their two scans) could be considered an outlier – repeating the analysis with that participant removed reduces the r^2 and makes the slopes less negative, but trends between pipelines remain (Supplementary Fig. 2; Supplemental Table 11). We repeated the analysis using stability rather than individualization (Supplementary Fig. 3; Supplemental Table 12). Here we did not find significant differences between pipelines differing in GSR or censoring, but ICA-AROMA pipelines tended to have modestly higher r^2 values on this metric (e.g., R2: $r^2 = 0.36$ vs. I2: r^2 0.41, $p = 0.038$).

3.4. Intra-class correlation (ICC)

For between-session ICC (Fig. 5a; Supplemental Table 13), pipelines that regressed HMP had higher average ICC than corresponding pipelines using ICA-AROMA (e.g., R1: 0.278 vs. I1: 0.228, $p \approx 0$, $d = 0.32$). Across pipelines, the use of GSR lowered ICC coefficients (e.g., R1: 0.278 vs. R2: 0.239, $p \approx 0$, $d = 0.34$). Meanwhile, the use of censoring raised ICC coefficients (e.g., R1: 0.278 vs. R3: 0.303, $p \approx 0$, $d = 0.27$). The two minimal preprocessing pipelines (M1 and M2) had low average ICC coefficients, of 0.163 and 0.216 respectively.

Within-session ICC showed a much smaller range across pipelines, ranging from 0.46 to 0.537 (Fig. 5b; Supplemental Table 14). M1, the pipeline with no regressors, had the third highest mean ICC (0.513). Both GSR and censoring resulted in lower ICC coefficients, with the value dropping from 0.537 for R1 (regress HMP+no GSR+no censor) to 0.479 for R4 (regress HMP+GSR+censor). Pipelines differing in censoring and regressing HMP vs ICA-AROMA had small effect sizes ($p \approx 0$ for all comparisons, $d < 0.27$), while pipelines differing in use of GSR had larger effect sizes ($p < 10^{-25}$, $d > 0.37$).

3.5. Intersubject correlation (ISC)

Fig. 6a (Supplemental Table 15) shows each scan's mean ISC across pipelines, averaged across the 10 nodes with the highest values (See Supplemental Fig. 4 for location of these nodes; Xia et al., 2013). Both GSR and censoring improved ISC, with large effect sizes ($d > 2$). Fig. 6b (Supplemental Table 16) shows each participant's "intrasubject correlation" across their two scans. While intrasubject correlations showed a greater range than mean ISC, this is likely related to being a comparison between only two scans, rather than an average across many scans. Average intrasubject correlations were higher than inter-SC, but trends between pipelines were the same, with GSR and censoring raising the correlation ($p < 10^{-7}$, $d > 0.89$).

When only lower- motion scans were compared using ISC (Fig. 6c; Supplemental Table 17) the benefits of censoring were reduced, with or without GSR. GSR itself increased ISC (e.g., R1: $z = 0.19$ vs. R2: $z = 0.22$, $p \approx 10^{-22}$, $d = 3.14$). However, when only higher- motion scans were compared (Fig. 6d; Supplemental Table 18), the benefits of censoring were more apparent, especially in combination with GSR (e.g., R2: $z = 0.17$ vs. R4: $z = 0.19$, $p \approx 10^{-24}$, $d = 3.77$).

3.6. Intrascan inter-pipeline correlation

With the exception of the two minimally preprocessed pipelines, correlations between connectomes from different pipelines were high overall, ranging from 0.67 to 0.97 (Fig. 7a). A correlation as low as 0.67 for the same scans preprocessed in two different ways – in this case between pipelines R3 (regress HMP+no GSR+censor) and I2 (ICA-AROMA+GSR+no censor) – highlights concerns about replicability

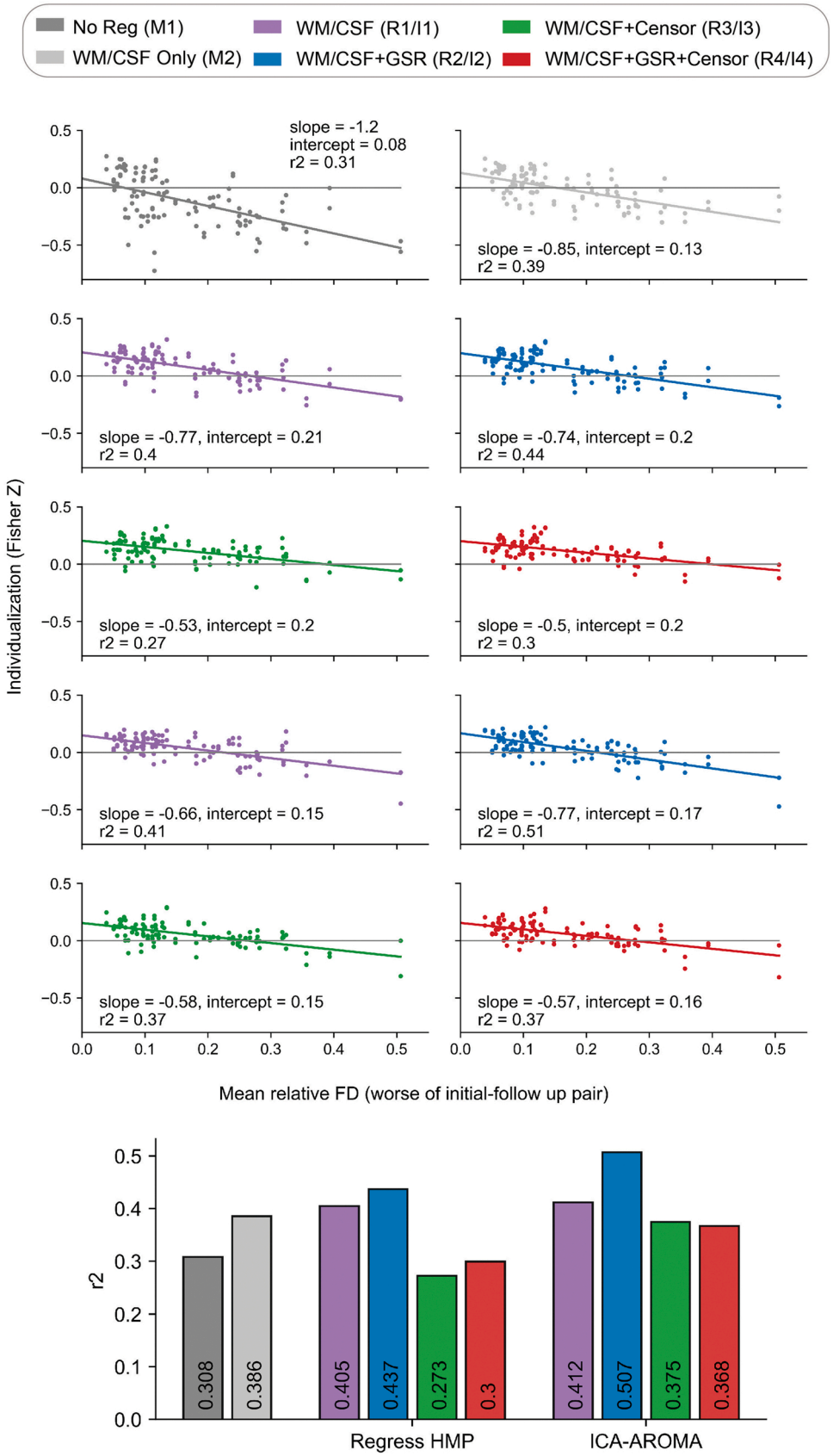


Fig. 4. Individualization as a function of head motion. The difference between each scan's stability (self-correlation) and the highest correlation to a scan from another participant, plotted by motion. Motion was calculated by taking the worse of each participant's two scans' mean relative framewise displacement; each participant has two points for their two scans. Any point below 0 on the y axis fails to successfully match.

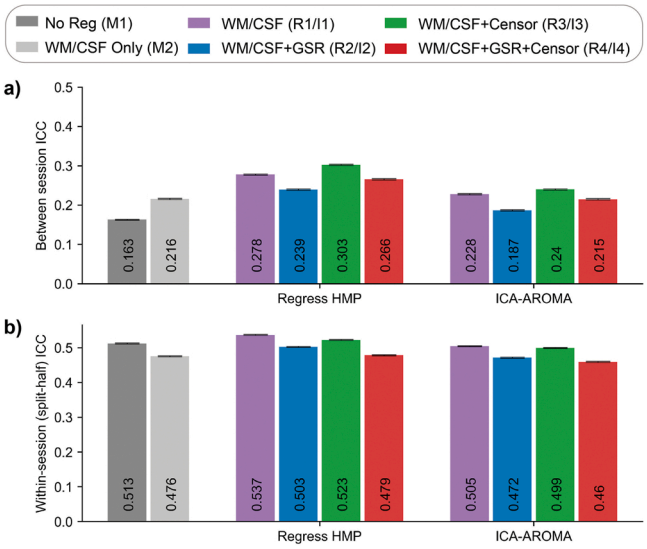


Fig. 5. Intra-class correlation (ICC). a) Mean and 99% confidence interval for between-session ICC across pipelines. b) Mean and 99% confidence interval for the within session (split-half) ICC by pipeline.

given the influence of preprocessing choices. Fig. 7 also highlights how each preprocessing choice progressively alters functional connectivity estimates. Pipelines that differed only in the use of one of GSR or censoring were similar (e.g., R1 vs. R3: r = 0.97) while pipelines that differed in multiple preprocessing steps showed incremental differences (e.g., R1 vs. R4: r = 0.82). The choice to use ICA-AROMA vs. regressing HMP had a relatively larger effect, with correlations ranging from 0.73 to 0.83. Again, these correlations decreased further if the pipelines also differed in use of GSR or censoring. Unsurprisingly, preprocessing choices had a smaller effect on lower-motion scans (Fig. 7b), with correlations between pipelines ranging from 0.72 to 0.99 for pipelines R1-R4, I1-I4, though were as low as 0.61 between M1 (regress nothing) and I2. For higher-motion scans (Fig. 7c), correlations ranged from 0.61 to 0.93 for pipelines R1-R4, I1-I4, dropping as low as 0.48 between M1 and I4.

3.7. Filtering comparison

For highpass relative to bandpass filtering, overall trends between pipelines were quite similar (see Supplemental Figs. 5–7). With a highpass filter, the number of edges significantly correlated with motion ranged from 25% to 50% for pipelines R1-R4, I1-I4 (Supplemental Fig. 5a). Other than for pipeline I1 (which fared poorly on this metric regardless), bandpass filtering resulted in fewer edges significantly correlated with motion than highpass filtering (e.g., R4: 19% for bandpass vs. 26% for highpass). Similarly, pipelines that used bandpass filtering had a smaller mean absolute correlation with motion (e.g., R4: $z = 0.12$ for bandpass vs. $z = 0.13$ for highpass, $p \approx 0$, $d = 0.24$; Supplemental Fig. 5b; Supplemental Table 19). Highpass filtering resulted in more negative motion vs. edge distance correlations than bandpass filtering (e.g., R4: $r = -0.304$ for highpass vs $r = -0.175$ for bandpass, $p \approx 0$; Supplemental Fig. 5c; Supplemental Table 20), suggesting a bandpass filter leads to reduced distance dependent effects.

For fingerprinting, running the analysis with a highpass filter usually led to a lower fingerprinting match rate (Supplemental Fig. 6a), such as 67% vs 77% for pipeline R2 (regress HMP+GSR+no censor). However, pipeline R4 (regress HMP+GSR+censor; 90% match rate) was equally successful using either style of filter. Pipelines that used highpass filtering showed increased stability and group similarity (e.g., for mean stability, R4: $z = 0.87$ for highpass, $z = 0.79$ for bandpass, $p \approx 10^{-128}$, $d = 11.22$; Supplemental Fig. 6b; Supplemental Tables 21 and 22), suggesting a highpass filter makes all scans' FC estimates more similar to each other, relative to bandpass. Filtering choice had a variable effect on individualization (Supplemental Fig. 6c; Supplemental Table 23). For some pipelines, using a highpass filter led to a lower fingerprinting margin (e.g., R2: $z = 0.057$ for highpass, $z = 0.077$ for bandpass, $p \approx 10^{-7}$, $d = 0.051$), but for pipelines R3 and R4 this margin of fingerprinting success improved with a highpass filter (e.g., R4: $z = 0.13$ for highpass, $z = 0.12$ for bandpass, $p \approx 10^{-10}$, $d = 0.27$). Bandpass filtering increased ISC and intrasubject correlations for all pipelines compared to highpass filtering (e.g., for ISC, R4: $z = 0.18$ for highpass, $z = 0.20$ for bandpass, $p \approx 10^{-74}$, $d = 3.74$; Supplemental Fig. 7; Supplemental Tables 24 and 25).

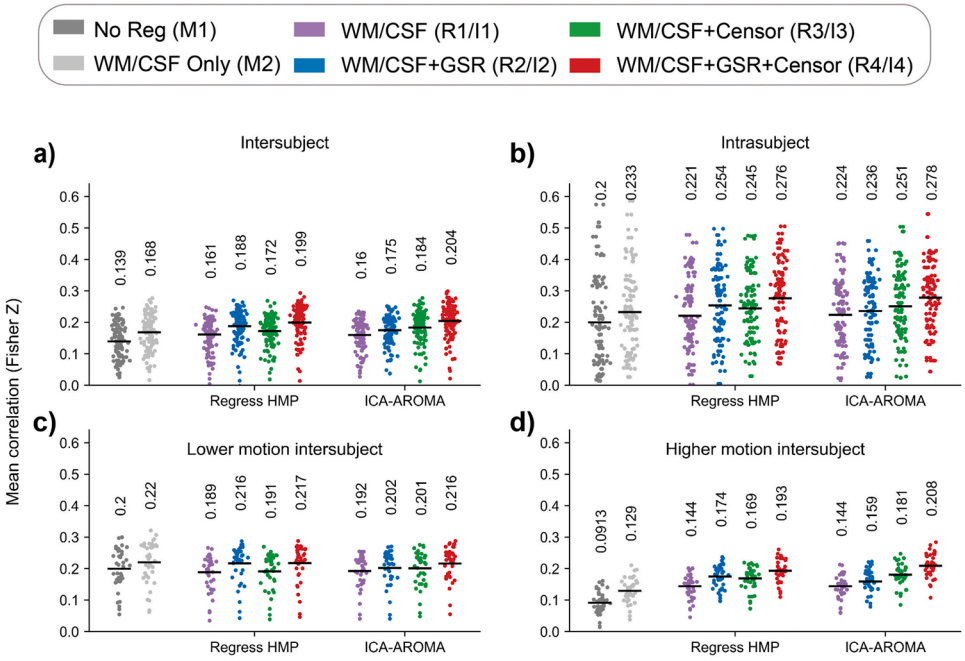


Fig. 6. Mean intersubject correlation (ISC) values. Each point represents the mean ISC for a scan to all other scans, averaged across the 10 nodes with the highest ISCs. Lines represent mean values. a) Mean ISCs for all 112 scans, each compared to all other scans. b) Each participant's intrasubject correlation, the correlation between time series from the two scans' time series. c) Mean ISCs for each of the 40 below-median-motion scans. d) Mean ISCs for each of the 40 above-median-motion scans, compared to the other 39.

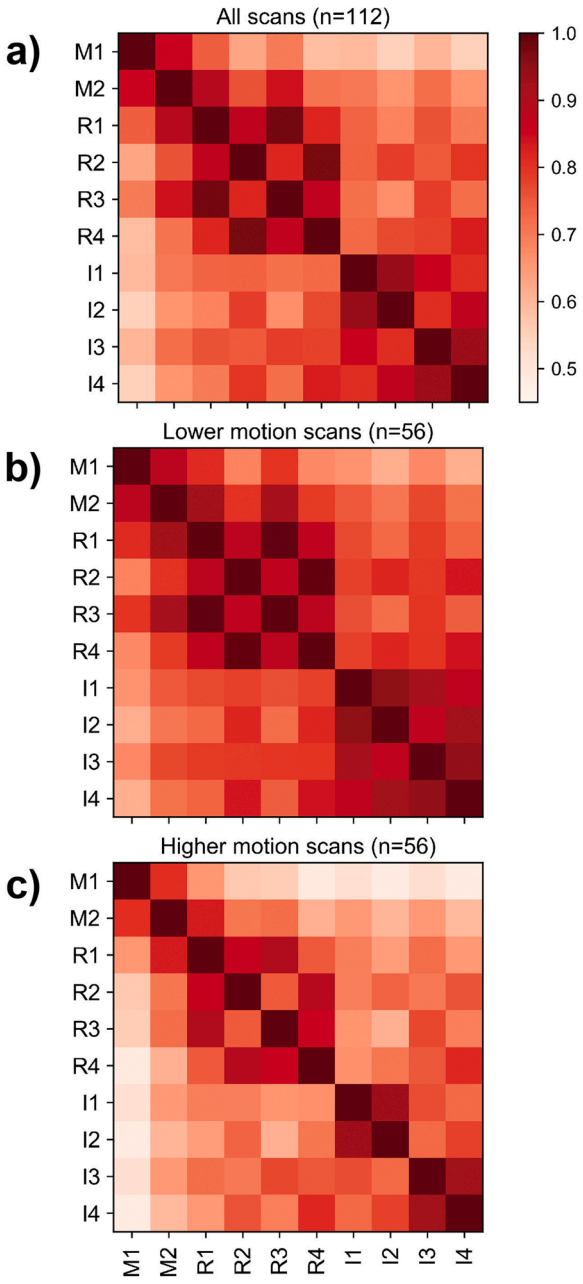


Fig. 7. Intrascan inter-pipeline correlations. For each pair of pipelines, the correlation between connectomes from the same individual across the two pipelines was calculated, Fisher z-transformed, averaged across scans, then converted back to correlations. Pipelines are listed in Table 1, briefly: M1 – no regression; M2 – regress WM/CSF; R1 – regress HMP + WM/CSF; R2 – regress HMP + WM/CSF + GSR; R3 – regress HMP + WM/CSF + censor; R4 – regress HMP + WM/CSF + GSR + censor; I1 – ICA-AROMA + WM/CSF; I2- ICA-AROMA + WM/CSF + GSR; I3 – ICA-AROMA + WM/CSF + censor; I4 – ICA-AROMA + WM/CSF + GSR + censor. a) Average correlation based on all 112 scans. b) Average correlation using the 56 scans below the median average framewise displacement. c) Average correlation using the 56 scans above the median average framewise displacement.

3.8. Censoring comparison

Starting from the pipeline that performed best on most metrics (R4), we repeated our analyses with variable censoring. When one volume was censored per motion artifact, a lower (more stringent) threshold for censoring resulted in fewer edges correlating with motion, lowering from 20% of edges at a threshold of 0.30 mm, to 17% of edges at a threshold of 0.15 mm (Fig. 8a). However, the censoring threshold had a minimal effect on the number of edges significantly correlated with motion when 3 or 4 volumes were censored per artifact, ranging from 13% to 15%. Effects were similar in the mean absolute correlation with motion (Fig. 8b; Supplemental Table 26), with a lower threshold at one volume per artifact reducing the correlation from $z = 0.116$ at 0.30 mm to $z = 0.108$ at 0.15 mm, but having minimal effect at 3 or 4 volumes (ranging from $z = 0.099$–0.104 mm). However, effect sizes between pipelines were comparatively small ($d < 0.37$) for this analysis.

The fingerprinting match-rate decreased as censoring became more stringent (Fig. 8c), suggesting that although censoring can help with identifiability, there is an optimal range, above which individual information is reduced. At 1 vol censored per motion artifact, the decrease was from 90% to 88% across thresholds (0.3–0.15 mm), but at 4 volumes censored per artifact the match-rate fell from 88% to 80%. With stricter censoring, both stability and group similarity decreased (Fig. 8d; Supplemental Tables 27 and 28), for example at 3 volumes per motion artifact the mean stability dropped from $z = 0.778$ to $z = 0.720$ across thresholds ($p \approx 10^{-20}$, $d = 1.40$). Mean individualization was less affected by censoring (Fig. 8e; Supplemental Table 29), though effect sizes remained large for most comparisons between pipelines. Censoring had a minimal effect on mean ISC (Fig. 8f; Supplemental Table 30), except for the strictest censoring. With ISC, effect sizes remained large for most comparisons between pipelines.

4. Discussion

This study extended preprocessing benchmark comparisons, which have previously been performed in youth and adult resting data, to an early childhood passive viewing sample. Our results support and extend previous findings, suggesting that the highest performing pipelines include both GSR and censoring and that these steps have greater impact in higher-motion data, but remained beneficial in lower-motion data. Further extending previous work, we used connectome fingerprinting to estimate how individual-specific information was retained, or enhanced, by preprocessing steps. In fingerprinting, we found that volume censoring conferred more benefit than GSR, and pipelines with HMP regression outperformed pipelines with ICA-AROMA. When examining ISC, GSR offered the greatest improvement, with censoring showing a specific benefit for high motion scans. Censoring multiple volumes per motion artifact improved QC-FC metrics but had minimal or even a negative effect on other metrics, suggesting an important tradeoff between removal of noise and signal of interest. When examining overall effects of preprocessing choices on connectomes, we found that connectomes differed substantially depending on preprocessing choices, particularly in our higher motion subgroup, which is a concern for cross-study replicability in early childhood fMRI. Overall, our work suggests that for relatively high motion data from early childhood the best preprocessing pipeline includes a bandpass filter, GSR, regressing HMP rather than using ICA-AROMA, and moderate censoring.

When comparing ICA-AROMA pipelines to pipelines that regressed HMPs, we found ICA-AROMA to be less effective on most metrics. This is somewhat surprising, given that ICA-AROMA primarily works by removing independent components strongly associated with motion estimates (Pruim et al., 2015). Previous benchmarking studies found that ICA-AROMA pipelines were among the most effective of pipelines that did not include censoring in terms of reducing the quantity of edges correlated with motion – more effective than regressing HMP – though the difference was small between ICA-AROMA+GSR and HMP+GSR (Ciric et al., 2017; Parkes et al., 2018). While our QC-FC results are consistent with previous findings, we found that ICA-AROMA performed relatively poorly on individualization metrics. Even when ICA-AROMA fared comparably well on stability (as in pipeline I4), a necessary component for individualization, it had an elevated similarity to others, decreasing individual identifiability relative to pipelines that regressed HMP. However, ICA-AROMA pipelines showed no obvious difference

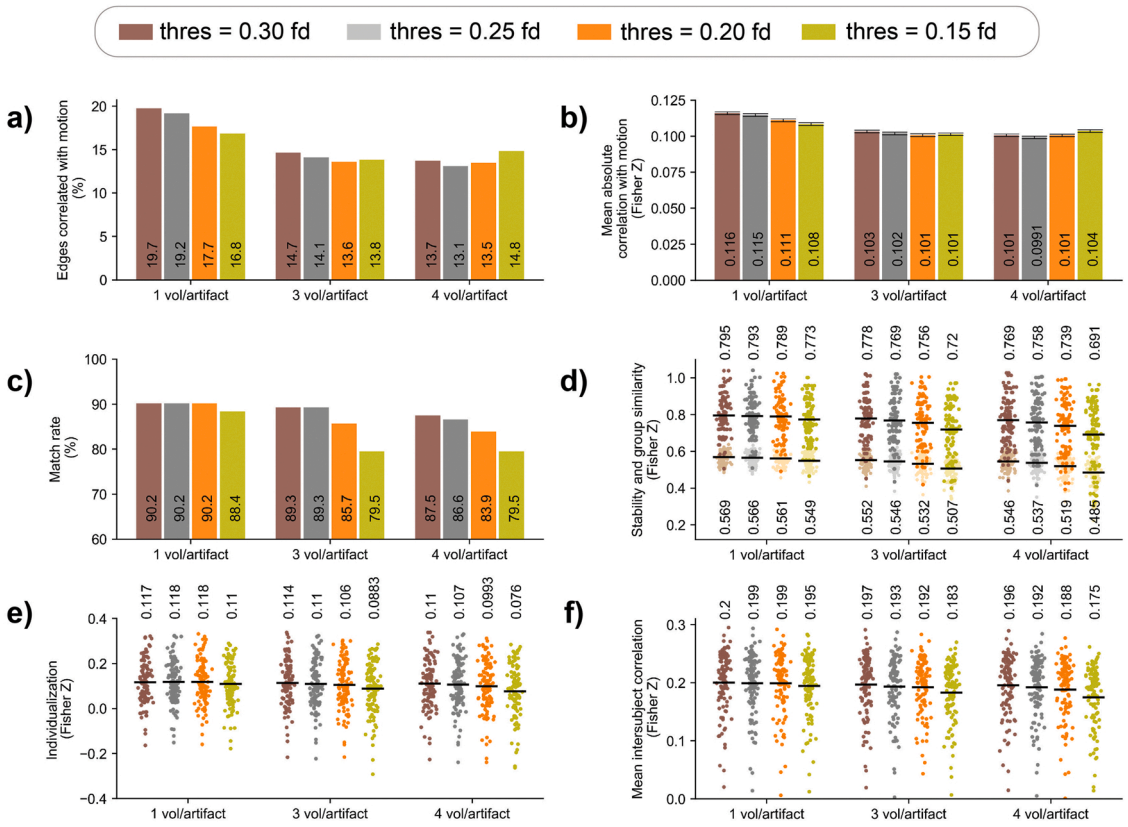


Fig. 8. Pipeline benchmarks with different censoring thresholds. Colored bars depict four threshold levels and across the x-axis a different number of volumes is censored around the high motion frame. a) Percentage of edges with a significant correlation between edge strength and head motion across all 112 scans (uncorrected $p < 0.05$). b) Mean and 99% confidence interval for absolute correlation between edge strength and motion across all 112 scans. c) Fingerprinting match rate across pipelines. d) Each scan's stability (darker points) and group similarity (lighter points), for each pipeline. Group similarity was assessed as the average correlation to scans from other participants; stability was assessed as the correlation between scans from the same individual. Lines represent mean values. e) Individualization across pipelines. Each dot represents one scan. Individualization was assessed as the difference between stability and the highest correlation to a scan from another participant. Lines represent mean values. f) Mean ISCs for all 112 scans.

from regressing HMP on ISC metrics, which measure the evoked response common across participants. One interpretation could be that ICA-AROMA removes both noise and individual-specific information although retaining signal common across participants. Further work is needed to clarify whether ICA-AROMA is less effective in samples of young children or high-motion samples, or whether connectome identifiability is generally poorer in ICA-AROMA relative to HMP pipelines. Other implementations of ICA-based preprocessing exist and are worth systematically investigating. ICA-AROMA is often used in conjunction with ICA-FIX, another ICA-based denoising strategy (Salimi-Khorshidi et al., 2014), and ICA-AROMA can also be run without their aggressive denoising option. Some studies have combined ICA-based denoising with regressing HMPs (Kaufmann et al., 2017; Jalbrzikowski et al., 2020), which may have advantages, though effectively removes aspects of the signal related to motion twice, contrary to the original intent of ICA-AROMA (Pruim et al., 2015).

While GSR remains contentious in the field, our results suggest there are advantages to using GSR in a high-motion dataset such as the early childhood sample used here. GSR reduced the number of edges significantly correlated with motion, especially when applied with censoring. Similarly, while there are distance-dependent effects when GSR is used, this effect was largely mitigated in pipelines that also included censoring, consistent with QC-FC results from previous studies. Ciric et al. (2017) found that their 6 best pipelines, based on fewest edges correlated with motion, all included GSR, while Parkes et al. (2018) found that all 6 pipeline-pairs that differed in the use of GSR benefited from its inclusion. Both studies found that GSR introduced distance-dependent effects but that censoring diminished the impact. Ciric et al. (2017) suggests that GSR introduces these effects by more effectively denoising long range connections; we speculate that censoring mitigates this by removing the worst motion damage that may be missed by GSR. While these aforementioned studies both considered resting state data from adults, a benchmarking paper on fetal fMRI (Taymourtash et al., 2020) found similar benefits to GSR, suggesting that GSR has benefits regardless of age range or scan protocol.

Notably, however, GSR had no obvious impact on connectome individualization metrics. This suggests that while GSR alters FC estimates, most notably via distance-dependent effects, these changes may have a more global and less regionally specific impact compared to censoring (Power et al., 2015). The finding that GSR had a minimal impact on fingerprinting, our metric of individual information, may alleviate concerns about its use when comparing healthy controls and clinical populations, or in developmental research, even in light of research that suggests the global signal is significantly related to life outcomes and psychological function (Li et al., 2019). We also found benefits to the use of GSR in improving signal-to-noise in ISC metrics, suggesting that the benefits in terms of motion and physiological noise removal in an early childhood sample may outweigh the cost of removing some signal of interest (Behzadi et al., 2007).

We found relatively small effects of a bandpass filter compared to a highpass filter, with the main advantage of bandpass filtering on QC-FC metrics. In connectome fingerprinting, while stability was higher in pipelines that used a highpass filter, so too was group similarity by comparable levels. This suggests that any resulting changes to the functional connectome by including higher frequencies are consistent across individuals and are unlikely to reflect individual-specific

information. Bandpass filtering also increased ISC metrics; this may suggest that bandpass filtering is effective for removing noise while retaining task signal, but this improvement may also be explained in part by signals composed of a narrower range of frequencies being inherently more similar across individuals.

We found benefits to volume censoring across metrics. On QC-FC benchmarks, we found that censoring improved both the quantity of edges associated with motion and distance-dependent effects. These effects are in line with Ciric et al. (2017) and Parkes et al. (2018) who both found that censoring pipelines outperformed alternatives on edges associated with motion, with both studies finding that the most effective pipelines included censoring. Our work extends these findings to early childhood passive viewing fMRI as well as to other benchmarks. In connectome fingerprinting, censoring increased accuracy more than any of the other preprocessing steps compared here, with censoring also conferring additional benefits to pipelines that used ICA-AROMA, despite ICA-AROMA being intended partially as an alternative to censoring (Pruim et al., 2015). Censoring had only a small effect on lower-motion scans for ISC comparisons, but censoring increased ISC between higher-motion scans.

When only one volume was censored per motion artifact, the more stringent the censoring threshold, the fewer edges significantly correlated with motion, dropping from 20% to 17% of edges. While this pattern was less appreciable when censoring multiple volumes per motion artifact, even at the most lenient threshold tested multivolume censoring outperformed single volume censoring on this metric. Stricter censoring tended to decrease both stability and group similarity. This suggests that removing data points removes information that is common across participants. Interestingly, the distribution of individualization values was relatively unaffected by the censoring threshold, except when very strict censoring was used, even though the overall match rate decreased when censoring became stricter. This suggests not all scans are affected in the same way by changes in censoring parameters. While censoring had a minimal effect on mean ISC, except for the strictest censoring, it should be noted that for the purposes of ISC, censoring one scan necessarily censors the scan it is being compared to, since the same time points need to be compared. A stricter censoring protocol will lead to a disproportionate reduction in shared time points, so benefits in the removal of noise may be balanced by the decrease in data being compared. Altogether, while we found censoring to be beneficial, and can advise against overly strict censoring, the exact implementation of censoring seems to involve tradeoffs, making it challenging to advocate a one-size-fits-all approach. Future work may further consider how censoring is implemented, with the possibility of utilizing a scan-specific approach.

Previous work has found that ICC values are generally higher in more minimal pipelines that perform worse on other benchmarks (Parkes et al., 2018; Kassinopoulos and Mitsis, 2021), suggesting a trade-off between reliability and validity (Noble et al., 2021) due to correlated noise across sessions. Here, we show that within-session ICC scores are consistent with previous work, but between-session ICC values (with ~12 months between sessions) were generally higher in pipelines with more aggressive denoising. Our interpretation of differences across studies is that greater time between scans in a developmental sample reduces the shared influence of motion and physiological artifacts. Indeed, in our sample the correlation between initial and follow-up scan head motion was non-significant. This interpretation is supported by analyses in Parkes et al. (2018), who similarly found that while poorer performing pipelines had higher ICC, this effect was reduced when scans were collected on average 90 days apart, rather than within the same session. Given the conceptual difficulty of separating reproducible head motion from its biological confounds (Engelhardt et al., 2017; Hodgson et al., 2016), our data presents an advantage. We further note that ICC or other measures of test-retest reliability may be imperfect benchmarks for addressing head motion, and a broader collection of benchmarks, such as those included here, may be more appropriate to best understand how preprocessing choices affect signal changes (Kassinopoulos and Mitsis, 2021).

Our findings on the intrascan interpipeline correlations are concerning towards the goal of replicable results in FC-MRI studies. While preprocessing choices have a smaller impact on lower-motion data, our work suggests that when analyzing high motion data, such as data from young children, each preprocessing choice has a measurable and cumulative effect on FC estimates (Li et al., 2021). While broader methodological differences or group-level participant differences are often cited as reasons for different findings among studies, we suggest that differences in preprocessing are likely to also play a role. Preprocessing choices should be both closely considered and accurately reported. While we compared the correlation between the entire functional connectome, future work should explore regionally-specific effects in more detail.

There are several limitations to our work. Our study was done in young children who were participating in a passive viewing task, and our results may not be directly applicable to other populations or other study protocols. Likewise, the benefits of any step we tested may not be transferable to a more traditional task-based fMRI study, especially censoring data points during the task. We note that several of the steps implemented here could have been applied in alternative forms, which also restricts generalizability. For example, we used a 24 HMP model rather than 6 or 12 HMP (Friston et al., 1996). Similarly, there are many other thresholds that can be used for temporal filtering or censoring, or conceptually similar strategies such as despiking (Patel et al., 2014). There are also other possible approaches to noise mitigation, such as CompCor (Behzadi et al., 2007) or ANATICOR (Jo et al., 2010), that warrant further investigation.

We also acknowledge that choices such as registration and parcellation may have impacted our findings. Previous studies using functional connectome fingerprinting have used a similar number or fewer nodes; for example, Finn et al. (2015) used a 268-node atlas while Miranda-Dominguez et al. (2018) used a 333-node atlas. Finn et al. (2015) also found lower fingerprinting success using a 68-node atlas; it is unknown if a higher number of nodes may offset the difference between preprocessing strategies on fingerprinting metrics, especially if considering a specific subset of edges rather than the full connectome (Byrge and Kennedy, 2019). While we chose metrics that we believe are meaningful to gauge the effect of preprocessing, other metrics could have instead been chosen. For example, we did not consider the effect preprocessing has on the magnitude of FC edges, and we also did not consider the effect preprocessing might have on individual networks within the broader connectome (Kassinopoulos and Mitsis, 2021), or other changes in functional organization.

This study implemented ISC by comparing nodes across scans, rather than a voxel-based approach (Hasson et al., 2004). This has the potential to average out the effect of the task, as not all voxels within a parcel will respond to the passive viewing task in a similar way. We also chose to focus on the nodes with the highest average ISC values. These choices limit the extent our findings can be generalized across the whole brain or to voxel-specific changes. Our analysis has assumed that a higher ISC value reflects greater recovery of true signal, based on the assumption that unremoved noise will lower the temporal correlation between scans. Parkes et al. (2018) suggest that reproducible, individual-specific noise can increase FC test-retest reliability. It is unknown whether noise has a similar effect between subjects in the context of ISC, though there is the potential for head movements to be more likely at specific times in a video (e.g., due to laughter). Future work is needed that specifically investigates the effect of preprocessing choices on ISC values across the brain, especially in high noise scans.

5. Conclusions

Due to the impact that both head motion and preprocessing choices have on FC estimates, benchmarking preprocessing steps in high-motion

early childhood samples is critical to support researchers in making informed choices. While in different datasets the optimal preprocessing choices may vary, our results suggest that GSR, censoring, a bandpass filter, and HMP regression are preferable in high motion datasets from early childhood populations engaging in a passive viewing task. In particular, GSR and censoring showed few disadvantages across our metrics, and ICA-AROMA showed no major improvement compared to regressing out HMPs, especially when used without censoring.

All preprocessing choices have unintended effects on data; in light of the major effect that preprocessing choices has on FC estimates, we urge awareness of the effect of preprocessing choices within any given research protocol. Ideally, studies should aim to reduce head motion at the time of scan, for instance through the use of passive viewing, and new preprocessing strategies should aim to improve the signal-to-noise ratio with fewer tradeoffs.

Software versions used

Nipype version 1.1.5 (Gorgolewski et al., 2011).
FSL version 6.0.0 (Smith et al., 2004).
ANTs version 3.0.0.0 (Avants et al., 2011).
AFNI version 18.3.03 (Cox, 1996).

Funding

This work was supported by: an Alberta Graduate Excellence Scholarship, and a Natural Sciences and Engineering Research Council of Canada (NSERC)-CREATE Training Scholarship awarded to KG; and an NSERC Discovery Award, Canadian Institutes of Health Research (CIHR) – Institute of Neurosciences, Mental Health and Addiction (INMHA) Bridge Award and CIHR Project Grant to SB.

CRediT authorship contribution statement

Kirk Graff: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Ryann Tansey:** Investigation, Writing – review & editing. **Amanda Ip:** Resources, Investigation. **Christiane Rohr:** Investigation, Writing – review & editing. **Dennis Dimond:** Investigation, Writing – review & editing. **Deborah Dewey:** Methodology, Writing – review & editing. **Signe Bray:** Conceptualization, Investigation, Methodology, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The authors do not have permission to share data. Raw data used in this study are not publicly available, in agreement with the terms of consent given by participants and the Conjoint Health and Research Ethics Board at the University of Calgary. Benchmarking results (e.g., QC-FC, fingerprinting, ISC values) for each pipeline, or other pertinent data, will be made available upon request. The software packages utilized (i.e., Nipype, FSL, AFNI, ANTs) are available on their respective websites. Python scripts are available at https://github.com/BrayNeuroimagingLab/BNL_open/tree/main/fMRI_preprocessing.

Acknowledgments

We thank all the children and their families who participated in this study, and the staff at the Alberta Children's Hospital Imaging Centre. Thanks to Taylor Johnson for assistance with the artwork.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.dcn.2022.101087.

References

Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998. The inferential impact of global signal covariates in functional neuroimaging analyses. Neuroimage 8, 302–306. https://doi.org/10.1006/nimg.1998.0367.

Alexander, L.M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., Vega-Potler, N., Langer, N., Alexander, A., Kovacs, M., Litke, S., O'Hagan, B., Andersen, J., Bronstein, B., Bui, A., Bushey, M., Butler, H., Castagna, V., Camacho, N., Chan, E., Citera, D., Clucas, J., Cohen, S., Dufek, S., Eaves, M., Fradera, B., Gardner, J., Grant-Villegas, N., Green, G., Gregory, C., Hart, E., Harris, S., Horton, M., Kahn, D., Kabotyanski, K., Karmel, B., Kelly, S.P., Kleinman, K., Koo, B., Kramer, E., Lennon, E., Lord, C., Mantello, G., Margolis, A., Merikangas, K.R., Milham, J., Minniti, G., Neuhaus, R., Levine, A., Osman, Y., Parra, L.C., Pugh, K.R., Racanello, A., Restrepo, A., Saltzman, T., Septimus, B., Tobe, R., Waltz, R., Williams, A., Yeo, A., Castellanos, F.X., Klein, A., Paus, T., Leventhal, B.L., Craddock, R.C., Koplewicz, H.S., Milham, M.P., 2017. Data descriptor: an open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci. Data 4, 1–26. https://doi.org/10.1038/sdata.2017.181.

Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54, 2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025.

Ball, G., Aljabar, P., Zebari, S., Tusor, N., Arichi, T., Merchant, N., Robinson, E.C., Ogundipe, E., Rueckert, D., Edwards, A.D., Counsell, S.J., 2014. Rich-club organization of the newborn human brain. Proc. Natl. Acad. Sci. U.S.A. 111, 7456–7461. https://doi.org/10.1073/pnas.1324118111.

Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37, 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042.

Bijsterbosch, J.D., Valk, S.L., Wang, D., Glasser, M.F., 2021. Recent developments in representations of the connectome. Neuroimage 243, 118533. https://doi.org/10.1016/j.neuroimage.2021.118533.

Bray, S., Arnold, A.E.G.F., Levy, R.M., Iaria, G., 2015. Spatial and temporal functional connectivity changes between resting and attentive states. Hum. Brain Mapp. 36 (2), 549–565. https://doi.org/10.1002/hbm.22646.

Byrge, L., Kennedy, D.P., 2019. High-accuracy individual identification using a "thin slice" of the functional connectome. Netw. Neurosci. 3, 363. https://doi.org/10.1162/NETN_A_00068.

Carp, J., 2012. The secret lives of experiments: methods reporting in the fMRI literature. Neuroimage 63, 289–300. https://doi.org/10.1016/j.neuroimage.2012.07.004.

Chai, X.J., Castañán, A.N., Öngür, D., Whitfield-Gabrieli, S., 2012. Anticorrelations in resting state networks without global signal regression. Neuroimage 59, 1420–1428. https://doi.org/10.1016/j.neuroimage.2011.08.048.

Churchill, N.W., Raamana, P., Spring, R., Strother, S.C., 2017. Optimizing fMRI preprocessing pipelines for block-design tasks as a function of age. Neuroimage 154, 240–254. https://doi.org/10.1016/j.neuroimage.2017.02.028.

Ciric, R., Wolf, D.H., Power, J.D., Roalf, D.R., Baum, G.L., Ruparel, K., Shinohara, R.T., Elliott, M.A., Eickhoff, S.B., Davatzikos, C., Gur, R.C., Gur, R.E., Bassett, D.S., Satterthwaite, T.D., 2017. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. Neuroimage 154, 174–187. https://doi.org/10.1016/j.neuroimage.2017.03.020.

Ciric, R., Rosen, A.F.G., Erus, G., Cieslak, M., Adebimpe, A., Cook, P.A., Bassett, D.S., Davatzikos, C., Wolf, D.H., Satterthwaite, T.D., 2018. Mitigating head motion artifact in functional connectivity MRI. Nat. Protoc. 13, 2801–2826. https://doi.org/10.1038/s41596-018-0065-y.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173. https://doi.org/10.1006/cbmr.1996.0014.

Damoiseaux, J.S., Rombouts, S.A.R.B., Barkhof, F., Scheltens, P., Stam, C.J., Smith, S.M., Beckmann, C.F., 2006. Consistent resting-state networks across healthy subjects. Proc. Natl. Acad. Sci. U.S.A. 103, 13848–13853. https://doi.org/10.1073/pnas.0601417103.

Davey, C.E., Grayden, D.B., Egan, G.F., Johnston, L.A., 2013. Filtering induces correlation in fMRI resting state data. Neuroimage 64, 728–740. https://doi.org/10.1016/j.neuroimage.2012.08.022.

Dimond, D., Rohr, C.S., Smith, R.E., Dhollander, T., Cho, I., Lebel, C., Dewey, D., Connelly, A., Bray, S., 2020a. Early childhood development of white matter fiber density and morphology. Neuroimage 210. https://doi.org/10.1016/j.neuroimage.2020.116552.

Dimond, D., Heo, S., Ip, A., Rohr, C.S., Tansey, R., Graff, K., Dhollander, T., Smith, R.E., Lebel, C., Dewey, D., Connelly, A., Bray, S., 2020b. Maturation and interhemispheric asymmetry in neurite density and orientation dispersion in early childhood. Neuroimage 221. https://doi.org/10.1016/j.neuroimage.2020.117168.

Dosenbach, N.U.F., Koller, J.M., Earl, E.A., Miranda-Dominguez, O., Klein, R.L., Van, A. N., Snyder, A.Z., Nagel, B.J., Nigg, J.T., Nguyen, A.L., Wesevich, V., Greene, D.J., Fair, D.A., 2017. Real-time motion analytics during brain MRI improve data quality

and reduce costs. Neuroimage 161, 80–93. https://doi.org/10.1016/j.neuroimage.2017.08.025.
Engelhardt, L.E., Roe, M.A., Juranek, J., DeMaster, D., Harden, K.P., Tucker-Drob, E.M., Church, J.A., 2017. Children's head motion during fMRI tasks is heritable and stable over time. Dev. Cogn. Neurosci. 25, 58–68. https://doi.org/10.1016/j.dcn.2017.01.011.
Fair, D.A., Nigg, J.T., Iyer, S., Bathula, D., Mills, K.L., Dosenbach, N.U.F., Schlaggar, B.L., Mennes, M., Gutman, D., Bangaru, S., Buitelaar, J.K., Dickstein, D.P., Martino, A., Di, Kennedy, D.N., Kelly, C., Luna, B., Schweitzer, J.B., Velanova, K., Wang, Y.F., Mostofsky, S., Castellanos, F.X., Milham, M.P., 2013. Distinct neural signatures detected for ADHD subtypes after controlling for micro-movements in resting state functional connectivity MRI data. Front. Syst. Neurosci. 6, 1–31. https://doi.org/10.3389/fnsys.2012.00080.
Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat. Neurosci. 18, 1664–1671. https://doi.org/10.1038/nn.4135.
Fleming, S., Thompson, M., Stevens, R., Heneghan, C., Plüddemann, A., Maconochie, I., Tarassenko, L., Mant, D., 2011. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. Lancet 377, 1011–1018. https://doi.org/10.1016/S0140-6736(10)62226-X.
Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E., 2005. From the cover: the human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc. Natl. Acad. Sci. 102, 9673–9678. https://doi.org/10.1073/pnas.0504136102.
Fox, M.D., Zhang, D., Snyder, A.Z., Raichle, M.E., 2009. The global signal and observed anticorrelated resting state brain networks. J. Neurophysiol. 101, 3270–3283. https://doi.org/10.1152/jn.90777.2008.
Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S., Turner, R., 1996. Movement-related effects in fMRI time-series. Magn. Reson. Med. 35, 346–355.
Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., Ghosh, S.S., 2011. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front. Neuroinform. 5, 13. https://doi.org/10.3389/fninf.2011.00013.
Gotts, S.J., Saad, Z.S., Jo, H.J., Wallace, G.L., Cox, R.W., Martin, A., 2013. The perils of global signal regression for group comparisons: a case study of autism spectrum disorders. Front. Hum. Neurosci. 7. https://doi.org/10.3389/fnhum.2013.00356.
Grayson, D.S., Fair, D.A., 2017. Development of large-scale functional networks from birth to adulthood: a guide to the neuroimaging literature. Neuroimage 160, 15–31. https://doi.org/10.1016/j.neuroimage.2017.01.079.
Greene, D.J., Koller, J.M., Hampton, J.M., Wesevich, V., Van, A.N., Nguyen, A.L., Hoyt, C.R., McIntyre, L., Earl, E.A., Klein, R.L., Shimony, J.S., Petersen, S.E., Schlaggar, B.L., Fair, D.A., Dosenbach, N.U.F., 2018. Behavioral interventions for reducing head motion during MRI scans in children. Neuroimage 171, 234–245. https://doi.org/10.1016/J.NEUROIMAGE.2018.01.023.
Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303, 1634–1640. https://doi.org/10.1126/science.1089506, 80.
Hodgson, K., Poldrack, R.A., Curran, J.E., Knowles, E.E., Mathias, S., Göring, H.H., Yao, N., Olvera, R.L., Fox, P.T., Almasy, L., Duggirala, R., Barch, D.M., Blangero, J., Glahn, D.C., 2016. Shared genetic factors influence head motion during MRI and body mass index. Cereb. Cortex 27, 5539–5546. https://doi.org/10.1093/cercor/bhw321.
Huang, C.-M., Lee, S.-H., Hsiao, I.-T., Kuan, W.-C., Wai, Y.-Y., Ko, H.-J., Wan, Y.-L., Hsu, Y.-Y., Liu, H.-L., 2010. Study-specific EPI template improves group analysis in functional MRI of young and older adults. J. Neurosci. Methods 189, 257–266. https://doi.org/10.1016/j.jneumeth.2010.03.021.
Jalbrzikowski, M., Liu, F., Foran, W., Klei, L., Calabro, F.J., Roeder, K., Devlin, B., Luna, B., 2020. Functional connectome fingerprinting accuracy in youths and adults is similar when examined on the same day and 1.5–years apart. Hum. Brain Mapp. 41, 4187–4199. https://doi.org/10.1002/hbm.25118.
Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841.
Jo, H.J., Saad, Z.S., Simmons, W.K., Milbury, L.A., Cox, R.W., 2010. Mapping sources of correlation in resting state FMRI, with artifact detection and removal. Neuroimage 52, 571–582. https://doi.org/10.1016/j.neuroimage.2010.04.246.
Kassinopoulos, M., Mitsis, G.D., 2021. A multi-measure approach for assessing the performance of fMRI preprocessing strategies in resting-state functional connectivity. BioRxiv. https://doi.org/10.1101/837609.
Kaufmann, T., Alnæs, D., Doan, N.T., Brandt, C.L., Andreassen, O.A., Westlye, L.T., 2017. Delayed stabilization and individualization in connectome development are related to psychiatric disorders. Nat. Neurosci. 20, 513–515. https://doi.org/10.1038/nn.4511.
Kauppi, J.P., Jääskeläinen, I.P., Sams, M., Tohka, J., 2010. Inter-subject correlation of brain hemodynamic responses during watching a movie: localization in space and frequency. Front. Neuroinform. 4. https://doi.org/10.3389/fninf.2010.00005.
Li, J., Bolt, T., Bzdok, D., Nomi, J.S., Yeo, B.T.T., Spreng, R.N., Uddin, L.Q., 2019. Topography and behavioral relevance of the global signal in the human brain. Sci. Rep. 9, 14286. https://doi.org/10.1038/s41598-019-50750-8.
Li, X., Ai, L., Giavasis, S., Jin, H., Feczko, E., Xu, T., Clucas, J., Franco, A., Heinsfeld, A.S., Adebimpe, A., Vogelstein, J.T., Yan, C., Esteban, O., Poldrack, R.A., Craddock, C., Fair, D., Satterthwaite, T., Kiar, G., Milham, M.P., 2021. Moving beyond processing and analysis-related variation in neuroscience. BioRxiv. https://doi.org/10.1101/2021.12.01.470790.
Lindquist, M.A., Geuter, S., Wager, T.D., Caffo, B.S., 2019. Modular preprocessing pipelines can reintroduce artifacts into fMRI data. Hum. Brain Mapp. 40, 2358–2376. https://doi.org/10.1002/hbm.24528.
Marek, S., Hwang, K., Foran, W., Hallquist, M.N., Luna, B., 2015. The contribution of network organization and integration to the development of cognitive control. PLoS Biol. 13, e1002328 https://doi.org/10.1371/journal.pbio.1002328.
Miranda-Dominguez, O., Feczko, E., Grayson, D.S., Walum, H., Nigg, J.T., Fair, D.A., 2018. Heritability of the human connectome: a connectotyping study. Netw. Neurosci. 02, 175–199. https://doi.org/10.1162/netn_a_00029.
Moraczewski, D., Chen, G., Redcay, E., 2018. Inter-subject synchrony as an index of functional specialization in early childhood. Sci. Rep. 8, 1–12. https://doi.org/10.1038/s41598-018-20600-0.
Murphy, K., Fox, M.D., 2017. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. Neuroimage 154, 169–173. https://doi.org/10.1016/j.neuroimage.2016.11.052.
Niazy, R.K., Xie, J., Miller, K., Beckmann, C.F., Smith, S.M., 2011. Spectral characteristics of resting state networks. Prog. Brain Res. 259–276. https://doi.org/10.1016/B978-0-444-53839-0.00017-X.
Noble, S., Scheinost, D., Constable, R.T., 2021. A guide to the measurement and interpretation of fMRI test-retest reliability. Curr. Opin. Behav. Sci. 40, 27–32. https://doi.org/10.1016/j.cobeha.2020.12.012.
Parkes, L., Fulcher, B., Yücel, M., Fornito, A., 2018. An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. Neuroimage 171, 415–436. https://doi.org/10.1016/j.neuroimage.2017.12.073.
Patel, A.X., Kundu, P., Rubinov, M., Jones, P.S., Vértes, P.E., Ersche, K.D., Suckling, J., Bullmore, E.T., 2014. A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. Neuroimage 95, 287–304. https://doi.org/10.1016/j.neuroimage.2014.03.012.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59, 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018.
Power, J.D., Schlaggar, B.L., Petersen, S.E., 2015. Recent progress and outstanding issues in motion correction in resting state fMRI. Neuroimage 105, 536–551. https://doi.org/10.1016/j.neuroimage.2014.10.044.
Power, J.D., Plitt, M., Kundu, P., Bandettini, P.A., Martin, A., 2017. Temporal interpolation alters motion in fMRI scans: magnitudes and consequences for artifact detection. PLoS One 12. https://doi.org/10.1371/journal.pone.0182939.
Pruim, R.H.R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J.K., Beckmann, C.F., 2015. ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. Neuroimage 112, 267–277. https://doi.org/10.1016/j.neuroimage.2015.02.064.
Reynolds, J.E., Long, X., Paniukov, D., Bagshawe, M., Lebel, C., 2020. Calgary preschool magnetic resonance imaging (MRI) dataset. Data Br. 29. https://doi.org/10.1016/j.dib.2020.105224.
Rohr, C.S., Vinette, S.A., Parsons, K.A.L., Cho, I.Y.K., Dimond, D., Benischek, A., Lebel, C., Dewey, D., Bray, S., 2017. Functional connectivity of the dorsal attention network predicts selective attention in 4–7 year-old girls. Cereb. Cortex 27, 4350–4360. https://doi.org/10.1093/cercor/bhw236.
Rohr, C.S., Dimond, D., Schuetze, M., Cho, I.Y.K., Lichtenstein-Vidne, L., Okon-Singer, H., Dewey, D., Bray, S., 2019. Girls' attentive traits associate with cerebellar to dorsal attention and default mode network connectivity. Neuropsychologia 127, 84–92. https://doi.org/10.1016/j.neuropsychologia.2019.02.011.
Saad, Z.S., Gotts, S.J., Murphy, K., Chen, G., jo, H.J., Martin, A., Cox, R.W., 2012. Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. Brain Connect. 2, 25–32. https://doi.org/10.1089/brain.2012.0080.
Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M., 2014. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90, 449–468. https://doi.org/10.1016/j.neuroimage.2013.11.046.
Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C., Gur, R.E., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. Neuroimage 60, 623–632. https://doi.org/10.1016/j.neuroimage.2011.12.063.
Satterthwaite, T.D., Elliott, M.A., Gerraty, R.T., Ruparel, K., Loughead, J., Calkins, M.E., Eickhoff, S.B., Hakonarson, H., Gur, R.C., Gur, R.E., Wolf, D.H., 2013. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. Neuroimage 64, 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052.
Satterthwaite, T.D., Ciric, R., Roalf, D.R., Davatzikos, C., Bassett, D.S., Wolf, D.H., 2019. Motion artifact in studies of functional connectivity: characteristics and mitigation strategies. Hum. Brain Mapp. 40, 2033–2051. https://doi.org/10.1002/hbm.23665.
Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86, 420–428. https://doi.org/10.1037//0033-2909.86.2.420.
Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155. https://doi.org/10.1002/hbm.10062.
Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. https://doi.org/10.1016/j.neuroimage.2004.07.051.
Taymourtash, A., Schwartz, E., Nenning, K., Sobotka, D., Diogo, M., Kasprian, G., Prayer, D., Langs, G., 2020. Quantifying Residual Motion Artifacts in Fetal fMRI Data. arXiv: 2001.03739.

Thomas, C.G., Harshman, R.A., Menon, R.S., 2002. Noise reduction in BOLD-based fMRI using component analysis. Neuroimage 17, 1521–1537. https://doi.org/10.1006/nimg.2002.1200.

Urchs, S., Armoza, J., Moreau, C., Benhajali, Y., St-Aubin, J., Orban, P., Bellec, P., 2019. MIST: a multi-resolution parcellation of functional brain networks. MNI Open Res. 1, 3. https://doi.org/10.12688/mniopenres.12767.2.

Van Dijk, K.R.A., Sabuncu, M.R., Buckner, R.L., 2012. The influence of head motion on intrinsic functional connectivity MRI. Neuroimage 59, 431–438. https://doi.org/10.1016/j.neuroimage.2011.07.044.

Vanderwal, T., Eilbott, J., Castellanos, F.X., 2019. Movies in the magnet: naturalistic paradigms in developmental functional neuroimaging. Dev. Cogn. Neurosci. 36, 100600 https://doi.org/10.1016/j.dcn.2018.10.004.

Vanderwal, T., Eilbott, J., Kelly, C., Frew, S.R., Woodward, T.S., Milham, M.P., Castellanos, F.X., 2021. Stability and similarity of the pediatric connectome as developmental measures. Neuroimage 226, 117537. https://doi.org/10.1016/j.neuroimage.2020.117537.

Xia, M., Wang, J., He, Y., 2013. BrainNet viewer: a network visualization tool for human brain connectomics. PLoS One 8 (7). https://doi.org/10.1371/journal.pone.0068910.

Zeng, L., Wang, D., Fox, M.D., Sabuncu, M., Hu, D., Ge, M., Buckner, R.L., Liu, H., 2014. Neurobiological basis of head motion in brain imaging. Proc. Natl. Acad. Sci. U.S.A. 111 (16), 6058–6062. https://doi.org/10.1073/pnas.1317424111.