Benchmarking common preprocessing strategies in early childhood functional connectivity MRI

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

Corresponding author:

Kirk Graff

kirk.graff@ucalgary.ca

Alberta Children's Hospital Research Institute

B4-290, 28 Oki Drive NW, Calgary, AB, T3B 6A8, Canada

1   **Highlights**

2   -We evaluated 27 preprocessing pipelines in passive viewing data from young children

3   -Pipelines were evaluated on noise-removed and information retained

4   -Pipelines that included censoring and GSR outperformed alternatives across benchmarks

5   -For high-motion scans, preprocessing choices substantially alter connectomes

6

7

8   **Keywords**

9   fMRI, functional connectivity, motion, noise, connectome fingerprinting, intersubject

10   correlation

1   **Abstract**

2   Functional connectivity magnetic resonance imaging (FC-MRI) has been widely used to

3   investigate neurodevelopment. However, FC-MRI is vulnerable to head motion, which is

4   associated with age and distorts FC estimates. Numerous preprocessing strategies have been

5   developed to mitigate confounds, each with advantages and drawbacks. Preprocessing

6   strategies for FC-MRI have typically been validated and compared using resting state data from

7   adults. However, FC-MRI in young children presents a unique challenge due to relatively high

8   head motion and a growing use of passive viewing paradigms to mitigate motion. This

9   highlights a need to compare processing choices in pediatric samples. To this end, we leveraged

10   longitudinal, passive viewing fMRI data collected from 4 to 8-year-old children. We

11   systematically investigated combinations of widely used and debated preprocessing strategies,

12   namely global signal regression, volume censoring, ICA-AROMA, and bandpass filtering. We

13   implemented commonly used metrics of noise removal (i.e. quality control-functional

14   connectivity), metrics sensitive to individual differences (i.e. connectome fingerprinting), and,

15   because data was collected during a passive viewing task, we also assessed the impact on

16   stimulus-evoked responses (i.e. intersubject correlations; ISC). We found that the most

17   efficacious pipeline included censoring, global signal regression, bandpass filtering, and head

18   motion parameter regression. Despite the drawbacks of noise-mitigation steps, our findings

19   show benefits for both noise removal and information retention in a high-motion early

20   childhood sample.

1. Introduction

Functional connectivity magnetic resonance imaging (FC-MRI) enables investigation of

functional networks in the brain. 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 is particularly problematic in

developmental neuroimaging studies, which face the challenge of age-correlated head motion

(Power et al., 2012; Fair et al., 2012). To reduce head motion and increase compliance, FC-MRI

studies in children are increasingly conducted using passive viewing tasks such as movies (Rohr

et al., 2017; Greene et al., 2018). However, head motion noise remains a major concern, and

preprocessing choices are therefore important in studies of young children. There is currently

no gold-standard preprocessing pipeline in the literature, for pediatric studies or otherwise, and

noise mitigation steps involve tradeoffs and remain widely debated (Murphy and Fox, 2017;

Satterthwaite et al., 2019). Common strategies include regression of motion estimates,

independent component analysis (ICA)-based approaches, regression of the global signal,

censoring volumes of high framewise displacement, and temporal filtering.

Regressing out motion estimates remains one of the most common denoising

approaches (Satterthwaite et al., 2019). However, based on pipeline 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

1   classifies components as noise by both temporal features (high frequency content, correlation

2   with realignment parameters) and spatial features (near CSF or the edge of the brain) (Pruim et

3   al., 2015). These structured noise components can then be regressed from the data (Thomas et

4   al., 2002).

5   Global signal regression (GSR) is an often used but widely debated preprocessing step

6   (Murphy and Fox, 2017; Chai et al., 2012; Gotts et al., 2013). GSR is a simple and arguably

7   effective (Ciric et al., 2017; Parkes et al., 2018) denoising technique, improving the specificity of

8   positive correlations and showing results that are more consistent with anatomical connectivity

9   (Fox et al., 2009). However, the use of GSR tends to increase the apparent strength of short-

10   range connections while decreasing that of long-range connections (Saad et al., 2012; Ciric et

11   al., 2017; Parkes et al., 2018), decreases positive BOLD responses, and creates anti-correlations

12   which may not exist (Aguirre et al., 1998). The global signal has been shown to resemble

13   established networks and to be significantly related to life outcomes and psychological

14   function, suggesting it is rich in information on cognition and behavior (Li et al., 2019), and that

15   by extension GSR should be used with caution as it removes information specific to the

16   individual.

17   To mitigate the effect of specific motion contaminated volumes, it has become common

18   to either remove, or interpolate over, specific time points by censoring or 'scrubbing' (Power et

19   al., 2012). Despite its apparent efficacy in removing noise (Ciric et al., 2017), censoring has

20   concerns, such as disrupting temporal autocorrelations, and leaving variable amounts of scan

21   data between participants. Further, even if the number of censored volumes is matched, not all

22   censored volumes are equally rich in information (Power et al., 2015). Censoring regimes vary

1   widely across studies (Satterthwaite et al., 2013; Power et al., 2012; Power et al., 2014) and it is

2   a challenge to determine which level of censoring optimizes the recovery of individually specific

3   FC information.

4   Another preprocessing strategy is temporal filtering. fMRI signals at both very low

5   frequencies (under 0.01 Hz) and high frequencies (above 0.1 Hz) are often filtered out to

6   remove noise (Satterthwaite et al., 2013). However, filtering above 0.1 Hz may also be

7   removing connectivity information (Niazy et al., 2011), or artificially increasing correlations by

8   introducing sample dependence (Davey et al., 2013), leading to concern on the appropriateness

9   of a bandpass filter compared to a highpass filter (Satterthwaite et al., 2019).

10   In response to the proliferation of preprocessing approaches (Carp, 2012), several

11   studies have compared the effectiveness of different pipelines (Churchill et al., 2017; Ciric et al.,

12   2017; Parkes et al., 2018). However, to our knowledge, no studies have systematically

13   compared preprocessing pipelines in pediatric samples nor in studies that have used passive

14   viewing tasks, which is important for several reasons. Passive viewing tasks alter FC relative to

15   rest (Bray et al., 2015; Vanderwal et al., 2017), and have been shown to provide more stable FC

16   estimates (Wang et al., 2017), which in turn may result in enhanced ability to identify brain-

17   behavior correlations (Vanderwal et al., 2019). It remains unknown to what extent these FC

18   changes are impacted by preprocessing choices. Young children typically have higher head

19   motion than adults (Dosenbach et al., 2017), however, their smaller head size could alter the

20   impact of rotational motion. Similarly, Bolton et al. (2020) found that in-scanner movement

21   patterns are related to anthropometric and cognitive factors, suggesting it may be

22   inappropriate to treat motion in children the same as motion in adults. Children also have

1   faster respiratory rates and heart rates than adults (Fleming et al., 2011), which likely alters the

2   physiological noise present in scans, and which is not accounted for with benchmarks that only

3   account for head motion, such as the most commonly used previous metric, quality control –

4   functional connectivity (QC-FC) (Satterthwaite et al., 2013; Power et al., 2012). At the same

5   time, there are fundamental differences in the brain structure of children relative to adults

6   (Grayson and Fair, 2017), and since GSR shows distance dependent effects, this step warrants

7   investigation in children specifically.

8   Preprocessing steps that aggressively remove signal related to motion involve tradeoffs.

9   Ciric et al. (2017) speculated that de-noising methods could improve QC-FC metrics and

10   improve reliability by removing both signal and noise, but in the process lose sensitivity to

11   individual differences of interest. These individual differences may include FC correlates of fluid

12   intelligence (Finn et al., 2015), developmental changes (Kaufmann et al., 2017), or symptom

13   load (Byrge and Kennedy, 2020), which may themselves be correlated with motion (Power et

14   al., 2015). A complementary metric that provides information about the individual-specific

15   information remaining following preprocessing is functional connectome fingerprinting (Finn et

16   al., 2015; subsequently referred to simply as "fingerprinting"), which aims to match scans from

17   the same participant through FC estimates. Preprocessing strategies that either fail to remove

18   noise, or remove too much signal of interest, will have poor individual identifiability.

19   Additionally, previous benchmarking studies using resting-state FC-MRI scans have not

20   assessed the effect of noise mitigation steps on stimulus evoked functional responses. Such

21   responses can be analyzed using intersubject correlation (ISC; Hasson et al., 2004), the

22   temporal signal correlation of the same brain region between individuals during passive viewing

1   of a video stimulus. Although distinct from FC, which is the main focus of this paper, this

2   analysis provides information about how stimulus evoked responses are enhanced or

3   suppressed by noise mitigation steps, thereby providing a more comprehensive picture of how

4   these steps impact functional time series.

5   The present study systematically compares common noise-mitigation steps in data

6   collected from young children, using QC-FC, fingerprinting, and ISC. The aim of this study is to

7   support researchers conducting FC-MRI in young samples who use naturalistic imaging methods

8   to consider the tradeoffs and effectiveness of several common preprocessing steps.

9

10

11   **2. Methods**

12   *2.1. Participants*

13   Data were from a study of early childhood brain development that included both single

14   timepoint and longitudinal measurements (Dimond et al., 2020a; Dimond et al., 2020b; Rohr et

15   al., 2019; Rohr et al., 2017). Participants were recruited from the local community through

16   advertisements and through existing databases. All procedures were approved by the

17   University of Calgary Conjoint Health Research Ethics Board. Parents provided informed

18   consent and children provided assent to participate. Participants were children between 4 and

19   7 years of age at baseline without any major health concerns and were excluded if they had full-

20   scale IQ more than 2 standard deviations below the mean (100), a history of neurodevelopment

21   or psychiatric disorders, or any neurological diagnoses. 168 participants were scanned at

22   baseline. Of those, 139 provided usable baseline data (i.e. free of excessive head motion), and

1 of those 59 (15 male) produced usable 12-month follow-up data. From the sample of

2 participants with both baseline and follow-up data, participants were included if after volume-

3 wise censoring of the fMRI data (described in more detail below), both scans retained at least

4 11 minutes of uncensored (retained) data. 56 of the 59 children (14 male) reached this

5 threshold, contributing 112 scans.

6

7 *2.2. Data collection*

8   Children were scanned at the Alberta Children's Hospital. During fMRI scans a passive-

9 viewing task was used, where participants watched clips from a children's television show

10 (Elmo's World) for 1100 s. Prior to scans, children underwent a practice scan in an MRI

11 simulator during which they watched the video that would be used during the fMRI scan and

12 practiced staying still. Data was acquired with a 3T GE MR750 w (Waukesha, WI) scanner using

13 a 32-channel head coil. An anatomical scan was acquired using a T1w 3D BRAVO sequence (TR =

14 6.764 ms, TE = 2.908 ms, FA = 10°, voxel size 0.8x0.8x0.8 mm^3). fMRI scans were collected using

15 a gradient-echo EPI sequence (TR = 2.5 s, TE = 30 ms, FA = 70°, voxel size 3.5x3.5x3.5 mm^3).

16

17 *2.3. Higher- vs. lower-motion subgroups*

18 Preprocessing pipelines were compared using three categories of metrics using the whole

19 sample of 112 scans. To further assess whether preprocessing pipelines were particularly

20 advantageous for scans with relatively lower or higher amounts of head motion, we median-

21 split the 112 scans into two groups of 56 scans, then re-calculated QC-FC metrics and ISC

22 metrics described below.

1

2.4. Data and code availability

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. Software packages mentioned (Nipype, FSL, AFNI, ANTs) provided on their respective websites. Python scripts available at https://github.com/BrayNeuroimagingLab/BNL_open/tree/main/fMRI_preprocessing.

2.5. Basic preprocessing steps common across pipelines

The following preprocessing steps were run on all scans prior to the custom elements of pipelines, 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. This included bias field correction, brain extraction, and tissue segmentation. Tissue segments 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 average relative framewise displacement (Jenkinson et al., 2002). b) FSL *slicetimer* for slice time correction. c) FSL *MCFLIRT* for rigid body realignment. We followed the recommendation

1   in Power et al. (2017) to estimate HMP on raw data but carry out slice time correction prior to

2   rigid body realignment, necessitating two FSL *MCFLIRT* steps. d) FSL *BET* to skull-strip EPI

3   images (Smith, 2002). e) ANTs *Registration* (Avants et al., 2011) to generate a transformation

4   matrix to warp the EPI image to a study-specific EPI template. This template was produced

5   based on the procedure described by Huang et al. (2010). Specifically, a 3D EPI reference image

6   was taken from each fMRI scan, chosen as a volume of low motion approximately in the middle

7   of the scan. These references were warped to MNI space, then averaged together and

8   smoothed to create the final study-specific template. f) FSL *FLIRT* boundary-based registration

9   (Jenkinson et al., 2002) was used to generate a transformation matrix to warp the EPI image to

10   the T1w image, then the inverse transformation matrix was used to warp tissue segmentations

11   to functional image space. All preprocessing and confound mitigation steps were carried out in

12   native space. g) A linear regression to remove the mean and linear and quadratic trends from

13   each voxel was conducted. For pipelines that include censoring, time points marked for

14   censoring were excluded from the model to calculate linear and quadratic trends. In pipelines

15   without censoring, all timepoints were included.

16

17   *2.6. Preprocessing pipelines tested*

18   After first-stage preprocessing, pipelines varied systematically in whether they used

19   GSR, censoring, and ICA-AROMA, as shown in Table 1, creating pipelines A1 through A8.

Pipeline	Primary motion artifact removal	Global signal regression (GSR)	Censoring
A1	Regress head motion parameters (HMPs)		
A2		Yes	
A3			Yes
A4		Yes	Yes
A5	ICA-AROMA		
A6		Yes	
A7			Yes
A8		Yes	Yes

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

1 *2.7. Additional preprocessing*

2 *2.7.1. ICA-AROMA*

3   For pipelines A5 through A8, ICA-AROMA was applied immediately following the first-

4 stage preprocessing steps described above. Given that this is a relatively high motion sample,

5 we used aggressive denoising but otherwise default options. The ICA-AROMA code was

6 modified to warp to MNI space via our ANTs transformation matrix rather than a FNIRT

7 transformation.

8

9 *2.7.2. Temporal filtering*

10 Following the common preprocessing steps (pipelines A1 through A4) or ICA-AROMA (A5

11 through A8), all pipelines underwent bandpass temporal filtering (0.01 – 0.08 Hz) done via a

12 fast Fourier transformation. In follow up comparisons we tested the effect of a highpass filter;

13 see section *2.9.1. Filtering.* To avoid reintroducing artifacts, HMPs were also filtered, and WM,

14 CSF, and the global signal were calculated following temporal filtering (Lindquist et al., 2019).

15

1   *2.7.3. Nuisance regression*

2   Following filtering, nuisance regression was applied. All pipelines, including ICA-AROMA

3   pipelines, included WM + CSF regression. Pipelines A1 through A4 included regressing the six

4   HMPs, while pipelines A2, A4, A6 and A8 included GSR. For all nuisance parameters, both linear

5   and quadratic terms were included, along with the first temporal derivative of those terms (4

6   regressors per parameter). Censoring was carried out as part of the regression step for

7   pipelines A3, A4, A7 and A8. We censored volumes above a FD threshold of 0.25 mm (based on

8   FSL *MCFLIRT*, i.e. $FD_{Jenkinson}$ as described in Ciric et al., 2018), and censored only the identified

9   frames.

10

11   *2.8. Connectome generation*

12   Following the above preprocessing steps, each scan was registered to the study specific

13   template using the previously generated ANTs transformation matrix. Each voxel was then

14   assigned to one of 325 nodes (regions) within the MIST 325 parcellation (Urchs et al., 2019).

15   The mean time course was calculated for each node by averaging the time courses for all voxels

16   within the node. The Pearson correlation between each pair of nodes was then calculated,

17   generating 52650 edges. Correlation values were Fisher z- transformed to better approximate a

18   normal distribution.

19

20   *2.9. Pipeline comparison metrics*

21

22   *2.9.1. Motion correlated edges (QC-FC)*

1   Typically, preprocessing benchmark studies have looked at the amount of noise

2   remaining in a dataset following preprocessing by QC-FC, which is the correlation between the

3   strength of a given network edge across participants, with participant motion (as measured by

4   average framewise displacement) (Satterthwaite et al., 2013; Power et al., 2012). QC-FC is used

5   to examine the quantity of network edges where FC significantly relates to motion (which

6   should theoretically be zero, notwithstanding edges functionally involved in motion); better

7   preprocessing pipelines will have fewer edges with a significant relationship with head motion.

8   Previous work has also assessed whether a network edge's QC-FC is related to the inter-node

9   distance. This addresses whether short-range connections are more vulnerable to head-motion

10   than long-range connections, and if preprocessing steps alter the slope of this relationship

11   (Power et al., 2015).

12   Using the approach of Satterthwaite et al. (2013), for each edge the correlation was

13   calculated between scan motion (calculated as the average relative framewise displacement, as

14   estimated from the first FSL *MCFLIRT* step) and edge strength (i.e. the Fisher z-value), across all

15   112 scans. Due to the concern that preprocessing steps such as GSR affects edges differently

16   depending on the edge-length, we also assessed distance-dependent effects by plotting the

17   edge strength-motion correlation vs. Euclidian edge length for each edge (Satterthwaite et al.,

18   2013). Euclidian edge length was found by calculating the distance between the center of mass

19   of the two nodes, as defined by the MIST parcellation.

20   Three QC-FC metrics were extracted: (1) the percentage of edges correlated with head

21   motion at a p-value <0.05 uncorrected; (2) the mean of the absolute value of the correlation

22   between head motion and edge strength; absolute value was used to assess the magnitude of

1 the motion-effect, rather than the direction; and (3) the slope of the best-fit line of edge

2 correlation with motion vs Euclidian edge length.

3

4 *2.9.2. Fingerprinting*

5   While measures such as QC-FC can estimate noise removed during preprocessing, there

6 is a concern that pipelines that aggressively remove noise may also be removing signal of

7 interest (Ciric et al., 2017). To assess this, we used functional connectome fingerprinting to

8 gauge whether information unique to an individual remains after preprocessing. Following the

9 approach used by Finn et al. (2015) and related work (Byrge and Kennedy, 2019; Kaufmann et

10 al., 2017; Miranda-Domingeuz et al., 2018), connectomes were correlated for each pair of

11 scans. In this context, a scan's "complement" is the other scan from the same individual. For

12 example, the complement to the scan for child 1's initial scan is child 1's 12-month follow up

13 scan, and vice versa. A scan from child 2 would be a "non-complement" scan.

14   The fingerprinting match rate is calculated by comparing each scan against every other

15 scan and dividing the number of times a scan's highest correlation is its complement by the

16 total number of scans. Here we had 56 individuals, each with two scans, for a denominator of

17 112. We expressed this as a percentage. For each scan we also calculated the fingerprinting

18 margin by taking the difference in correlation between a scan and its complement, and the

19 highest correlation to a non-complement. This metric is indicative of the degree of

20 individualization. We also calculated the mean correlation between complement scans and

21 mean correlation between all non-complement scans.

1   To assess associations between individualization and head motion across pipelines, we

2   plotted the fingerprinting margin vs. head motion, then calculated the slope, intercept, and R^2.

3   For each scan-complement pair, we used the higher of the two motion values for this analysis.

4

5   *2.9.3. Intersubject correlation*

6   Because data was collected while participants watched the same set of videos, our data

7   offered the possibility to simultaneously assess the impact of preprocessing on functional

8   connectivity and task-evoked responses. To this end, we correlated the time series for each

9   node between all pairs of scans to find the intersubject correlation (ISC; Hasson et al., 2004).

10   Preprocessing pipelines that fail to adequately remove noise, or remove task-evoked signal will

11   have lower ISC values. For pipelines that included censoring, we did not include time points that

12   were censored for either scan in a given pair. From there, we calculated the mean ISC for each

13   node within each scan as the average of that individual's correlations to all other individuals.

14   Since ISCs have significant correlations in particular brain regions (Hasson et al., 2004; Kauppi et

15   al., 2010), such as the visual and auditory networks, we identified the 10 nodes with the highest

16   average ISCs across all pipelines and used the average of these 10 nodes to generate a single

17   ISC score for each scan for each pipeline. While many ISC studies only use one scan per

18   participant (Vanderwal et al., 2019), our analysis considers how the similarity between scans

19   might be impacted by preprocessing decisions, and we therefore included both scans per child.

20

21   *2.9.4. Intra-scan inter-pipeline correlation*

1   Preprocessing choices may have relatively large or relatively small effects on

2   connectomes, and these may influence downstream analyses and convergence across studies.

3   To assess the magnitude of effects of preprocessing choices on FC estimates, for each scan we

4   calculated the correlation between edge strengths across pipelines A1-A8. These were Fisher z-

5   transformed, averaged across scans, then converted back to correlations for ease of

6   comparison.

7

8   *2.10. Follow-up comparisons*

9

10   *2.10.1. Filtering*

11   Using a lowpass filter has become common in the preprocessing literature due to

12   frequencies above approximately 0.1 Hz being more highly associated with noise than signal of

13   interest (Satterthwaite et al., 2019). However, this remains controversial, as connectivity

14   information at higher frequencies will be lost (Niazy et al., 2011). We therefore repeated our

15   analysis using a highpass (>0.01 Hz) rather than a bandpass filter, allowing us to compare the

16   same eight pipelines. All other preprocessing steps were identical.

17

18   *2.10.2. Varying censoring parameters*

19   Following our initial comparison of pipelines, we conducted an additional comparison

20   focusing specifically on the effect of censoring at different thresholds. There is variation in

21   thresholds used in the literature, with older studies generally using more lenient censoring than

22   more recent studies (Satterthwaite et al., 2019). Similarly, the literature has advocated both

1   censoring a single volume per motion artifact (Satterthwaite et al., 2013) and censoring

2   multiple volumes per motion artifact (Power et al., 2012). Thresholding decisions are

3   challenging because there is limited guidance in pediatric populations. Stricter censoring

4   thresholds will reduce the number of time points, which in turn can result in less reliable

5   functional connectivity estimates (Gordon et al., 2017).

Pipeline	Volumes censored per artifact	Censoring threshold (mm FD)
B1	1	0.3
B2 (A4)		0.25
B3		0.2
B4		0.15
B5	3	0.3
B6		0.25
B7		0.2
B8		0.15
B9	4	0.3
B10		0.25
B11		0.2
B12		0.15

Table 2: Preprocessing pipelines tested to compare effects of censoring at different thresholds and volumes censored per motion artifact

6

7   We used pipeline A4 above, which performed well on all metrics, and modified the

8   approach to censoring using 11 additional pipelines that varied in the censoring threshold and

9   the volumes censored per motion artifact (Table 2). Pipelines that censor 3 volumes per motion

10   artifact censor volumes immediately before and after any region of motion, despite neither

11   reaching the FD threshold for censoring. When two or more volumes in a row both meet the FD

12   threshold for censoring, only 2 additional volumes are censored, one before and one after the

13   motion region. Similarly, for pipelines that censor 4 volumes per motion artifact, 1 volume prior

14   and 2 volumes after the motion-contaminated volume(s) are censored. The same 56

1   participants (112 scans) were included across comparisons, even if stricter censoring caused

2   scans to fall below the threshold of 11 minutes of uncensored data.

3

4

5   **3. Results**

6   *3.1. Participant characteristics*

7   The mean age of participants for their baseline scan was 5.47 years old (standard

8   deviation: 0.76 years), and 6.52 for their second scan. Time from initial to follow up scan ranged

9   from 0.88 to 1.19 years (mean: = 1.05 years, SD = 0.069 years). The average motion of our

10   sample was high, with a mean average relative FD of 0.126 mm across all 112 scans (median

11   mean relative FD of 0.095 mm). The correlation in mean motion (average relative FD) between

12   scans from the same participant was not significant (r = 0.18, p = 0.19). Across the 112 scans,

13   age and mean FD were not significantly correlated (r = -0.14, p = 0.14).

14   For analyses that divided scans into two groups based on median head motion, the 56

15   low motion scans had a mean average FD of 0.062 mm (range: 0.035 to 0.094 mm, std = 0.016

16   mm). The 56 high motion scans had a mean average FD of 0.191 mm (range: 0.097 to 0.506

17   mm, std = 0.088 mm). Mean age at time of scan was similar in both groups (low motion: 6.01

18   years, std = 0.920 years; high motion: 5.98 years, std = 0.935 years).

19

20   *3.2. QC-FC*

21   QC-FC associations were assessed for each pipeline (Figure 1). Figure 1a shows the

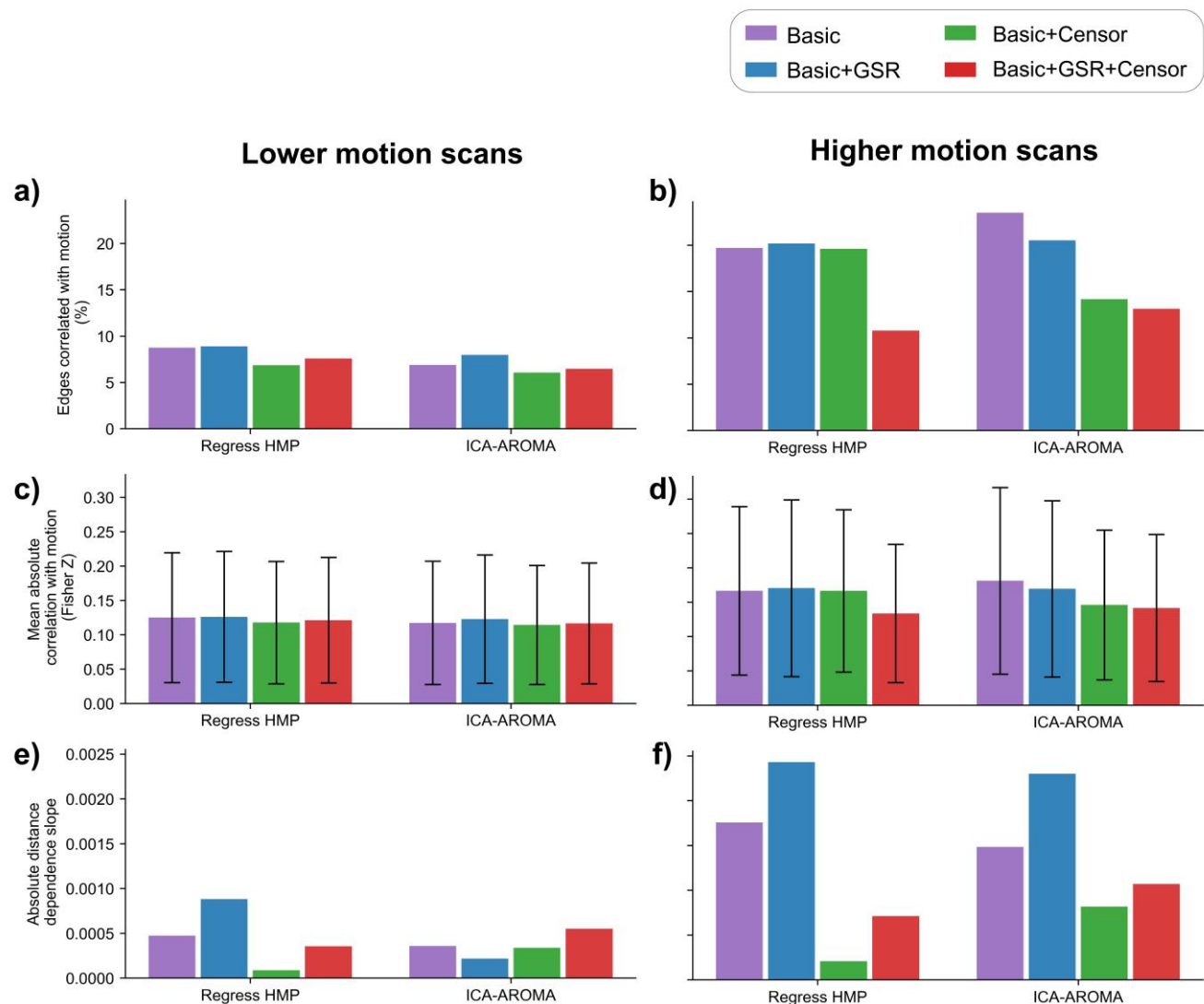22   percentage of edges significantly correlated with motion, ranging from 42% to 20%. Both GSR
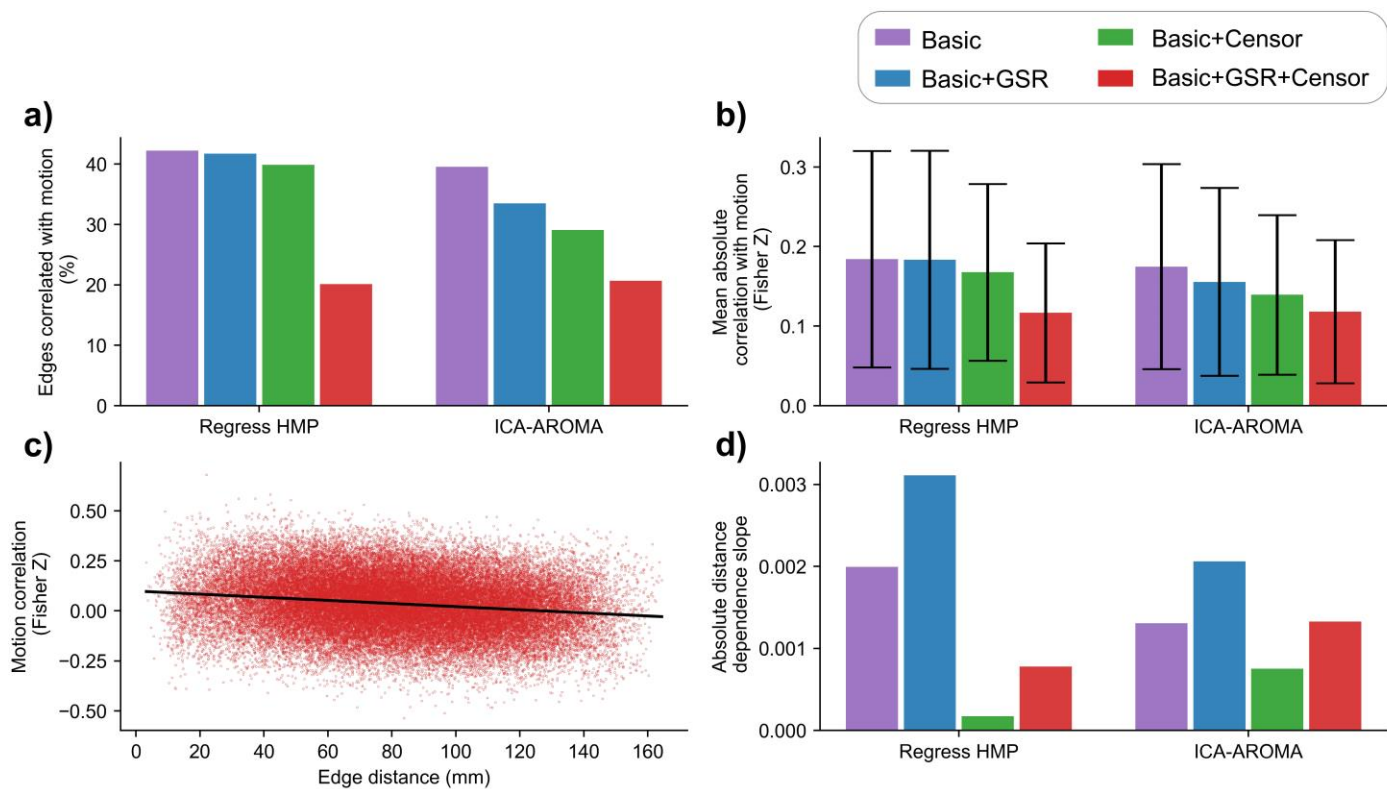
Figure 1. **Quality control-functional connectivity (QC-FC) across pipelines.** Plots are arranged from pipeline A1 through A8, as described in Table 1. **a)** Percentage of edges with a significant correlation between edge strength and subject motion across all 112 scans (uncorrected p < 0.05). **b)** Mean and standard deviation of absolute correlation between edge strength and subject motion across all 112 scans. **c)** Example QC-FC distance dependence plot, from pipeline A4 (regress HMP + 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 subjection motion. **d)** Absolute value of the slope of each pipeline's distance-dependence plot.

1   and censoring reduced the number of edges correlated with motion, and pipelines that paired

2   these two steps had the smallest number of motion-associated edges. Pipelines A1 (regress

3   HMP) and A5 (ICA-AROMA), lacking both GSR and censoring, had 42% and 40% of edges

4   correlated with motion respectively, which dropped to 20% (A4) and 21% (A8) respectively with

5   the inclusion of combined GSR and censoring. In general, ICA-AROMA pipelines fared better

6   than comparable pipelines that regressed HMP, though the pipeline with the fewest QC-FC

7   associated edges (A4 – regress HMP+GSR+censoring) did not use ICA-AROMA.

1   Figure 1b shows the mean absolute correlation with motion for all edges. Here the same

2   trends are observed, with GSR and censoring reducing the effect of motion. A1 and A5 had

3   mean absolute correlations with motion of z = 0.184 and z = 0.175 respectively. When adding

4   both GSR and censoring (A4 and A8), these dropped to z = 0.117 and z = 0.118 respectively,

5   with only smaller improvements when using only one of GSR or censoring.

6   Figure 1c shows an example from pipeline A4 of edge correlation with motion vs. edge

7   distance with the slope indicated. Figure 1d shows the slope values for each pipeline. Our

8   results agreed with previous literature that GSR is associated with a more negative slope,

9   suggesting GSR affects shorter edges differently than longer edges, with shorter edges having

10   more remaining motion influence, but longer edges more likely to be negatively correlated with

11   motion. For example, A1 (regress HMP+no GSR+no censoring) had a slope of -0.0020, which

12   increased to -0.0031 when adding GSR (A2). However, we found that censoring reduces the

13   slope, partially compensating for the effect of GSR. The two pipelines with the smallest slope

14   were A3 (regress HMP+no GSR+censoring), with a slope of -0.0002 and A4 (regress

15   HMP+GSR+censoring) with a slope of -0.0008.

16   When scans were split into lower- and higher- motion subsets, the higher-motion scans

17   fared worse on all metrics (Figure 2a vs 2b, 2c vs 2d, 2e vs 2f). Both the lower- and higher-

18   motion groups had fewer edges significantly correlated with motion than the entire group

19   (Figures 2a and 2b vs 1a), which we attribute to a smaller sample size and a smaller range of

20   possible motion values. In the low motion group, the choice of pipeline had minimal effect on

21   the percent of edges correlated with motion (ranging from 6.1% to 8.9%; Figure 2a) and the

22   mean absolute correlation with motion (ranging from z = 0.11 to z = 0.13; Figure 2c). Small

Figure 2. **Quality control-functional connectivity (QC-FC) across pipelines, separately for lower or higher motion scans.** Higher and lower motion participants were identified as above or below median average framewise displacement. Plots are arranged from pipeline A1 through A8, as described in Table 1. **a, b)** Percentage of edges that have a significant correlation between edge strength and subject motion (uncorrected p < 0.05). **c, d)** Mean and standard deviation of absolute correlation between edge strength and subject motion. **e, f)** Absolute value of the slope of each pipeline's distance-dependence plot.

1   improvements were seen with censoring, and pipelines that included GSR fared slightly worse.

2   Most of the effects noted in the entire sample (Figure 1) were only seen in the high motion

3   group (Figures 2b and 2d), where censoring and GSR used together improve the percentage of

4   edges significantly correlated with motion (e.g. pipeline A1 to A4 dropping from 20% to 11%)

1   and the mean absolute correlation with motion (e.g. A1 to A4 dropping from z = 0.17 to 0.13).

2   Interpipeline slope trends were similar for both lower- and higher-motion groups as in the

3   entire sample (2e and 2f vs 1d), with GSR making the slope more negative, but censoring

4   mitigating the effect.

5

6   *3.3. Fingerprinting*

7   Fingerprinting match-rate was assessed for each pipeline (Figure 3a). Overall, the match

8   rate was high for all pipelines tested, ranging from 71% to 90% (chance level is <1%). GSR had a

9   minimal effect on the overall match rate (e.g. 71.4% for A1 – regress HMP+no GSR+no censor,

10   and 70.5% for A2 – regress HMP+GSR+no censor), but pipelines that included censoring were

11   more successful. The two pipelines with the highest match rate included censoring and

12   regressing HMP rather than using ICA-AROMA (A3 – no GSR and A4 - GSR; 89.3% and 90.2%

13   respectively).

14   Scans' mean correlation to non-complement scans varied minimally between

15   participants and across pipelines (Figure 3b), suggesting group network structure can be

16   consistently found regardless of preprocessing choices. While the correlation between a scan

17   and its complement showed a much larger range of values, this is likely an artifact of being a

18   single correlation, rather than an average. The two pipelines with the highest fingerprinting

19   match-rate, A3 and A4, were the ones with the highest mean correlation between scan-

20   complement pairs (A3 – z = 0.82 and A4 – z = 0.80), though interestingly pipeline A8 (ICA-

21   AROMA+GSR+censoring) had a comparable mean correlation between scan-complement pair (z

22   = 0.79), though a far larger mean correlation to non-complement scans (z = 0.61, vs z = 0.57 for
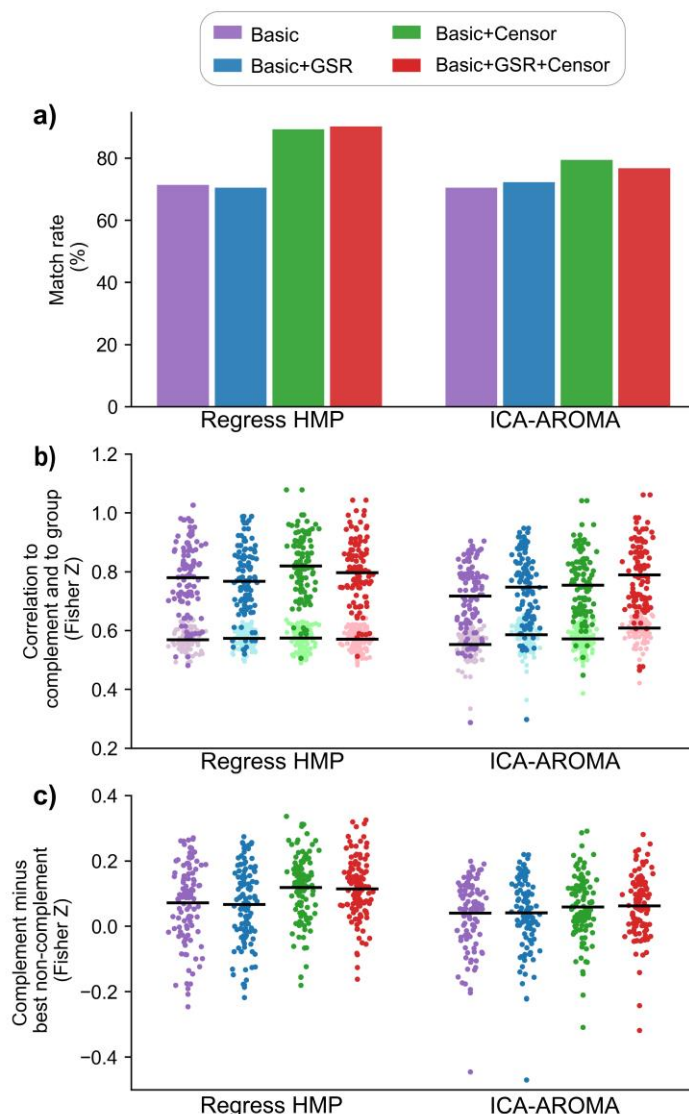
Figure 3. Functional connectome fingerprinting by pipeline. Plots are arranged from pipeline A1 through A8, as described in Table 1. **a)** Match rate across pipelines. A scan matched if its highest correlation was to its complement scan (other scan from the same individual). **b)** Each scan's correlation to their complement scan (darker points) and their average correlation to every other scan (lighter points), for each pipeline. Lines represent mean values. **c)** Fingerprinting margin, i.e. difference between each scan's correlation to their complement and to their best (highest correlation) non-complement. Any point below 0 fails to successfully match. Lines represent mean values.

1   both A3 and A4).

2   Figure 3c shows the fingerprinting margin (i.e. complement minus best non-

3   complement) for each of the 112 scans. As expected, pipelines with a higher match-rate (Figure

4   3a) have a higher average fingerprinting margin, which ranged from z = 0.040 to z = 0.072 for

1   the less successful pipelines (A1, A2, A5 through A8), while reaching z = 0.12 and z = 0.11 for A3

2   and A4 respectively. The group distribution for this metric was relatively consistent, with more

3   successful pipelines shifting the distribution upwards. This suggests that with certain pipelines

4   scans that fail to match with their complement still see greater recovery of individual

5   information via a reduced margin of failure (e.g. a scan's fingerprinting margin may improve

6   from -0.2 to -0.1), and scans that successfully match pass with an improved margin (e.g.

7   improving from +0.1 to +0.2).

8   Figure 4 shows the fingerprinting margins (from Figure 3c) plotted against head motion.

9   A more negative slope suggests that the pipeline performs progressively worse at identifying

10   individuals with more motion, and a higher r^2 value suggests that data run through the pipeline

11   is more sensitive to motion noise. In general, pipelines that regress HMP rather than use ICA-

12   AROMA have less negative slopes, lower r^2 values, and higher intercepts. For example, pipeline

13   A1 (regress HMP+no GSR+no censor) has a slope = -0.51, intercept = 0.20, r^2 = 0.36, while A5

14   (ICA-AROMA+no GSR+no censor) has slope = -0.62, intercept = 0.19, r^2 = 0.41. Pipelines that

15   include GSR have more negative slopes and higher r^2 values, for example going from A1 to A2

16   increases the slope from -0.51 to -0.56, and r^2 increases from 0.36 to 0.40. Censoring has

17   minimal effect on slope, but raises the intercept and lowers the r^2, for example going from A1

18   to A3 raises the intercept from 0.20 to 0.22, and lowers r^2 from 0.36 to 0.25. These suggest that

19   censoring benefits all scans, but GSR has minimal benefit for fingerprinting, especially at higher

20   motion. The participant with the most motion (mean relative FD of 0.51 mm for one of their

21   two scans) could be considered an outlier – repeating the analysis with that participant

22   removed reduces the r^2 and makes the slopes less negative, but trends between pipelines
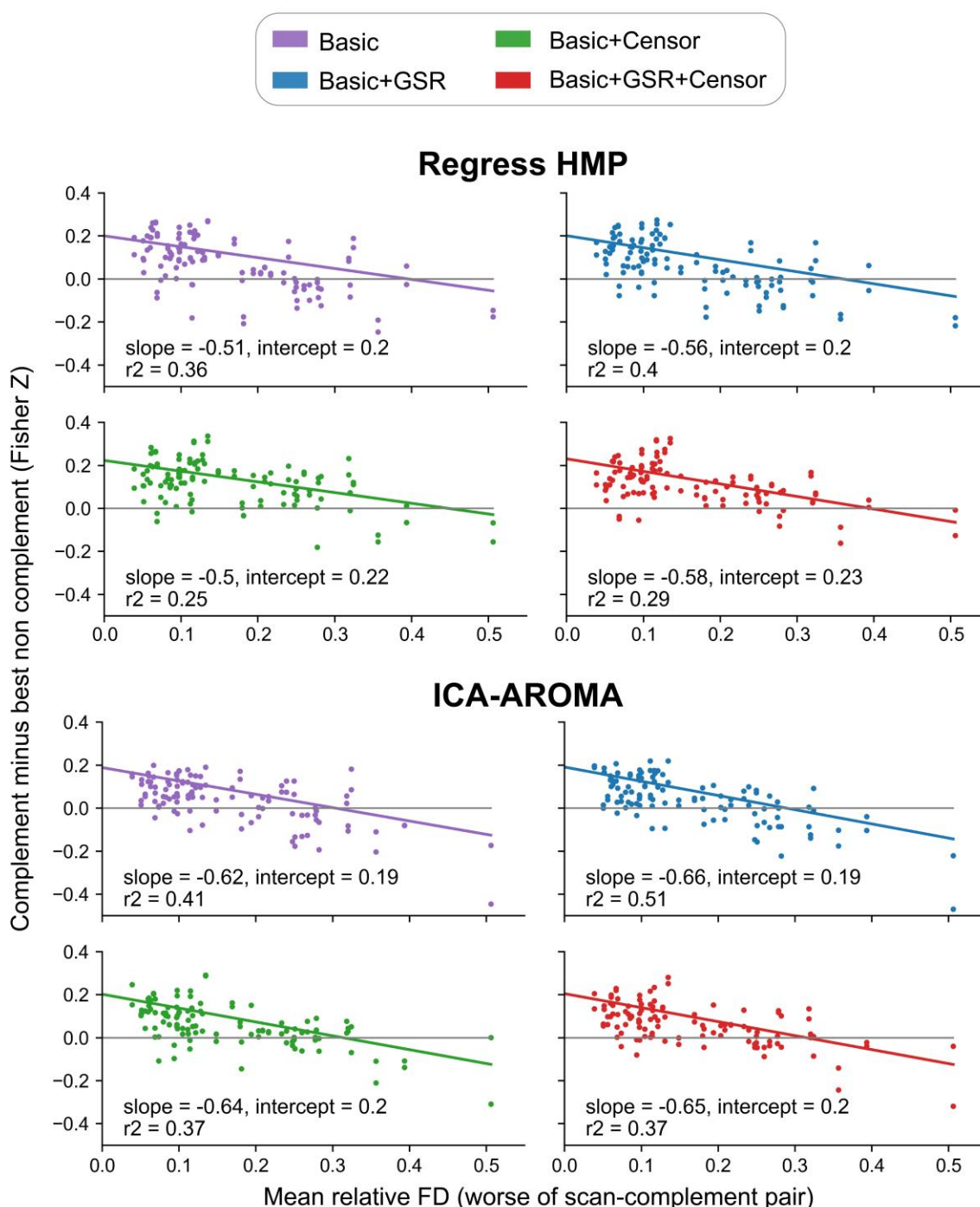
Figure 4. Fingerprinting margin as a function of head motion. The difference between each scan's correlation to their complement scan minus their correlation to the best (highest correlation) non-complement, plotted by motion. Motion was calculated by taking the worse of each participant's two scans' 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. Plots are arranged, left to right, from pipeline A1 through A8, as described in Table 1.

1   remain (Supplementary Figure 1).
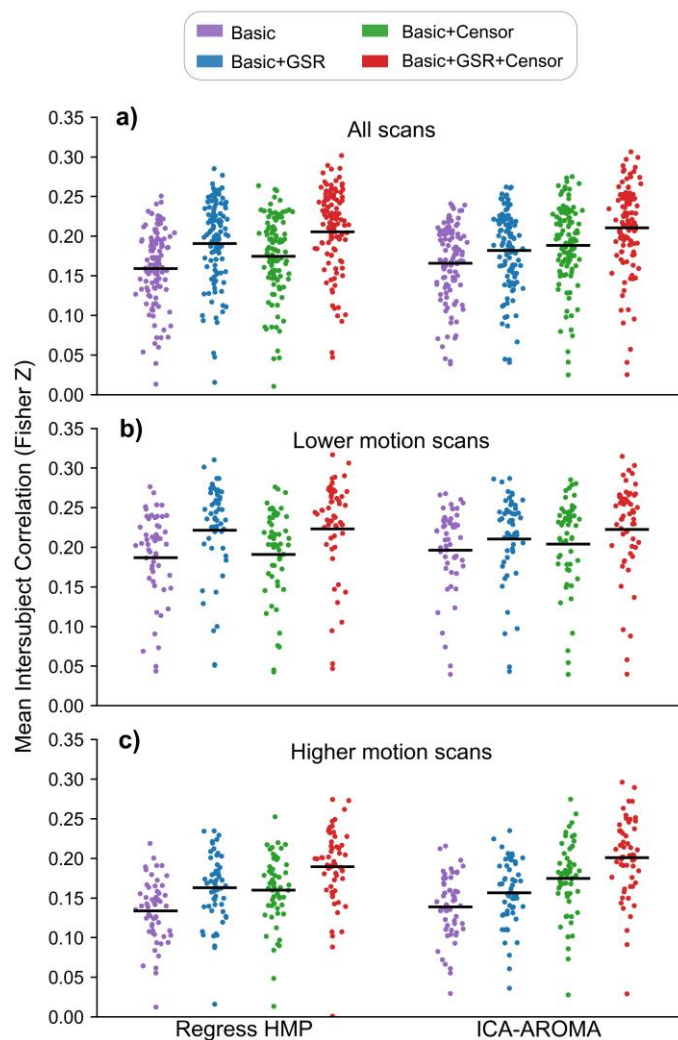
Figure 5. Mean intersubject correlation (ISC) values.
Each point represents the mean ISC for a scan to all other scans, averaged across the 10 parcels with highest ISC. Lines represent mean values. Plots are arranged from pipeline A1 through A8, as described in Table 1. **a)** Mean ISCs for all 112 scans, each compared to all other scans. **b)** Mean ISCs for each of the 56 below-median-motion scans, compared to the other 55. **c)** Mean ISCs for each of the 56 above-median-motion scans, compared to the other 55.

2   *3.4. Intersubject correlation*

3   Figure 5 shows each participant's mean intersubject correlation scores (ISCs) across

4   pipelines, based on the 10 parcels with the highest scores. Both GSR and censoring improved

1   ISCs, incrementally when applied together. Pipelines A1 (regress HMP) and A5 (ICA-AROMA),

2   both lacking GSR and censoring, had mean ISCs of z = 0.16 and z = 0.17 respectively, which

3   increased to z = 0.21 for both pipelines when both strategies were added (pipelines A4 and A8).

4   Pipelines that used ICA-AROMA had slightly higher ISCs than pipelines that did not, though the

5   gain was small.

6

7   When only low motion scans were compared using ISC (Figure 5b) the benefits of

8   censoring were minimal, with or without GSR. GSR itself increased ISCs, for example A2 (regress

9   HMP+GSR+no censor) had a mean ISC of z = 0.22, compared with z = 0.19 for A1 (regress

10   HMP+no GSR+no censor). However, when only high motion scans were compared (Figure 5c),

11   the benefits of censoring were more apparent, especially in combination with GSR. For

12   example, pipeline A2 (regress HMP+GSR+no censor) had a mean ISC of z = 0.16, which

13   increased to z = 0.19 for A4 (regress HMP+GSR+censor).

14

15   *3.5. Intrascan inter-pipeline correlation*

16   Figure 6 shows the within-scan between-pipeline correlation of whole brain functional

17   connectivity estimates. These correlations were converted to z-scores for averaging, then

18   converted back to correlations for ease of interpretability. While correlations were high overall,

19   ranging from 0.67 to 0.97, a correlation as low as 0.67 for the same scans preprocessed in two

20   different ways (in this case between pipelines A3 (regress HMP+no GSR+censor) and A6 (ICA-

21   AROMA+GSR+no censor) highlights concerns about replicability given the overall influence of

22   preprocessing.
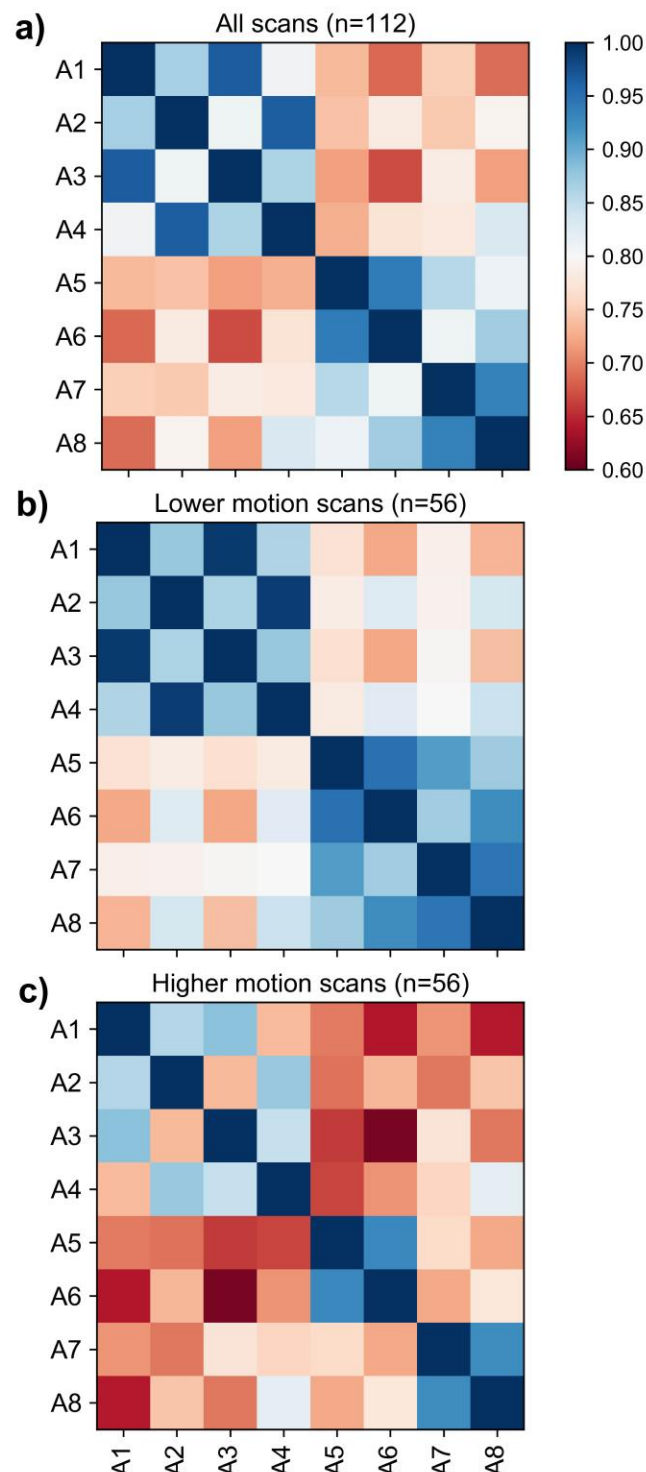
Figure 6. Intrascan inter-pipeline correlations. For each scan, the correlation between edge strengths was compared across pipelines, Fisher z-transformed, averaged across scans, then converted back to correlations. Pipelines are listed in Table 1, briefly: A1 – regress HMP; A2 – regress HMP + GSR; A3 – regress HMP + censor; A4 – regress HMP + GSR + censor; A5 – ICA-AROMA; A6- ICA-AROMA + GSR; A7 – ICA-AROMA + censor; A8 – ICA-AROMA + 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

1   Figure 6 also shows how each preprocessing choice progressively changes functional

2   connectivity estimates. When comparing pipelines that were identical but either included GSR

3   or not, correlations were 0.87, 0.86, 0.94 and 0.94 (for comparisons A1-A2, A3-A4, A5-A6, A7-

4   A8 respectively). When comparing pipelines that were identical but either included censoring or

5   not, correlations were 0.97, 0.96, 0.85, and 0.87 (for comparisons A1-A3, A2-A4, A5-A7, A6-A8

6   respectively). Individually these changes were small but comparing pipelines A1 and A5 to A4

7   and A8, respectively, i.e. neither GSR nor censoring, and with both GSR and censoring, the

8   correlation was only 0.81 for both comparisons. The choice to use ICA-AROMA vs. regressing

9   HMP had a larger effect, with correlations ranging from 0.74 to 0.83. Again, these correlations

10   decreased further if the pipelines also differed in use of GSR or censoring. Unsurprisingly,

11   preprocessing choices had a smaller effect on low motion scans (Figure 6b), with correlations

12   between pipelines ranging from 0.72 to 0.99. For high motion scans (Figure 6c), correlations

13   ranged from 0.61 to 0.93.

14

15   *3.6. Filtering comparison*

16   When a highpass filter was used rather than a bandpass filter, overall trends between

17   pipelines were quite similar (see Supplemental Figures 2-4). With a highpass filter, the number

18   of edges significantly correlated with motion ranged from 25-50% (Supplemental Figure 2a).

19   Other than for pipeline A5 (which fared poorly on this metric regardless), bandpass filtering

20   resulted in fewer edges significantly correlated with motion than highpass filtering, for example

21   20% vs 27% of edges for pipeline A4 (regress HMP+GSR+censor). Similarly, pipelines that used

22   bandpass filtering had a smaller mean absolute correlation with motion (Supplemental Figure

1   2b), for example z = 0.12 vs z = 0.14 for A4. Highpass filtering resulted in more negative edge

2   correlation with motion vs. edge distance slopes than bandpass filtering (e.g. -0.0017 vs -0.0008

3   for A4; Supplemental Figure 2c), suggesting a bandpass filter leads to reduced distance

4   dependent effects.

5   For fingerprinting, running the analysis with a highpass filter usually led to a lower

6   fingerprinting match rate (Supplemental Figure 3a), such as 65% vs 71% for pipeline A2 (regress

7   HMP+GSR+no censor). However, pipelines A3 (regress HMP+no GSR+censor) and A4 (regress

8   HMP+GSR+censor) were equally successful using either style of filter (89% match rate for A3,

9   90% for A4). Pipelines that used highpass filtering showed increased correlations to both non-

10   complement scans and to complement scans (Supplemental Figure 3b) (e.g. for A4: mean

11   complement correlation z = 0.87 for highpass, z = 0.80 for bandpass), suggesting a highpass

12   filter makes all scans' FC estimates more similar to each other, relative to bandpass. This

13   suggests that signals above 0.08 Hz, whether noise or connectivity information, are more

14   consistent across people and longitudinally within people. Filtering choice had a variable effect

15   on the fingerprinting margin (Supplemental Figure 3c). For some pipelines, using a highpass

16   filter led to a lower mean complement minus non-complement score (e.g. A2, z = 0.044 for

17   highpass, z = 0.067 for bandpass), but for pipelines A3 and A4 this margin of fingerprinting

18   success improved with a highpass filter (e.g. A4, z = 0.12 for highpass, z = 0.11 for bandpass).

19   Bandpass filtering increased ISCs for all pipelines compared to highpass filtering

20   (Supplemental Figure 4) (e.g. A4, z = 0.18 for highpass, z = 0.21 for bandpass), suggesting that

21   bandpass filtering is effective for removing noise while retaining task signal. However, at least

1   part of this improvement with bandpass filtering may be explained by signals composed of a

2   narrower range of frequencies being inherently more similar.

3

4   *3.7. Censoring comparison*

5   We repeated our initial analysis by varying the amount of censoring with a pipeline that

6   used a bandpass filter, regressed HMP and included GSR (Figure 6). When one volume was

7   censored per motion artifact, a lower threshold for censoring resulted in fewer edges being

8   correlated with motion, lowering from 21% of edges at a threshold of 0.30 mm, to 17% of edges

9   at a threshold of 0.15 mm (Figure 6a). However, the censoring threshold had minimal effect on

10   the number of edges significantly correlated with motion when 3 or 4 volumes were censored

11   per artifact, ranging from 13% to 15%. These effects were repeated in the mean absolute

12   correlation with motion (Figure 6b), with a lower threshold at one volume per artifact reducing

13   the correlation from $z = 0.119$ at 0.30 mm to $z = 0.109$ at 0.15 mm, but having minimal effect at

14   3 or 4 volumes (ranging from $z = 0.099$ mm to 0.105 mm).

15   The fingerprinting match-rate decreased as censoring became more stringent (Figure

16   6c), suggesting that although censoring can help with identifiability, information about

17   individuals is lost with more stringent approaches. At 1 volume censored per motion artifact,

18   the decrease was from 91% to 88%, but at 4 volumes censored per artifact the match-rate fell

19   from 88% to 79% when the threshold changed from 0.30 mm to 0.15 mm. With stricter

20   censoring, both the mean correlation to a complement scan and the mean correlation to a non-

21   complement scan decreased (Figure 6d), for example at 3 volumes per motion artifact the

22   mean complement-correlation dropped from $z = 0.784$ to $z = 0.731$ for a threshold of 0.30 mm
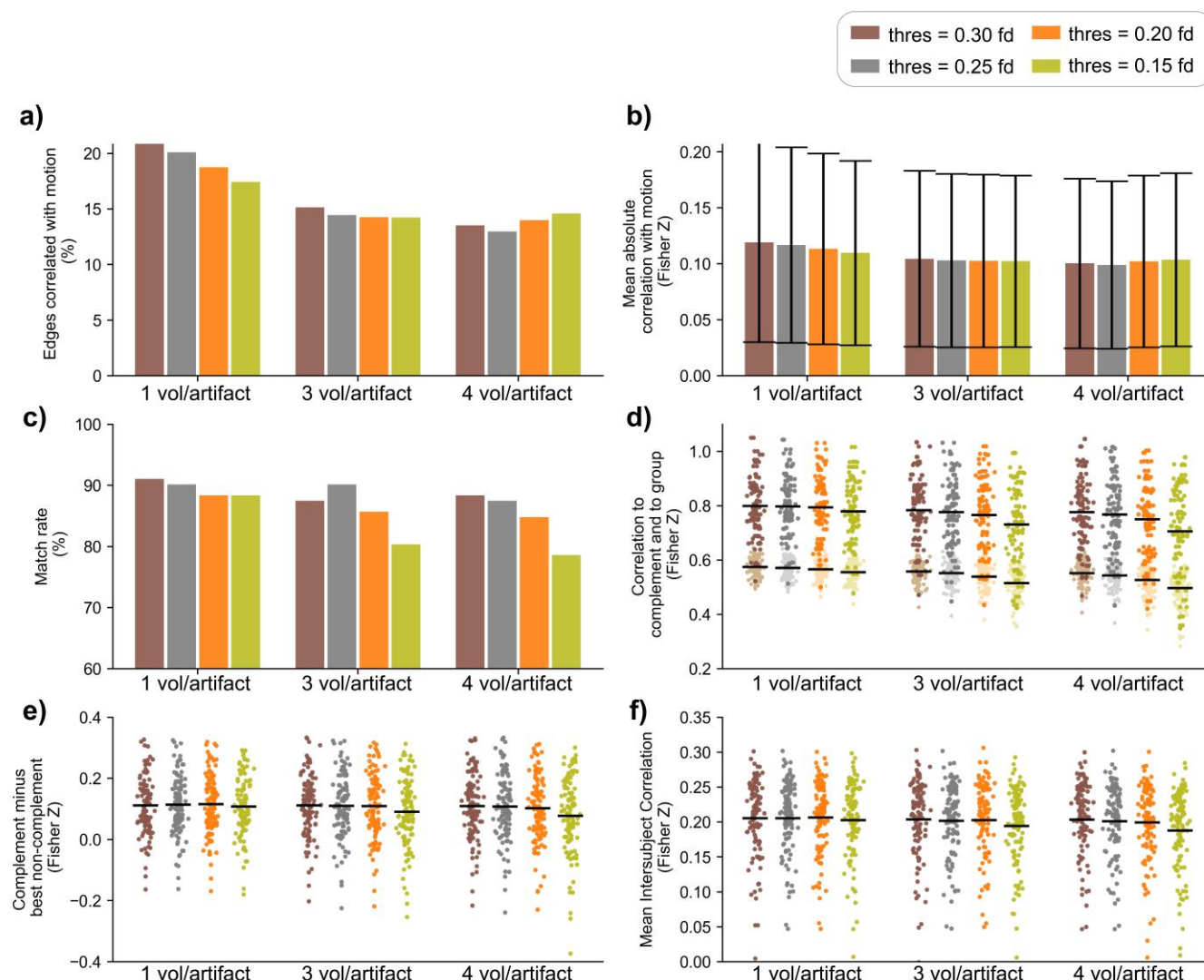
Figure 7. Pipeline benchmarks with different censoring thresholds. Colored bars depict four threshold levels and across the x-axis a different number of volumes is censored around the high motion frame. Plots are arranged from pipeline B1 through B12, as described in Table 2. **a)** Percentage of edges with a significant correlation between edge strength and subject motion across all 112 scans (uncorrected $p < 0.05$). **b)** Mean absolute correlation between edge strength and subject motion across all 112 scans. **c)** Fingerprinting match rate across pipelines. **d)** Each scan's correlation to their complement scan (darker points) and their average correlation to every other scan (lighter points), for each pipeline. Lines represent mean values. **e)** Difference between each scan's correlation to their complement and to their best (highest correlation) non-complement. Lines represent mean values. **f)** Mean ISCs for all 112 scans.

1   and 0.15 mm respectively. Surprisingly, the fingerprinting margin was largely unaffected by

2   censoring (Figure 6e), ranging from a mean of z = 0.102 to z = 0.116, except for the most

1   stringent censoring of 3 volumes at 0.15 mm and 4 volumes at 0.15 mm having a margin mean

2   of z = 0.091 and z = 0.077 respectively.

3   Censoring has a minimal effect on ISC (Figure 6f), except for the strictest censoring. At a

4   threshold of 0.15 mm, 3 volumes per motion artifact had a mean ISC of 0.194; 4 volumes per

5   motion artifact had a mean ISC of 0.188. All other censoring pipelines tested ranged from 0.199

6   to 0.207. It should be noted that for the purposes of ISC, censoring one scan necessarily censors

7   the scan it is being compared to, since the same time points need to be compared. A slightly

8   stricter censoring will lead to a much larger decrease in the number of shared time points, so

9   benefits in the removal of noise may be cancelled by the decrease in data being compared.

10

11   **4. Discussion**

12   Here we compared several preprocessing pipelines in relatively high motion longitudinal

13   data from young children. We found that preprocessing strategies that include GSR and

14   censoring together outperformed other pipelines on QC-FC metrics. We used connectome

15   fingerprinting to estimate how individually unique information was retained, or enhanced, by

16   preprocessing steps and found that volume censoring conferred more benefit than GSR, and

17   that the best pipelines regressed out HMPs rather than using ICA-AROMA. When examining ISC,

18   GSR offered the most improvement, with censoring showing a specific benefit for high motion

19   scans. Censoring multiple volumes per motion artifact improved QC-FC metrics but had minimal

20   or even a negative effect on other metrics. When examining overall effect of preprocessing

21   choices on connectomes, we found that pipelines had a major effect on individual estimates of

22   FC, which is a concern for cross-study replicability.

1   When comparing ICA-AROMA pipelines to pipelines that regressed HMPs, we found ICA-

2   AROMA to be less effective on most metrics. This is somewhat surprising, given that ICA-

3   AROMA primarily works by removing independent components strongly associated with motion

4   estimates (Pruim et al., 2015). Given that ICA-AROMA does reasonably well on QC-FC metrics,

5   which estimate noise-contaminated edges, but relatively poorly on fingerprinting metrics,

6   which estimate individual information, one interpretation is that ICA-AROMA is removing both

7   noise and signal of interest. However, we note that ICA-AROMA pipelines showed no obvious

8   difference on ISC metrics. Other implementations of ICA-based preprocessing exist and are

9   worth systematically investigating. ICA-AROMA is often used in conjunction with ICA-FIX,

10   another ICA-based denoising strategy (Salimi-Khorshidi et al., 2014), and ICA-AROMA can also

11   be run with less aggressive denoising. Some studies have combined ICA-based denoising with

12   regressing HMPs (Kaufmann et al., 2017; Jalbrzikowski et al., 2020), which may have

13   advantages, though effectively removes aspects of the signal related to motion twice, contrary

14   to the original intent of ICA-AROMA (Pruim et al., 2015).

15   While GSR remains contentious in the field, our results suggest there are advantages in

16   a high-motion dataset such as the early childhood sample used here. GSR reduced the number

17   of edges significantly correlated with motion, especially when applied with censoring. Similarly,

18   while there are distance-dependent effects when GSR is used, this effect was largely mitigated

19   in pipelines that also included censoring. GSR had no obvious impact on connectome

20   fingerprinting, both in terms of the match rate and the fingerprinting margin. This suggests that

21   while GSR effectively centers, and has distance-dependent effects, these changes are less edge-

22   specific impact than censoring. The finding that GSR had a minimal impact on fingerprinting,

1   our metric of individual information, may alleviate concerns about its use when comparing

2   healthy controls and clinical populations, or in developmental research, even in light of

3   research that suggests the global signal is significantly related to life outcomes and

4   psychological function (Li et al., 2019). We also found benefits to the use of GSR in improving

5   signal-to-noise in ISC metrics, suggesting that the benefits in terms of motion and physiological

6   noise removal may outweigh the cost of removing some signal of interest (Behzadi et al., 2007).

7   We found relatively small effects of a bandpass filter compared to a highpass filter.

8   Pipelines that used a bandpass filter fared better on both QC-FC metrics and ISC metrics. In

9   connectome fingerprinting, while complement correlations were higher in pipelines that used a

10   highpass filter, so too were non-complement correlations by comparable levels. This suggests

11   that any resulting changes to the functional connectome by including higher frequencies are

12   consistent across all scans and are unlikely to reflect individual-specific information.

13   We found benefits to volume censoring across metrics, but especially in connectome

14   fingerprinting, where censoring increased accuracy more than any of the other preprocessing

15   steps compared here. Censoring also conferred additional benefits to pipelines that used ICA-

16   AROMA, despite ICA-AROMA being intended partially as an alternative to censoring (Pruim et

17   al., 2015). Censoring had only a small effect on low-motion scans for ISC comparisons, but

18   censoring increased ISC scores on high-motion scans. When only one volume was censored per

19   motion artifact, the more stringent the censoring threshold, the fewer edges were significantly

20   correlated with motion. However, censoring multiple volumes per motion artifact

21   outperformed single volume censoring, suggesting that censoring more volumes per motion

22   artifact at a less conservative threshold may be preferable. Stricter censoring tended to

1   decrease correlations to both complement scans and non-complement scans. This suggests that

2   removing data points removes information that is common across participants. Interestingly,

3   the distribution of fingerprinting margin values was relatively unaffected by the censoring

4   threshold, except when very strict censoring was used, even though the overall match rate

5   decreased when censoring became stricter. This suggests not all scans are affected in the same

6   way by changes in censoring parameters.

7   Our findings on the intrascan interpipeline correlations are concerning towards the goal

8   of having replicable results in FC-MRI studies. While preprocessing choices have a smaller

9   impact on low motion data, our work suggests that when analyzing high motion data (which

10   data from early childhood tends to be), each preprocessing choice has a drastic effect on FC

11   estimates. While more broad methodological differences or group-level participant differences

12   are often cited as reasons for different findings between studies, we suggest that differences in

13   preprocessing also play a role. Preprocessing choices should be both closely considered and

14   accurately reported. While we compared the correlation between the entire functional

15   connectome, future work should explore regional specific effects in more detail. This can both

16   provide more insight into how preprocessing affects FC estimates, but also provide greater

17   clarification when different studies come to different conclusions.

18   There are several limitations to our work. Our study was done in young children who

19   were participating in a passive viewing task, and our results may not be directly applicable to

20   other populations or other study protocols. Likewise, the benefits of any step we tested may

21   not be transferable to a more traditional task-based fMRI study, especially censoring data

22   points during the task. We note that several of the steps implemented here could have been

1   applied in alternative forms, which also restricts generalizability. For example, we used a 24

2   HMP model rather than 6 or 12 HMP (Friston et al., 1996). Similarly, there are many other

3   thresholds that can be used for censoring or temporal filtering. There are also other possible

4   approaches to noise mitigation, such as CompCor (Behzadi et al., 2007) or ANATICOR (Jo et al.,

5   2010), that warrant further investigation.

6   We also acknowledge that choices such as parcellation may have impacted our findings.

7   Previous studies using functional connectome fingerprinting have used a similar number or

8   fewer nodes; for example, Finn et al (2015) used a 268-node atlas while Miranda-Dominguez et

9   al. (2018) used a 333-node atlas. Finn et al. (2015) also found lower fingerprinting success using

10   a 68-node atlas; it is unknown if a higher number of nodes may offset the difference between

11   preprocessing strategies on fingerprinting metrics. While we chose metrics that we believe are

12   meaningful to gauging the effect of preprocessing, other metrics could have instead been

13   chosen. For example, we did not consider the effect preprocessing has on the magnitude of FC

14   estimates within a scan, and we also did not consider the effect preprocessing might have on

15   individual networks within the broader connectome.

16   Here we implemented ISC by comparing parcels across scans, rather than a voxel-based

17   approach (Hasson et al., 2004). This has the potential to average out the effect of the task, as

18   not all voxels within a parcel will respond to the passive viewing task in a similar way. We also

19   chose to focus on the parcels with the highest average ISC values. These choices limit the extent

20   our findings can be generalized across the whole brain or to voxel-specific changes. Our

21   analysis has assumed that a higher ISC value reflects greater recovery of true signal, based on

22   the idea that unremoved noise in one or both scans will lower the temporal correlation

1   between scans. Parkes et al. (2018) suggest that FC test-retest reliability can increase in a high

2   noise dataset, suggesting that noise has a highly reproducible component within an individual.

3   It is unknown whether noise could have a similar effect between subjects in the context of ISC.

4   Future work should investigate more specifically the effect of preprocessing choices on ISC

5   values across the brain, especially in high noise scans.

6

5. Conclusions

8   Our findings have implications for future work in the field, both in pediatric studies but

9   also fMRI studies in general. While in different datasets different preprocessing choices may be

10 more appropriate, our results suggest that GSR, censoring, a bandpass filter, and HMP

11 regression are preferable in high motion datasets from pediatric populations engaging in a

12 passive viewing task. In particular, GSR and censoring showed few disadvantages across our

13 metrics, and ICA-AROMA showed no major improvement compared to regressing out HMPs,

14 especially when used without censoring.

15   All preprocessing choices have unintended effects on data; in light of the major effect

16 that preprocessing choices has on FC estimates within a single scan, we urge awareness of the

17 effect of preprocessing choices within any given research protocol. Ideally, studies should aim

18 to reduce head motion at the time of scan, and new preprocessing strategies or other advances

19 in fMRI protocols should aim to improve the signal-to-noise ratio with fewer tradeoffs.

1 **Funding**

2 This work was supported by: an Alberta Graduate Excellence Scholarship, and a Natural

3 Sciences and Engineering Research Council of Canada (NSERC) Create-We-Trac Training

4 Scholarship awarded to KG; and an NSERC Discovery Award, Canadian Institutes of Health

5 Research (CIHR) – Institute of Neurosciences, Mental Health and Addiction (INMHA) Bridge

6 Award and CIHR Project Grant to SB.

7

8 **CRediT authorship contribution statement**

9 **Kirk Graff:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing -

10 original draft, review & editing, Visualization, Funding acquisition. **Ryann Tansey:** Investigation,

11 Writing - review & editing. **Amanda Ip:** Resources, Investigation. **Christiane Rohr:** Investigation,

12 Writing - review & editing. **Dennis Dimond:** Investigation, Writing - review & editing. **Deborah**

13 **Dewey:** Methodology, Writing - review & editing. **Signe Bray:** Conceptualization, Investigation,

14 Methodology, Supervision, Project administration, Funding acquisition, Writing - review &

15 editing.

16

17 **Acknowledgments**

18 We thank all the children and their families who participated in this study, and the staff at the

19 Alberta Children's Hospital Imaging Centre. Thanks to Taylor Johnson for assistance with the

20 artwork.

21

22 **Disclosures**

1   The authors declare no conflict of interest.

Appendix. Supplementary materials
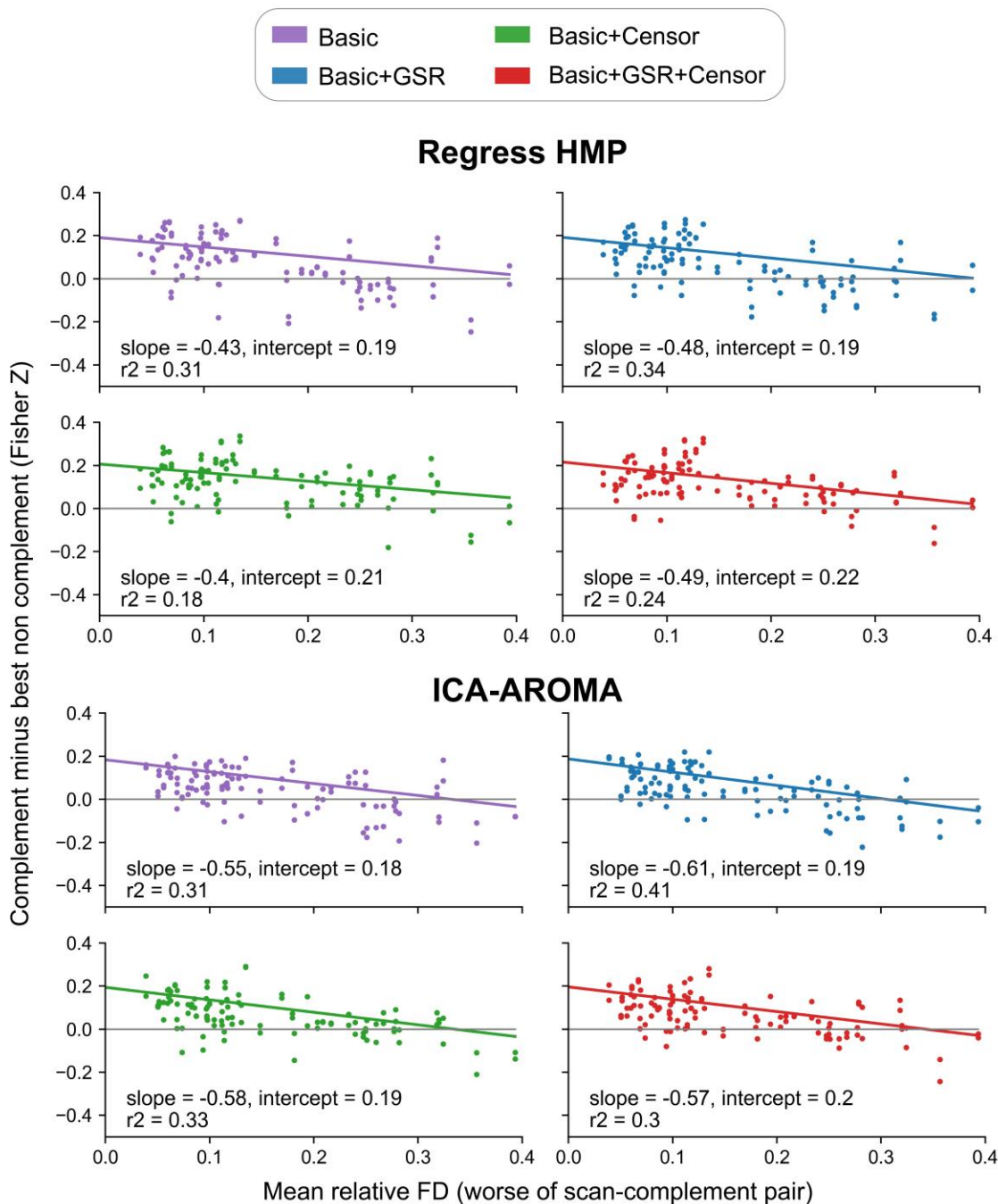
Supplemental Figure 1. Fingerprinting margin as a function of head motion. This is similar to Figure 4 with an outlier removed. The difference between each scan's correlation to their complement scan minus their correlation to the best (highest correlation) non-complement, plotted by motion. Motion was 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. Plots are arranged, left to right, from pipeline A1 through A8, as described in Table 1.
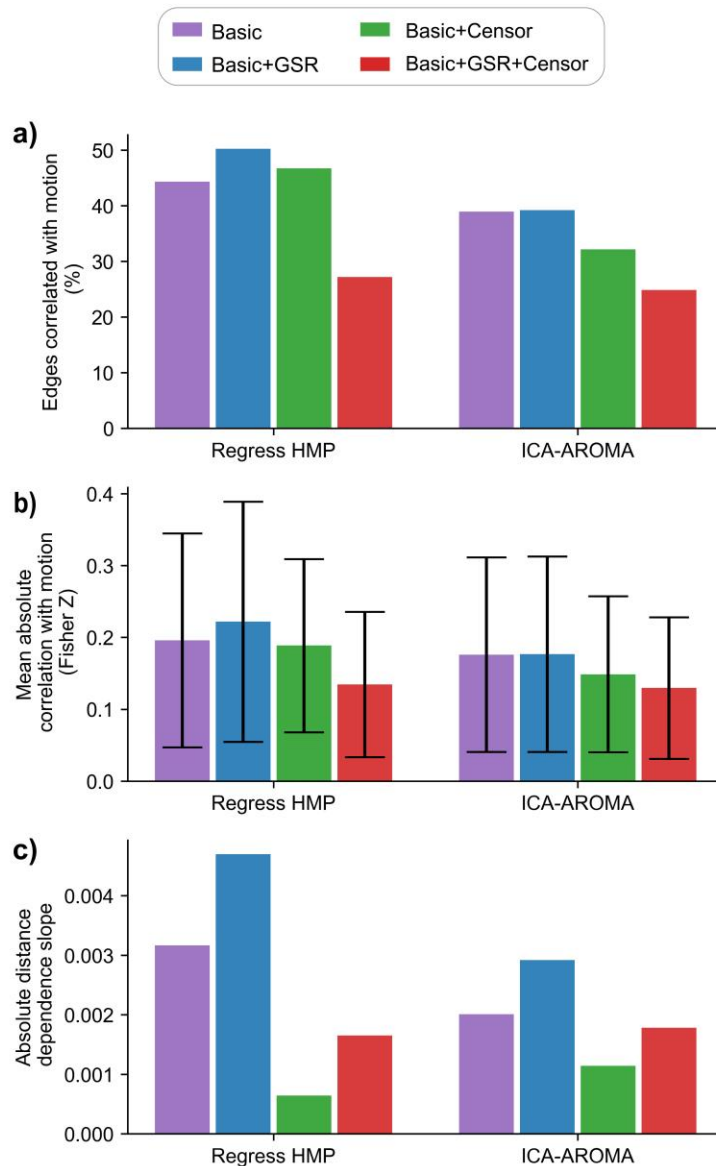
Supplemental Figure 2. **Quality control-functional connectivity (QC-FC) across pipelines with a highpass filter.** Plots are arranged from pipeline A1 through A8, as described in Table 1. **a)** Percentage of edges with a significant correlation between edge strength and subject motion across all 112 scans (uncorrected p < 0.05). **b)** Mean absolute correlation between edge strength and subject motion across all 112 scans. **c)** Absolute value of slope of each pipeline's distance-dependence plot.
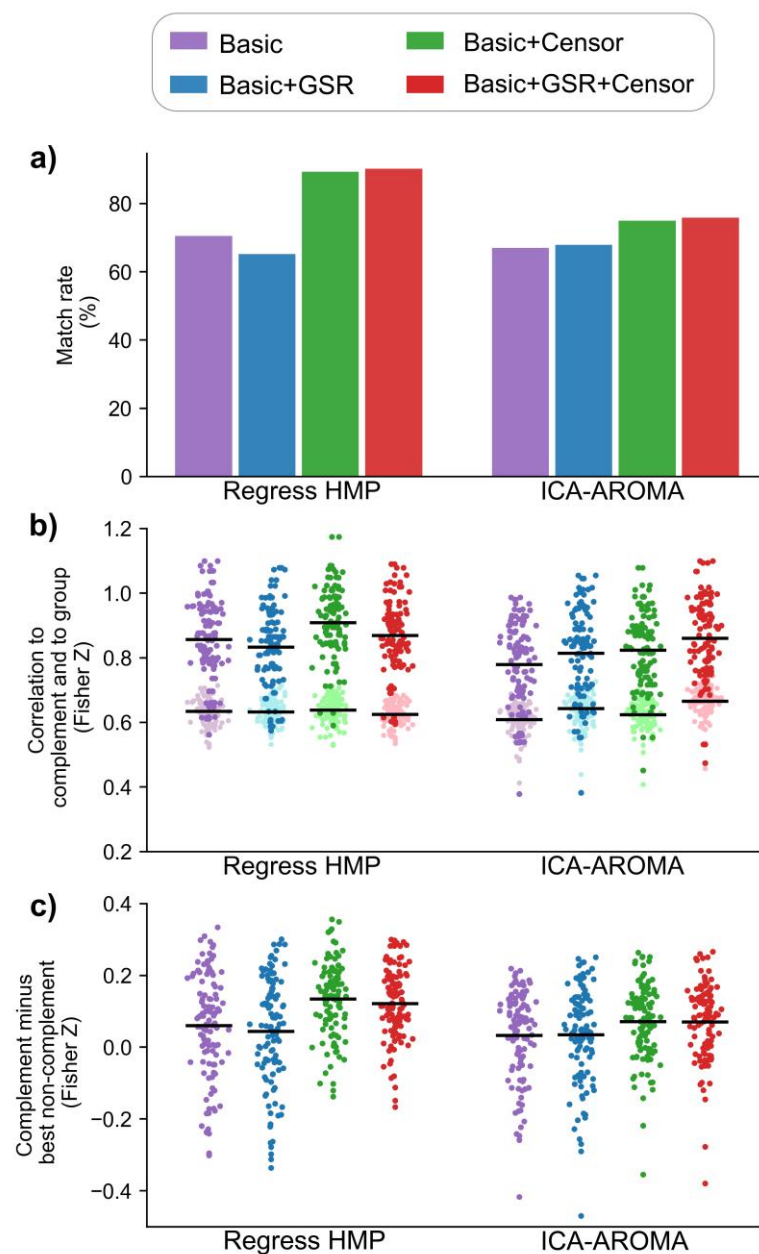
Supplemental Figure 3. Functional connectome fingerprinting by pipeline with a highpass filter. Plots are arranged from pipeline A1 through A8, as described in Table 1. **a)** Match rate across pipelines. A scan matched if its highest correlation was to its complement scan (other scan from the same individual). **b)** Each scan's correlation to their complement scan (darker points) and their average correlation to every other scan (lighter points), for each pipeline. Lines represent mean values. **c)** Difference between each scan's correlation to their complement and to their best (highest correlation) non-complement. Lines represent mean values.
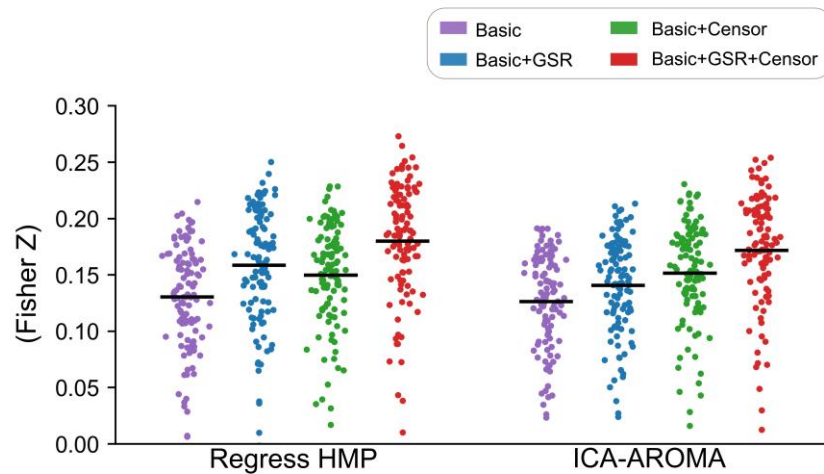
Supplemental Figure 4. Mean intersubject correlation (ISC) values with a highpass filter. Each point represents the mean ISC for a scan to all other scans, averaged across the 10 parcels with highest ISC. Lines represent mean values. Plot is arranged from pipeline A1 through A8, as described in Table 1.

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

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

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

Bolton, T.A.W., Kebets, V., Glerean., E., Zoller, D., Li., J., Yeo., B.T.T., Caballero-Gaudes., C., Van De Ville., D., 2020. *Agito ergo sum*: Correlates of spatio-temporal motion characteristics during fMRI. Neuroimage 209, 116433. https://doi.org/10.1016/j.neuroimage.2019.116433

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, 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

Byrge, L., Kennedy, D.P., 2020. Accurate prediction of individual subject identity and task, but not autism diagnosis, from functional connectomes. Hum. Brain Mapp. 41, 2249–2262. https://doi.org/10.1002/hbm.24943

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., 2020. 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., 2020. 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

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, 1664–71. 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, 346–55.

Gordon, E.M., Laumann, T.O., Gilmore, A.W., Newbold, D.J., Greene, D.J., Berg, J.J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., Hampton, J.M., Coalson, R.S., Nguyen, A.L., McDermott, K.B., Shimony, J.S., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., Nelson, S.M., Dosenbach, N.U.F., 2017. Precision Functional Mapping of Individual Human Brains. Neuron 95, 791-807.e7. https://doi.org/10.1016/j.neuron.2017.07.011

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 (80-.). 303, 1634–1640. https://doi.org/10.1126/science.1089506

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, 257–66. 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, 825–41.

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

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

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

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

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, in: Progress in Brain Research. pp. 259–276. https://doi.org/10.1016/B978-0-444-53839-0.00017-X

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

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, 2142–54. https://doi.org/10.1016/j.neuroimage.2011.10.018

Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage 84, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Power, J.D., Schlaggar, B.L., Petersen, S.E., 2015. Recent progress and outstanding issues in motion correction in resting state fMRI. Neuroimage 105, 536–51. 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

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, 623–32. 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, 240–56. 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

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, in: NeuroImage. Neuroimage. https://doi.org/10.1016/j.neuroimage.2004.07.051

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, 431–8. https://doi.org/10.1016/j.neuroimage.2011.07.044

Vanderwal, T., Eilbott, J., Finn, E.S., Craddock, R.C., Turnbull, A., Castellanos, F.X., 2017. Individual differences in functional connectivity during naturalistic viewing conditions. Neuroimage 157, 521–530. https://doi.org/10.1016/j.neuroimage.2017.06.027

Vanderwal, T., Eilbott, J., Castellanos, F.X., 2019. Movies in the magnet: Naturalistic paradigms in developmental functional neuroimaging. Developmental Cognitive Neuroscience 36, 100600. https://doi.org/10.1016/j.dcn.2018.10.004

Wang, J., Ren, Y., Hu, X., Nguyen, V.T., Guo, L., Han, J., Guo, C.C., 2017. Test–retest reliability of functional connectivity networks during naturalistic fMRI paradigms. Hum. Brain Mapp. 38, 2226–2241. https://doi.org/10.1002/hbm.23517