



## RESEARCH ARTICLE

WILEY

# Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data

Meichen Yu<sup>1</sup> | Kristin A. Linn<sup>1,2</sup> | Philip A. Cook<sup>1,3</sup> | Mary L. Phillips<sup>4</sup> |  
Melvin McInnis<sup>5</sup> | Maurizio Fava<sup>6</sup> | Madhukar H. Trivedi<sup>7</sup> |  
Myrna M. Weissman<sup>8,9,10</sup> | Russell T. Shinohara<sup>1,2</sup> | Yvette I. Sheline<sup>1,3,11</sup>

<sup>1</sup>Center for Neuromodulation in Depression and Stress, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

<sup>2</sup>Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

<sup>3</sup>Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

<sup>4</sup>Department of Psychiatry, University of Pittsburgh School of Medicine, Philadelphia, Pennsylvania

<sup>5</sup>Department of Psychiatry, University of Michigan School of Medicine, Ann Arbor, Michigan

<sup>6</sup>Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts

<sup>7</sup>Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, Texas

<sup>8</sup>Department of Psychiatry, Columbia University College of Physicians & Surgeons, New York, New York

<sup>9</sup>Division of Epidemiology, New York State Psychiatric Institute, New York, New York

<sup>10</sup>Mailman School of Public Health, Columbia University, New York, New York

<sup>11</sup>Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

## Correspondence

Yvette I. Sheline, MD, MS, Center for Neuromodulation in Depression and Stress, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

Email: sheline@mail.med.upenn.edu

## Funding information

National Institute of Neurological Disorders and Stroke, Grant/Award Numbers: R01 NS085211, R01 NS060910; National Institute of Mental Health, Grant/Award Number: R01 MH112847; National Multiple Sclerosis Society, Grant/Award Number: RG-1707-28586

## Abstract

Acquiring resting-state functional magnetic resonance imaging (fMRI) datasets at multiple MRI scanners and clinical sites can improve statistical power and generalizability of results. However, multi-site neuroimaging studies have reported considerable nonbiological variability in fMRI measurements due to different scanner manufacturers and acquisition protocols. These undesirable sources of variability may limit power to detect effects of interest and may even result in erroneous findings. Until now, there has not been an approach that removes unwanted site effects. In this study, using a relatively large multi-site (4 sites) fMRI dataset, we investigated the impact of site effects on functional connectivity and network measures estimated by widely used connectivity metrics and brain parcellations. The protocols and image acquisition of the dataset used in this study had been homogenized using identical MRI phantom acquisitions from each of the neuroimaging sites; however, intersite acquisition effects were not completely eliminated. Indeed, in this study, we found that the magnitude of site effects depended on the choice of connectivity metric and brain atlas. Therefore, to further remove site effects, we applied ComBat, a harmonization technique previously shown to eliminate site effects in multi-site diffusion tensor imaging (DTI) and cortical thickness studies. In the current work, ComBat successfully removed site effects identified in connectivity and network measures and increased the power to detect age associations when using optimal combinations of connectivity metrics and brain atlases. Our proposed ComBat harmonization approach for fMRI-derived connectivity measures facilitates reliable and efficient analysis of retrospective and prospective multi-site fMRI neuroimaging studies.

## KEYWORDS

aging, atlas, ComBat, fMRI, functional connectivity, graph theory, harmonization, multi-site, network efficiency

## 1 | INTRODUCTION

Functional magnetic resonance imaging (fMRI), a noninvasive neuroimaging modality with high spatial resolution, enables neural activity to be monitored. Functional connectivity and network measures derived from fMRI data have facilitated the study of the brain's function during development, in aging (Bressler & Menon, 2010; Fox & Raichle, 2007; Raichle 2015), and in the context of various neurological disorders (Bullmore & Sporns, 2009, 2012; Fornito et al., 2015; Fornito, Zalesky, & Bullmore, 2016; Stam, 2014).

Over the last decade, multi-site fMRI studies have become increasingly common (Biswal et al., 2010; Di Martino et al., 2014; Friedman et al., 2006, 2008; Gradin et al., 2010; Noble et al., 2017; Van Horn & Toga, 2009). Indeed, pooling fMRI data from multiple sites can accelerate participant recruitment rates and increase the total sample size of the study, thereby increasing statistical power. Pooling fMRI data is often critical when studying rare disorders and subtle effects and when aiming to generalize the study results to a diverse population (Dansereau et al., 2017; Keshavan et al., 2016; McGonigle, 2012; Suckling et al., 2010). Despite these advantages, multi-site studies are often plagued by nonbiological variability that can be attributed to differences in scanner manufacturers, nonstandardized imaging acquisition parameters, and other intrinsic factors (Shinohara et al., 2017). These additional sources of unwanted variability may decrease statistical power and lead to spurious results. Many multi-site studies have reported considerable site or scanner effects in fMRI data (Abraham et al., 2017; Brown et al., 2011; Dansereau et al., 2017; Feis et al., 2015; Friedman et al., 2006, 2008; Forsyth et al., 2014; Gountouna et al., 2010; Jovicich et al., 2016; McGonigle, 2012; Noble et al., 2017; Rath et al., 2016; Suckling et al., 2008, 2010; Turner et al., 2013; Van Horn & Toga, 2009). However, most of these studies only describe the problem or report the magnitude of site effects in fMRI measurements.

A few studies have attempted to mitigate site effects by standardizing protocols and image acquisition parameters (Chavez et al., 2018; Friedman et al., 2008; Glover et al., 2012; Kochunov et al., 2018; Oh et al., 2017; Shinohara et al., 2017). However, it has been shown that scanner-to-scanner variation arising from the use of scanners from different manufacturers is not eliminated completely by the standardization of acquisition parameters (Jovicich et al., 2016; Noble et al., 2017), for instance, by use of phantom-based imaging acquisitions (Delaparte et al., 2017). To our knowledge, until now, there has been only one attempt to diminish scanner differences in multi-site resting-state fMRI postacquisition. The authors used an independent component analysis (ICA) based approach that reduced differences across sites in some resting-state network connectivity measures but did not fully eliminate the structured noise arising from different scanners (Feis et al., 2015).

Recently, our group adapted ComBat harmonization (Johnson, Li, & Rabinovic, 2007) to model and remove site effects in multi-site DTI (Fortin et al., 2017) and cortical thickness (Fortin et al., 2018) measurements. ComBat was originally designed to correct so-called "batch effects" in genomic studies (Johnson et al., 2007) that arise due to processing high-throughput genomic data in different laboratories with different equipment at different times. In our previous studies, we

demonstrated that the ComBat harmonization technique successfully removed unwanted nonbiological variability, while preserving biological associations between participant age and DTI (fractional anisotropy and mean diffusivity) and the association between age and cortical thickness measurements.

In this study, we quantified the site effects in functional connectivity and several brain network measures in the multi-site Establishing Moderators and Biosignatures of Antidepressant Response in Clinical Care (EMBARC) dataset that was acquired at four clinical sites: Columbia University (CU), Massachusetts General Hospital (MGH), the University of Texas Southwestern Medical Center (TX), and the University of Michigan (UM). Our main objectives were to (1) remove any identified site effects using ComBat harmonization and (2) preserve the commonly reported negative correlation between age and functional connectivity within the default mode network (DMN; Damoiseaux et al., 2008; Damoiseaux, 2017; Ferreira & Busatto, 2013; Grady et al., 2010; Koch et al., 2010; Tomasi & Volkow, 2012), as well as preserve previously reported negative correlations between age and network efficiency measures (Achard & Bullmore, 2007; Ajilore, Lamar, & Kumar, 2014). Objective (2) was important to demonstrate that the ComBat technique did not remove important, biologically relevant information. A recently published multi-site autism study (Abraham et al., 2017) reported that the magnitude of site effects was influenced by the choice of functional connectivity metrics and brain parcellation. Therefore, we investigated the degree to which widely used functional connectivity and network metrics derived from a number of brain parcellations were affected by scanner-to-scanner variation and how ComBat harmonization performed in each setting. We hypothesized that (1) considerable site effects exist in both functional connectivity and network efficiency measures calculated from nonharmonized multi-site fMRI data; (2) the magnitude of site effects is not constant across different connectivity metrics and brain parcellations; and (3) ComBat harmonization can be used to remove site effects in connectivity and network measures while preserving age-related associations for numerous combinations of connectivity metrics and brain parcellations.

## 2 | MATERIALS AND METHODS

### 2.1 | Participants

This study considered 200 unmedicated depressed patients with major depressive disorder (MDD) and 40 healthy subjects recruited for EMBARC that have been analyzed in several previous studies (Fortin et al., 2018; Greenberg et al., 2015; Trivedi et al., 2016; Webb et al., 2016). This study concentrates on the harmonization of multi-site fMRI-based functional connectivity and network measures. The participants were recruited and scans acquired at four clinical sites: Columbia University (CU), Massachusetts General Hospital (MGH), the University of Texas Southwestern Medical Center (TX), and the University of Michigan (UM). All participants provided written informed consent and the institutional review boards from the four clinical sites approved all study procedures. For both patients and healthy individuals, the Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research

**TABLE 1** Imaging parameters for the four clinical sites: Columbia University (CU), Massachusetts General Hospital (MGH), the University of Texas Southwestern Medical Center (TX) and the University of Michigan (UM)

Scanner	CU General electric 3T	MGH Siemens 3T	TX Phillips 3T	UM Phillips 3T
<b>Structural</b>	FSPGR	MPRAGE	Turbo field echo (TFE)	MPRAGE
	TR = 6.0 ms	TR = 2,300 ms	TR = 8.2 ms	TR = 2,100 ms
	TE = 2.4 ms	TE = 2.54 ms	TE = 3.7 ms	TE = 3.7 ms
	TI = 900 ms	TI = 900 ms	TI = 1,100 ms	TI = 1,100 ms
	Flip angle = 9°	Flip angle = 9°	Flip angle = 12°	Flip angle = 12°
	FOV = 256 × 256 mm	FOV = 256 × 256 mm	FOV = 256 × 256 mm	FOV = 256 × 256 mm
	Slice thickness = 1 mm	Slice thickness = 1 mm	Slice thickness = 1 mm	Slice thickness = 1 mm
	Matrix = 256 × 256	Matrix = 256 × 256	Matrix = 256 × 256	Matrix = 256 × 256
	178 continuous slices (4 discarded)	176 continuous slices	178 continuous slices	178 continuous slices
<b>Functional</b>	TR/TE = 2,000/28 ms	TR/TE = 2,000/28 ms	TR/TE = 2,000/28 ms	TR/TE = 2,000/28 ms
	Flip angle 90°	Flip angle 90°	Flip angle 90°	Flip angle 90°
	FOV = 205 × 205 mm	FOV = 205 × 205 mm	FOV = 205 × 205 mm	FOV = 205 × 205 mm
	Slice thickness: 3.1 mm	Slice thickness: 3.1 mm	Slice thickness: 3.1 mm	Slice thickness: 3.1 mm
	Matrix 64 × 64	Matrix 64 × 64	Matrix 64 × 64	Matrix 64 × 64

Version, Patient Edition (SCID-I/P; First, Spitzer, & Gibbon, 2002) was used as inclusion criteria to diagnose the presence or absence of depressive symptoms. The Hamilton Depression Rating Scale (HAMD; Hamilton, 1960) and Quick Inventory for Depression Symptomatology (QIDS; Rush et al., 2003) depression scores were used to estimate depressive severity. Anxiety and depressive severity were also assessed using the Mood and Anxiety Symptom Questionnaire (MASQ; Watson & Clark, 1991), including three subscales: general distress (MASQ-GD), anhedonic depression (MASQ-AD), and anxious arousal (MASQ-AA). The individuals were eligible for the study if they met the following inclusion criteria: (1) age 18–65; (2) reported age of depression onset before age 30; and (3) fluent in English. Eleven depressed patients and one healthy individual were excluded due to excessive motion (>4 mm), low slice signal-to-noise ratio (<80), and severe slice artifacts in MRI data. The final sample included 189 MDD patients and 39 healthy individuals. The distribution of age, sex, handedness, and education level were matched between the two groups.

## 2.2 | Image acquisition and data preprocessing

All four sites used 3T scanners, however, the manufacturer differed from site to site: CU used a GE SIGNA HDx 3T scanner, MGH used a Siemens TIM Trio 3T scanner, TX used a Philips Achieva 3T scanner, and UM used a Philips Ingenia 3T scanner (Fortin et al., 2018; Greenberg et al., 2015; Trivedi et al., 2016). Imaging parameters at each site are described in Table 1.

Prior to the project's initiation, in close collaboration with MR physics teams at the acquisition sites, a homogenized imaging protocol was developed to minimize acquisition-related site differences. In

particular, data were collected using identical MRI phantom acquisitions from each of the neuroimaging sites. Well established routines for using phantoms were employed to perform quality assurance on the scanners used in this study. However, although the phantom-based approach minimized the inconsistency of signal-to-noise across scanners over the time and other variability in image acquisition and quality across sites, the intersite acquisition effects were not completely eliminated (Delaparte et al., 2017). Therefore, we employed a postprocessing procedure that further harmonized the fMRI functional connectivity matrices of subjects across the 4 sites.

T1-weighted (T1) images were processed using the ANTS Cortical Thickness pipeline available in the `antsCorticalThickness.sh` script in advanced normalization tools (ANTs; Avants et al., 2011a; Tustison et al., 2014). The workflow is sketched out as follows: (1) N4 bias correction to minimize field inhomogeneity (Tustison et al., 2010); (2) brain extraction using an optimal population-specific template created by a Symmetric Group Normalization framework (Avants et al., 2010); (3) Atropos probabilistic six-tissue segmentation (Avants, Tustison, Wu, Cook, & Gee, 2011b); (4) DiReCT-based cortical thickness estimation (Das, Avants, Grossman, & Gee, 2009); (5) SyN deformable spatial registration to the population-specific template (Klein et al., 2009).

Resting-state time series data from each participant were processed using the XCP Engine (Ciric et al., 2017), which uses an optimized confound regression procedure to reduce the influence of subject motion (Satterthwaite et al., 2017). Each subject contributed time series data from two resting-state fMRI sessions. The workflow of functional data preprocessing is summarized as follows: (1) removal of the four initial volumes of the Blood-oxygen-level Dependent (BOLD) signals to achieve signal stabilization; (2) realignment of functional

images using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002); (3) removal of nine confounding signals (six motion parameters + global/white matter/cerebral spinal fluid) as well as the temporal derivative, quadratic term, and temporal derivatives of each quadratic term (36 regressors total) (Satterthwaite et al., 2017); (4) co-registration of functional images to the T1 image using boundary-based registration (Greve & Fischl, 2009); (5) alignment of the co-registered images to template space using the ANTs-transform for the T1 image as above; and (6) temporal filtering of time series between 0.01 and 0.08 Hz as in previous studies (Biswal, Yetkin, Haughton, & Hyde, 1995) using a first-order Butterworth filter. In this study, all regressors, including motion parameters and confound time courses, were band-pass filtered to the same frequency range as the time series data to prevent frequency-dependent mismatch during confound regression (Hallquist, Hwang, & Luna, 2013). Functional images were smoothed using a Gaussian convolution at 6 mm full-width at half-maximum.

### 2.3 | Parcellation

To investigate the influence of different parcellations on functional connectivity measures across sites and subsequent harmonization, we partitioned the brain of each participant into cortical and subcortical ROIs using the following three different whole-brain atlases (one anatomical and two functional): (1) 78 cortical and 12 subcortical ROIs identified by automated anatomical labeling (AAL) (Tzourio-Mazoyer et al., 2002); (2) 264 cortical and subcortical ROIs of the widely-used functional Power atlas (Power et al., 2011); and (3) 333 cortical and subcortical ROIs from the functional Gordon atlas (Gordon et al., 2016). The ROIs and MNI space centroids of the AAL, Power, and Gordon atlases can be found in Supporting Information, Figure S1 and Material 2.

### 2.4 | Functional connectivity

For each participant, whole-brain functional connectivity between all brain regions was constructed pairwise from the preprocessed fMRI data. The fMRI time series were extracted from each voxel and averaged within each ROI of the three atlases (AAL, Power, and Gordon). The functional connectivity between time series for all pair-wise ROIs was estimated by calculating two commonly used connectivity metrics: Pearson correlation and wavelet coherence. For Pearson correlation, the correlation coefficients were Fisher-transformed to draw more statistically interpretable conclusions about the magnitude of the correlations (Cohen & D'Esposito, 2016; Doucet, Bassett, Yao, Glahn, & Frangou, 2017). Due to poor signal quality and signal dropout, we excluded 61 ROIs from the Power atlas and 26 ROIs from the Gordon atlas, which resulted in 203 and 307 ROIs for the Power and Gordon atlases, respectively. All subsequent analyses were performed using the  $90 \times 90$  AAL-atlas,  $203 \times 203$  Power-atlas, and  $307 \times 307$  Gordon-atlas connectivity matrices based on both Fisher-transformed Pearson correlation coefficients and raw wavelet coherence values from all participants.

### 2.5 | Model for functional connectivity matrix harmonization

Based on the literature (Dansereau et al., 2017; Friedman et al., 2008; Feis et al., 2015; Rath et al., 2016), we speculated that measurements such as DTI fractional anisotropy (Fortin et al., 2017), MRI cortical thickness (Fortin et al., 2018), and fMRI functional connectivity (the present study) would differ among the four sites (CU, MGH, TX, and UM) due to systematic bias and nonbiological variability attributable to the use of different scanners and different imaging parameters.

In this study, we used ComBat (Johnson et al., 2007) to reduce potential biases and non-biological variability induced by site and scanner effects. ComBat uses a multivariate linear mixed effects regression with terms for biological variables and scanner to model imaging feature measurements. The method uses empirical Bayes to improve the estimation of the model parameters for studies with small sample sizes. Here, we reformulate the ComBat model, so that it can be applied to functional connectivity matrices estimated using Pearson correlation and wavelet coherence in combination with the AAL, Power, and Gordon atlases (i.e., six combinations: Correlation-AAL, Coherence-AAL, Correlation-Power, Coherence-Power, Correlation-Gordon, and Coherence-Gordon). As all connectivity matrices are symmetric, we applied ComBat to connectivity values in the upper triangles of the matrices. Let  $y_{ijv}$  represents the connectivity values of imaging site  $i$  ( $i \in \{1, \dots, 4\}$ ), participant  $j$  ( $j \in \{1, \dots, 228\}$ ), and connectivity value  $v$  ( $v \in \{1, \dots, 4, 005\}$  for the AAL atlas,  $v \in \{1, \dots, 20, 503\}$  for the Power atlas, and  $v \in \{1, \dots, 46, 971\}$  for the Gordon atlas) between two ROIs. Then, the ComBat model can be written as

$$y_{ijv} = \alpha_v + X_{ij}^T \beta_v + \gamma_{iv} + \delta_{iv} \epsilon_{ijv}$$

where  $\alpha_v$  is the average connectivity value for a particular connectivity value  $v$  between two ROIs,  $X_{ij}^T$  is a design matrix for the covariates of interest (age, gender, and group), and  $\beta_v$  is a vector of regression coefficients corresponding to  $X$ . As in Fortin et al. (2018), we further assume that the residual terms  $\epsilon_{ijv}$  arise from a normal distribution with zero mean and variance  $\sigma_v^2$ . The terms  $\gamma_{iv}$  and  $\delta_{iv}$  represent the additive (or location parameter) and multiplicative (or scale parameter) site effects of site  $i$  for connectivity value  $v$ , respectively. The ComBat-harmonized functional connectivity values were then defined as

$$y_{ijv}^{\text{ComBat}} = \frac{y_{ijv} - \widehat{\alpha}_v - X_{ij}^T \widehat{\beta}_v - \gamma_{iv}^* + \widehat{\alpha}_v + X_{ij}^T \widehat{\beta}_v}{\delta_{iv}^*}$$

where  $\gamma_{iv}^*$  and  $\delta_{iv}^*$  are the empirical Bayes estimates of  $\gamma_{iv}$  and  $\delta_{iv}$ , respectively. Thus, ComBat simultaneously models and estimates biological and nonbiological terms and algebraically removes the estimated additive and multiplicative site effects. Of note, in the ComBat model, we included age, sex, and group as covariates to preserve important biological trends in the data and avoid overcorrection.

In this study, we performed the ComBat harmonization analyses for the six combinations of connectivity matrices in two sessions (S1 and S2), separately. ComBat harmonization analyses were performed using a publicly available MATLAB package hosted at <https://github.com/Jfortin1/ComBatHarmonization/tree/master/Matlab>.

**TABLE 2** Site effects in functional connectivity values for different connectivity measures (Pearson correlation and wavelet coherence) using different atlases (AAL, Power, and Gordon)

Six combinations of two connectivity measures and three atlases		Number of ROIs	Original	ComBat
Pearson correlation	AAL	90	1610 (40.2%)	0 (0%)
	Power	203	515 (2.5%)	0 (0%)
	Gordon	307	603 (1.3%)	0 (0%)
Wavelet coherence	AAL	90	7 (0.2%)	0 (0%)
	Power	203	12 (0.06%)	0 (0%)
	Gordon	307	17 (0.04%)	0 (0%)

Note that the numbers and percentages in the round brackets represent the numbers and percentages of connectivity values that were significantly different across 4 sites for six combinations of two connectivity measures and three atlases, respectively. Significant effects for each combination were computed using Kruskal–Wallis tests with FDR corrections. Here we only presented the results of the first-session data, as the two sessions showed extremely similar results.

## 2.6 | Visualization and evaluation of functional connectivity harmonization

We used Kruskal–Wallis tests to quantify the magnitude of site effects in functional connectivity between all pairwise ROIs before and after applying ComBat harmonization to each of the six metric-atlas combinations (Correlation-AAL, Coherence-AAL, Correlation-Power, Coherence-Power, Correlation-Gordon, and Coherence-Gordon). The  $p$  values were adjusted for multiple comparisons by controlling the false discovery rate (FDR) (Benjamini & Hochberg, 1995) at 5%, separately for each combination (AAL: 4,005 comparisons; Power: 20,503 comparisons; Gordon: 46,971 comparisons). The numbers and percentages of connectivity values that were significantly different (after FDR correction) across the 4 sites for the six combinations are summarized in Table 2. The FDR-corrected  $p$  values can be found in Supporting Information, Figure 1. We visualized site effects using boxplots of connectivity values between signals of two randomly selected ROIs for each atlas across the four sites (Figure 2 and Supporting Information, Figures S3–S6, subplot A). The selected ROIs were consistent for the same atlas using the two connectivity measures.

We also performed a principal component analysis (PCA) on the functional connectivity values in the upper triangle of the connectivity matrices for the six metric-atlas combinations before and after ComBat harmonization. In Supporting Information, Figures S3A–S6A, subplot A and S7, we plotted three-dimensional scatter plots of the first three PC scores from the PCA. If the connectivity values were significantly different (Kruskal–Wallis tests; FDR corrections) across sites before or after ComBat harmonization, the corresponding PC scores are likely to be associated with site, and we would expect to see data from the same scanner roughly clustered together in the scatter plots.

To evaluate whether the assumed empirical Bayes priors for the location ( $\gamma$ ) and scale ( $\delta$ ) parameters in the ComBat harmonization model reasonably reflect the observed data, we overlaid the empirical

and prior distributions of  $\gamma$  and  $\delta$  in Figure 3 and Supporting Information, Figures S3–S6, subplot C.

We applied ComBat harmonization to the connectivity matrices from each fMRI session, separately. We present visualizations of the site effects and plots of ComBat model parameters for the first session. Plots generated from the second session were similar and therefore not included. Figure 2 demonstrates differences in the distribution of functional connectivity across sites for the first session. Figure 3 provides a visualization of the goodness of fit of the ComBat model's prior assumptions to the observed data for the first session. Following ComBat, we extracted four network measures from the harmonized connectivity matrices and averaged these measures across the two sessions. Henceforth, we focus on analyzing the average network measures, which included weighted DMN connectivity, nodal strength, local efficiency, and global efficiency. We formally define these measures in Section 2.7.

## 2.7 | Calculation of network measures

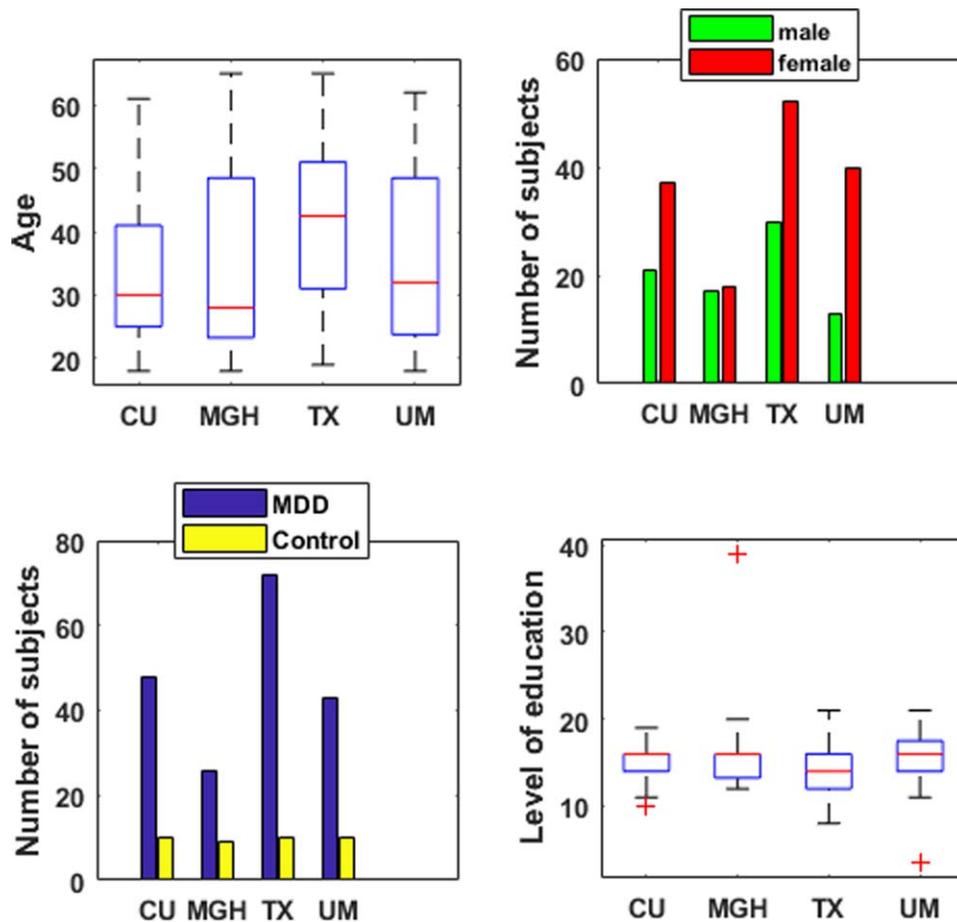
To ensure that our postprocessing harmonization did not remove meaningful biological variability along with the undesirable site effects, we conducted an additional analysis. As the default mode network (DMN) has been found to have larger negative associations between age and functional connectivity metrics than other resting-state networks (Damoiseaux, 2017; Ferreira & Busatto, 2013; Tomasi & Volkow, 2012), we selected it to conduct analysis of age-related effects. In this study, functional connectivity and local network metrics (quantified by weighted nodal strength and nodal efficiency) were thus calculated in the DMN. Global network topology was characterized by weighted global efficiency. The computation details of these connectivity and network metrics are described in the following paragraphs.

For each atlas, the DMN network connectivity was defined by the summation of the functional connectivity values within the DMN ROIs normalized by the number of DMN ROIs for each atlas. The DMN nodal strength was computed by first summing the functional connectivity values (link weights) for each pair of ROIs and then summing up the nodal level connectivity values within the DMN ROIs of each specific atlas. The weighted local and global efficiency (Latora & Marchiori, 2001) were computed using the weighted shortest path length ( $L^W$ ; Dijkstra, 1959), which is the shortest sum of connection length (inverse of the connectivity values or link weights) between two nodes (or ROIs; Rubinov & Sporns, 2010). The weighted nodal efficiency ( $E_{\text{nodal}}^W$ ) was calculated as the inverse of the harmonic mean of  $L^W$  from one node to all other nodes, as follows:

$$E_{\text{nodal}}^W = \frac{1}{N-1} \sum_{j \in G} \frac{1}{L_{ij}^W}$$

where  $N$  is the number of nodes in graph  $G$  (represented by the AAL, Power, and Gordon connectivity matrices in this study) and  $L_{ij}^W$  is the weighted shortest path length between node  $i$  and  $j$ .

Weighted local efficiency ( $E_{\text{local}}^W$ ) for a node is defined as the average weighted nodal efficiency among the neighboring nodes of that node (excluding the reference node), as follows:



**FIGURE 1** Distribution of subject demographic characteristics across 4 sites. Abbreviations: MDD = major depressive disorder; control = healthy controls; MGH = Massachusetts General Hospital; TX = University of Texas Southwestern Medical Center; UM = University of Michigan. Note that whiskers in the boxplots represent variability outside the upper and lower quartiles [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

$$E_{\text{local}}^W = \frac{1}{N_{G_i}(N_{G_i}-1)} \sum_{j,k \in G_i} L_{j,k}^W$$

where  $N_{G_i}$  is the number of nodes in subgraph  $G_i$  that consists of all neighboring nodes of node  $i$ , but excluding node  $i$ . For the weighted DMN local efficiency, weighted local efficiency values were computed for each ROI and then summed up within the DMN ROIs of each specific atlas.

The weighted global efficiency ( $E_{\text{global}}^W$ ) was calculated as the average weighted nodal efficiency of nodes in a graph  $G$ , as follows:

$$E_{\text{global}}^W = \frac{1}{N(N-1)} \sum_{i \neq j \in G} L_{i,j}^W$$

All the network efficiency measures were computed using the Brain Connectivity Toolbox (BCT) (Rubinov & Sporns, 2010).

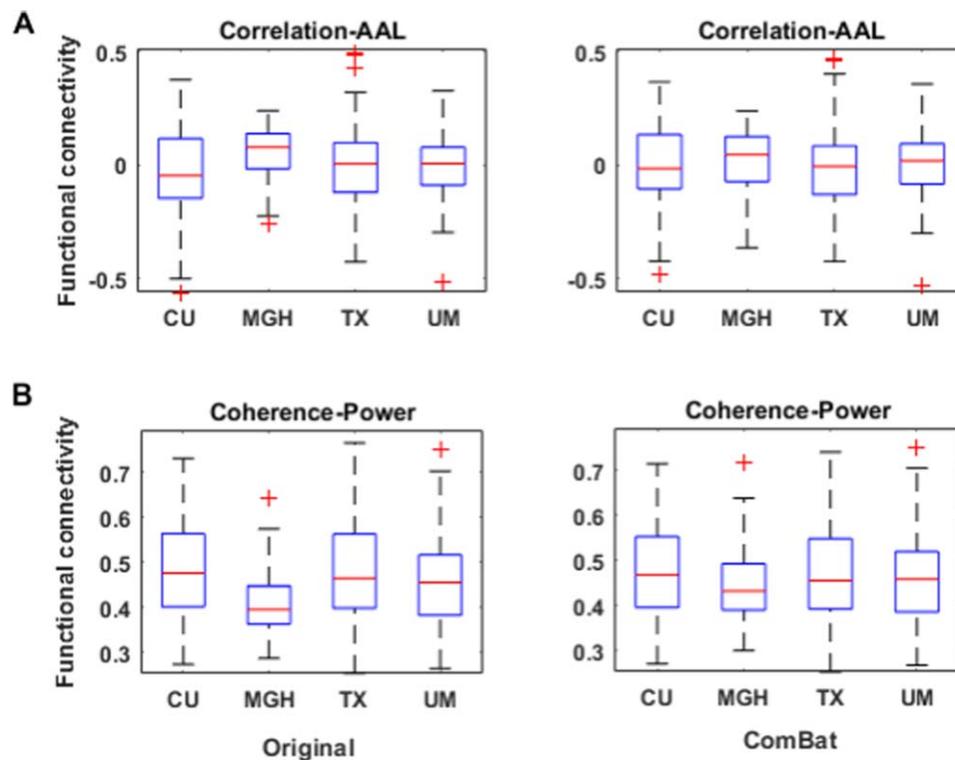
For each participant, we first computed DMN network connectivity, DMN nodal strength, weighted DMN local efficiency, and weighted global efficiency for each of the six combinations (3 atlases  $\times$  2 connectivity metrics) before and after applying ComBat harmonization. Next, we averaged the values of each participant's network connectivity or efficiency measures from the two sessions. Then, we tested the global null hypothesis of no differences across sites in the network

connectivity or efficiency measures using Kruskal–Wallis tests (in total, 2 conditions (before and after ComBat)  $\times$  2 connectivity metrics  $\times$  4 network measures  $\times$  3 atlases = 48 comparisons) with a separate FDR correction at 5% within each condition (2 connectivity metrics  $\times$  4 network measures  $\times$  3 atlases = 24 comparisons), separately.

## 2.8 | Preservation of biological variability

An optimal harmonization technique should be able to remove most or all non-biological sources of variability caused by site and scanner, yet preserve or increase statistical power to detect biological associations. In this study, there was a broad participant age range (18–65 years), enabling investigation of age-related associations. Therefore, we investigated whether negative associations between age and DMN network connectivity and associations between age and network efficiency measures were preserved or made stronger when estimated using ComBat-harmonized data.

We computed the Spearman correlation between each network (or connectivity) measure and age. The  $p$  values were adjusted for multiple comparisons (in total, 2 conditions (before and after ComBat)  $\times$  2 connectivity metrics  $\times$  4 network measures  $\times$  3 atlases = 48



**FIGURE 2** Site effects in functional connectivity estimated by “Correlation-AAL” (a) and “Coherence-Power” (b) before and after ComBat harmonization. Note that the functional connectivity values in (a) and (b) were computed from the time series of two randomly selected ROIs: for AAL, the TPOmid.R and ACG.R were selected; for Power, two regions in the visual cortex were selected; whiskers in the boxplots represent variability outside the upper and lower quartiles. Abbreviations: CU = Columbia University; MGH = Massachusetts General Hospital; TX = University of Texas Southwestern Medical Center; UM = University of Michigan; TPOmid.R = right temporal pole; middle temporal gyrus; ACG.R = right anterior cingulate and paracingulate gyri [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

comparisons) by controlling the false discovery rate (FDR; Benjamini & Hochberg, 1995). As before, the FDR corrections were applied separately within condition (2 connectivity metrics  $\times$  4 network measures  $\times$  3 atlases = 24 comparisons). A significance level of  $p < .05$  was used for these tests. Note that for the Power and Gordon atlases, we used the original definitions of DMN ROIs from Power et al. (2011) and Gordon et al. (2016), respectively; for the AAL atlas, we defined DMN ROIs according to a review article, Rosazza and Minati (2011). For details of the definition of the DMN ROIs for each atlas, refer to Supporting Information, Table 1, Figure S1, and Materials 2. Supporting Information, Figure S1, subplots b and c were visualized with BrainNet Viewer (version 1.5, Xia, Wang, & He, 2013; <http://www.nitrc.org/projects/bnv/>).

## 2.9 | Statistical analysis of demographic characteristics

Statistical analyses for demographic characteristics of participants were performed using MATLAB (R2017a). Age and educational level were compared among the four sites using Kruskal–Wallis tests followed by Mann–Whitney  $U$  tests when appropriate. All  $p$  values from the Mann–Whitney  $U$  tests were adjusted for multiple comparisons by controlling the false discovery rate (FDR) (Benjamini & Hochberg, 1995) at 5%. We tested for differences in the gender and clinical group distribution among the four sites using Pearson’s chi-squared ( $\chi^2$ ) tests.

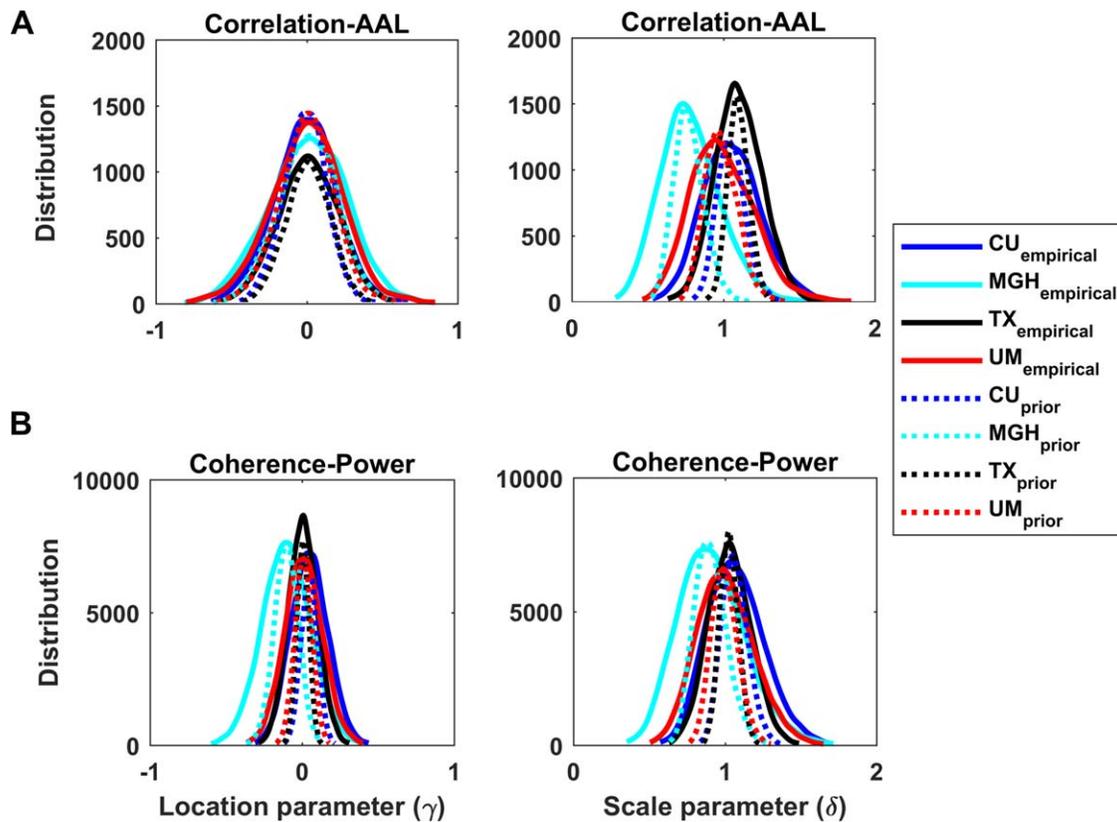
## 3 | RESULTS

### 3.1 | Demographic characteristics

The distribution of demographic characteristics across imaging sites is shown in Figure 1. The age distribution ( $p = .001$ ) was imbalanced across sites; subjects in the TX site were older than the other sites (TX–CU:  $p < .001$ ; TX–MGH:  $p = .04$ ; TX–UM:  $p = .04$ ; FDR correction). There were also weak site effects ( $p = .03$ ) in the level of education, but no differences were found regarding educational level between pairwise sites after FDR correction. Gender ( $p = .14$ ) and depressed/control ( $p = .34$ ) distributions were equally distributed across sites.

### 3.2 | Visualization and evaluation of ComBat harmonization

Functional connectivity values estimated by Pearson correlation showed much stronger site effects than those by wavelet coherence for analyses using the AAL and Power atlases as well as the Gordon atlas (Table 2). Moreover, the AAL atlas had a much larger percentage of connectivity values that differed significantly across the four sites than the Power and Gordon atlases (Table 2). Following ComBat harmonization, there were no statistically significant site effects in the functional connectivity values of the six metric–atlas combinations.



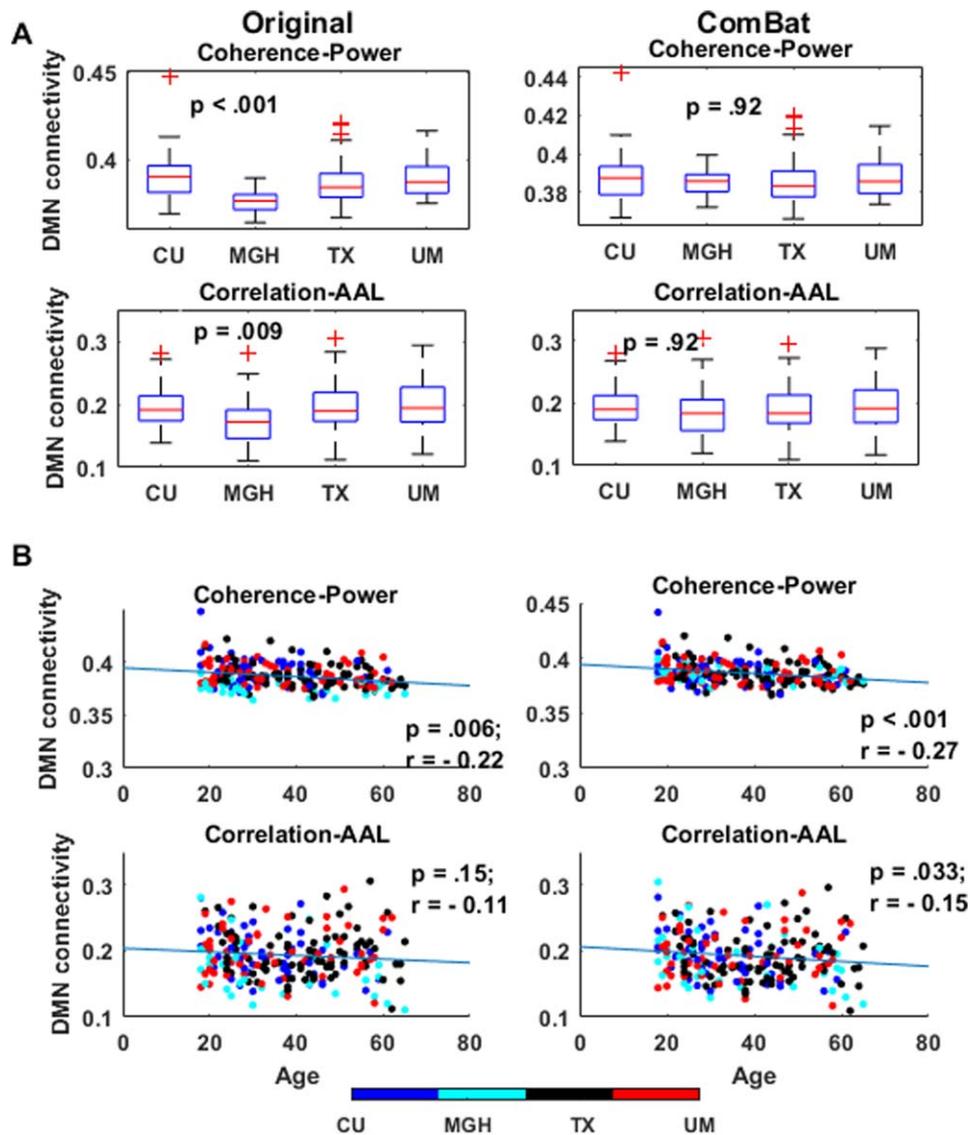
**FIGURE 3** Empirical (dashed lines) and ComBat-estimated (solid lines) prior distributions for the site-specific location ( $\gamma$ ) and scale ( $\delta$ ) parameters for “Correlation-AAL” (a) and “Coherence-Power” (b). Abbreviations: CU = Columbia University; MGH = Massachusetts General Hospital; TX = University of Texas Southwestern Medical Center; UM = University of Michigan

Figure 2 displays the boxplots of functional connectivity values between two randomly selected ROIs (for AAL: TPOmid.R and ACG.R; for Power: two regions in the visual cortex) for the Pearson correlation-AAL atlas and Wavelet coherence-Power atlas combinations. The connectivity values between these two ROIs showed consistent patterns across the four sites: for the AAL atlas, MGH values were generally higher than the other three sites; for the Power atlas, MGH values were generally lower than the other three sites. For the Gordon atlas, MGH and UM values were generally lower than the other two sites (see Supporting Information, Figures S5 and S6, subplot A, for the Gordon atlas results). After ComBat harmonization, these significant site effects were dramatically reduced in all six metric-atlas combinations. See Supporting Information, Figures S3–S6, subplot A for boxplots of functional connectivity for other metric-atlas combinations, before and after ComBat.

To further visualize site effects, we generated three-dimensional scatter plots of the first three principal component (PC) scores obtained from the functional connectivity matrices (Supporting Information, Figures S3–S6, subplot B and S7). For all three atlases, the second PC scores from CU and/or MGH patients showed distinct separation from those of TX and UM, particularly when using Pearson correlation. These visual site effects are much less noticeable in the ComBat-harmonized data: the first three PC scores were not clearly associated with site by visual inspection for any of the metric-atlas combinations (Supporting Information, Figures S3–S6, subplot B and S7).

For all metric-atlas combinations, the ComBat-harmonized prior distributions appear to fit the empirical distributions of both the location ( $\gamma$ ) and scale ( $\delta$ ) parameters well (Figure 3 and Supporting Information, Figures S3–S6, subplot C). Visual inspection of these overlaid distributions suggests that the ComBat model used appropriate prior information to capture the underlying site effects in the functional connectivity matrices. Furthermore, for each of the three atlases, the distributions of  $\gamma$  and  $\delta$  reflected the observed lower magnitudes in the distribution location and variability of MGH values compared with the values from other three sites, that is, for MGH,  $\delta < 1$  on average using Pearson correlation and  $\delta < 1$  and  $\gamma < 0$  using wavelet coherence.

Before performing ComBat harmonization on the functional connectivity matrices, all the network connectivity and efficiency measures estimated by Pearson correlation and a majority of the measures estimated by wavelet coherence displayed statistically significant site effects with similar patterns across sites (MGH values were generally lower than all the other sites). Figure 5a displays log-transformed  $p$  values from the global site effect tests for each metric-atlas combination before and after ComBat. The  $p$  values for network connectivity and the efficiency measures estimated by Pearson correlation were considerably more significant than those estimated by wavelet coherence (see Supporting Information, Figures S8–S10, subplot A, for corresponding boxplot visualizations of the site effects). After ComBat harmonization, there were no remaining statistically significant site effects for any metric-atlas combination.



**FIGURE 4** DMN connectivity estimated by “Coherence-Power” and “Correlation-AAL” across sites (a) and their anti-correlations with age (b). Note that DMN connectivity was computed by first summing the functional connectivity values within the DMN ROIs, and then normalizing by the number of DMN ROIs corresponding to each atlas (Power and AAL); the significant site effects in DMN connectivity were removed by ComBat harmonization (a). The strong anti-correlation between age and DMN connectivity estimated by “Coherence-Power” was preserved in ComBat-harmonized data; for “Correlation-AAL,” ComBat harmonization increased the detection power of the anti-correlation (b). Note that whiskers in the boxplots represent variability outside the upper and lower quartiles. Abbreviations: DMN = default mode network

Prior to functional connectivity harmonization using ComBat, there were statistically significant site effects across all network connectivity and efficiency measures estimated by both Pearson correlation and wavelet coherence when using the Gordon atlas (Figure 5a). In contrast, prior to harmonization using ComBat, when using the AAL or Power atlases, a number of the measures estimated by wavelet coherence did not display significant differences across sites (Figure 5a and Supporting Information, Figures S9 and S10, subplot A).

### 3.3 | Preservation of biological variability

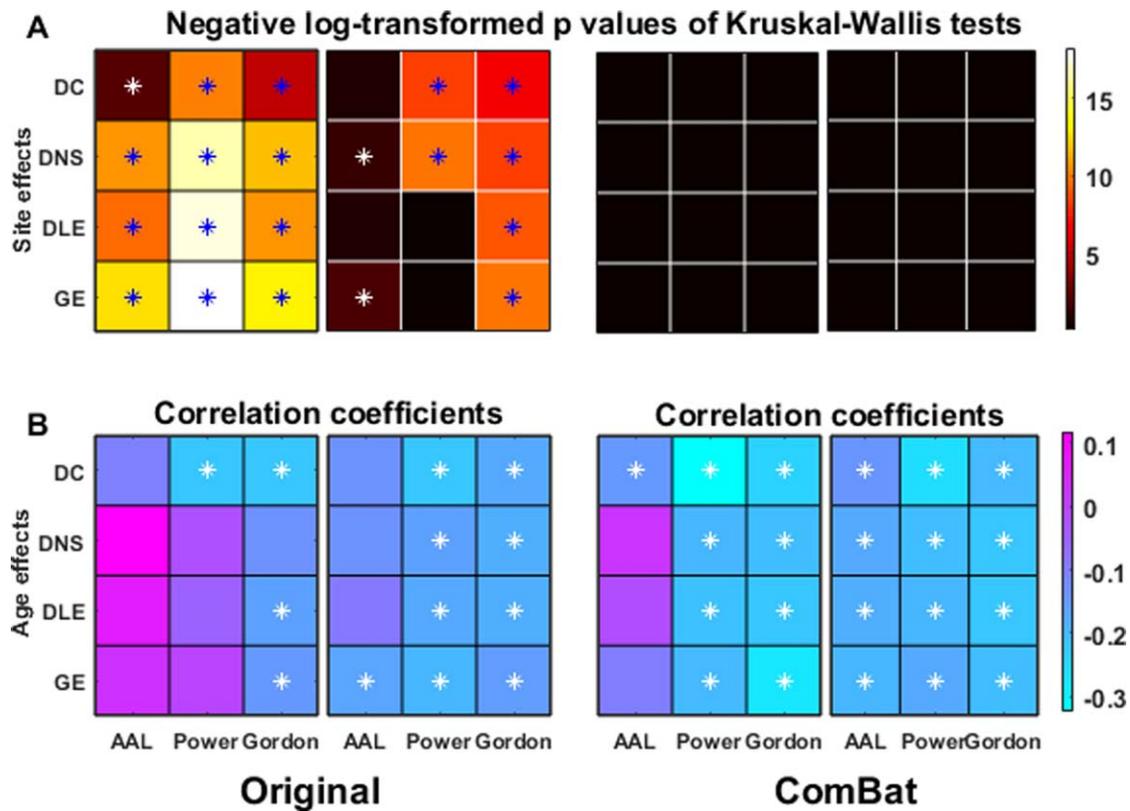
#### 3.3.1 | By connectivity metric

ComBat harmonization preserved or strengthened the anticorrelations between age and DMN functional connectivity and between age and

network efficiency measures. The  $p$  values and correlation values for each metric–atlas combination are displayed in Figure 5b,c, respectively, where we see more significant  $p$  values and stronger correlations post-ComBat. This result was true for both Pearson correlation and wavelet coherence connectivity, with wavelet coherence identifying the strongest anti-correlations both before and after ComBat harmonization. Supporting Information, Figures S8 and S9, subplot B display scatter plots associated with each correlation value.

#### 3.3.2 | By atlas

Using the original data without ComBat harmonization, the Gordon atlas showed significant site effects in all network connectivity and efficiency measures estimated by both Pearson correlation and wavelet



**FIGURE 5** Negative log-transformed  $p$  value heat maps for Kruskal–Wallis tests (e.g., Figure 4a; Supporting Information, Figures S8–S10, subplot A) of network connectivity and efficiency measures (a) and correlation coefficient (b) heat maps for correlation analyses between the network measures and age (e.g., Figure 4b; Supporting Information, Figures S8–S10, subplot B). Note that the asterisks (blue or white colors) represent  $p < .05$ , which corresponds to a negative  $\log_{10}$  transformed  $p$  value  $> 1.301$  in the color bar of (a). Abbreviations: DC = DMN connectivity; DNS = DMN nodal strength; DLE = DMN local efficiency; GE = global efficiency. Note that DMN connectivity was computed by first summing the functional connectivity values within the DMN ROIs, and then normalizing by the number of DMN ROIs corresponding to each atlas (AAL, Power, and Gordon); please see the details of the computation of DNS, DLE, and GE in Section 2.7.

coherence (Figure 5a). However, for AAL and Power atlases, there were no site effects even in some nonharmonized measures estimated by wavelet coherence (Figure 5a).

As shown in Figure 5b,c, ComBat harmonization strengthened the estimated anti-correlations between age and network measures across all three atlases. In particular, for the AAL and Power atlases, ComBat harmonization uncovered significant anti-correlations that were not detected when using the nonharmonized data (Figures 4b and 5b). Among the three atlases, the AAL atlas identified the fewest significant anti-correlations both before and after ComBat harmonization and the magnitudes were generally smaller than those identified by other atlases. The Power atlas identified stonger anti-correlations than the other two atlases post-ComBat. Moreover, a majority of the network measures estimated by the correlation–AAL combination were not negatively associated with age, even after performing ComBat harmonization (see Supporting Information, Figures S8B–S10B for scatter plots associated with the  $p$  and correlation values in Figure 5b,c).

Overall, ComBat harmonization not only removed unwanted site effects in network connectivity and efficiency measures calculated from functional connectivity matrices but also preserved or increased the estimated underlying correlations with age. Some specific combinations of atlases and connectivity metrics appear to be better than

others with respect to revealing significant relationships with age. When considering both site effect removal and correlation with age, we found that the coherence–Power combination performed optimally.

## 4 | DISCUSSION

In this study, we investigated the degree to which combining data from different scanners in a multi-site study could affect downstream analyses of fMRI-based functional connectivity and network efficiency measures. We implemented several visualization techniques and statistical tests to visualize and quantify the scanner effects. We performed ComBat harmonization on fMRI-based functional connectivity matrices to remove site effects before extracting DMN connectivity and network measures. We quantified the site effects and the performance of ComBat harmonization using two different metrics to compute connectivity and three different brain atlases. We demonstrated that ComBat harmonization can successfully remove site effects in the functional connectivity matrices, thereby leading to network connectivity and efficiency measures that are also not different across sites for any choice of connectivity metric and atlas. Moreover, we found that using wavelet coherence with the Power atlas resulted in the best power to detect anti-correlations between age and DMN functional connectivity and

network efficiency measures following ComBat harmonization, suggesting the best preservation of underlying biological signal with this combination.

#### 4.1 | ComBat harmonization removes site effects

As previous studies (Dansereau et al., 2017; Van Horn & Toga, 2009) have consistently reported the existence of considerable site effects in multi-site fMRI measurements that cannot be removed by performing ICA-based approaches (Feis et al., 2015), we tested whether ComBat harmonization could eliminate site effects in several fMRI-based functional connectivity and network measures. Of note, we only performed ComBat harmonization on the original functional connectivity matrices and then subsequently calculated network connectivity and efficiency measures from the harmonized connectivity matrices. Notably, we did not find statistically significant site effects in the downstream network measures.

Given the excellent performance of ComBat in DTI (Fortin et al., 2017), MRI-based cortical thickness (Fortin et al., 2018), and fMRI (this study) measurements, we conclude that this harmonization method is a reliable and powerful technique that can be widely applied to different neuroimaging modalities and summary measurements.

#### 4.2 | Wavelet coherence outperforms Pearson correlation

In this study, to investigate the effects of connectivity metrics on multi-site fMRI measurements and the performance of ComBat harmonization, we used both Pearson correlation and Wavelet coherence to estimate the fMRI functional connectivity. Previous studies have shown that wavelet coherence outperforms Pearson correlation with respect to sensitivity to outliers caused by motion artifact (Huber, 2004; Achard, Salvador, Whitcher, Suckling, & Bullmore, 2006). Additionally, using coherence avoids the need to remove negative correlation coefficients to calculate network measures (Achard & Bullmore, 2007; Bassett et al., 2011) and robustly extracts frequency-specific information from the time series without picking up on edge effects of band-pass filtering (Percival & Walden, 2000; Zhang, Telesford, Giusti, Lim, & Bassett, 2016). However, at present, there is no study comparing the sensitivity to scanner differences of the two connectivity metrics applied to fMRI data. Our results indicate that ComBat harmonization can remove scanner effects from the data, regardless of the choice of connectivity metric. However, wavelet coherence-based measures showed weaker differences across sites than Pearson correlation-based measures in nonharmonized data. Moreover, wavelet coherence measures generally resulted in stronger anti-correlations between age and the connectivity and network measures across all the three atlases (AAL, Power, and Gordon) both before and after harmonization. For multi-site fMRI studies, this result suggests that wavelet coherence may be preferable to Pearson correlation when extracting connectivity and network summary outcomes.

#### 4.3 | Power atlas outperforms AAL and Gordon atlases

We also studied the effects of three atlases (AAL, Power, and Gordon) on multi-site fMRI measurements and the performance of ComBat harmonization. A larger percentage of connections between ROIs were significantly affected by site in the AAL atlas than in the Power and Gordon atlases. These results are consistent with previous findings that in multi-site fMRI studies, functional atlases extracted from large resting-state fMRI datasets outperform traditional anatomical atlases (Abraham et al., 2017). For all three atlases, site effects in the functional connectivity and network measures were successfully removed by ComBat harmonization. However, all the network connectivity and efficiency measures using the AAL atlas were less correlated with age than those using the Power and Gordon atlases, suggesting that the AAL atlas may not be as sensitive to underlying biological variability (assessed using age in this study) when using multi-site fMRI data. Interestingly, we did not find significant site effects using the Power atlas among non-ComBat-harmonized network efficiency measures estimated by wavelet coherence. In contrast, the AAL and Gordon atlases demonstrated strong site effects in these non-ComBat-harmonized network measures. Overall, we concluded that the Power atlas outperforms the AAL and Gordon atlases with respect to post-ComBat analyses of biological variability.

#### 4.4 | Strengths, limitations, and future direction

This study has several strengths: (1) we investigated six combinations of two connectivity metrics and three atlases, and thus were able to explore the ability of ComBat harmonization to remove site effects and to identify combinations of connectivity metrics (wavelet coherence) and atlases (Power and Gordon) that best preserved age-related anti-correlations after harmonization; (2) we used a relatively large sample (228 participants), therefore providing relatively reliable and convincing results; (3) by using the ComBat model, which is generic in its formulations and thus could easily be generalized to additional imaging modalities, our findings may have implications for multi-site electroencephalography, magnetoencephalography and other neurophysiological and neuroimaging datasets; (4) ComBat has been implemented in MATLAB and R (<https://github.com/Jfortin1/ComBatHarmonization>) and in Python (<https://github.com/ncullen93/neuroCombat>) making the technique available and largely applicable to analysts using a variety of different software packages for image processing.

There are also several limitations that should be considered and improved in future studies. First, several previous fMRI studies (Brown et al., 2011; Forsyth et al., 2014; Friedman et al., 2006; Keshavan et al., 2016; Noble et al., 2017; Rath et al., 2016; Shinohara et al., 2017) used traveling-subject datasets in which the same participants were scanned across sites to reduce the subject effects. One recent study (Noble et al., 2017) using a small dataset (8 subjects scanned at each of 8 sites) found that the subject differences were stronger than potential site effects. Although, our ComBat harmonization technique was tested on different participants scanned across different sites with

heterogeneous protocols, we speculate that ComBat harmonization will also have excellent performance on removing any nonbiological variations when applied to traveling-subject datasets; however, this remains to be proven. Second, in some longitudinal datasets, the same participants may be scanned on different scanners over multiple time points. However, the current ComBat harmonization model cannot be directly applied to this type of longitudinal data. Therefore, in the future, we plan to develop a time-dependent ComBat algorithm to study longitudinal fMRI connectivity and network properties. Third, in this study, we tested the performance of ComBat harmonization on two functional connectivity metrics and three atlases (six combinations). Although the ComBat model does not require assumptions on connectivity metrics and atlases, we found that the choices of connectivity metrics and atlases had a strong influence on the magnitude of site effects in fMRI measurements and on preserving biological variability (age in this study). Therefore, future work exploring the performance of ComBat harmonization in other combinations of connectivity metrics (e.g., partial correlation) and atlases (anatomical atlas: Brodmann, 1909; Desikan et al., 2006; functional atlas: Glasser et al., 2016; Schaefer et al., 2017; Wig, Laumann, & Petersen, 2014; Yeo et al., 2011) is warranted. Finally, in this project we focused on the ability of ComBat harmonization to preserve age-related associations with several network connectivity and efficiency measures. However, previous studies (Bullmore & Sporns, 2012; Fornito, Zalesky, & Breakspear, 2015; Stam, 2014; Yu et al., 2016, 2017) have shown that functional brain network organizations are highly correlated with other demographic (e.g., gender, educational level), clinical phenotypes (e.g., disease severity for neurological disorders), and pathological biomarkers (e.g., amyloid- $\beta_{42}$  and tau proteins in Alzheimer's disease). In particular, the EMBARC functional dataset was originally designed to study the potential differences on fMRI measurements between MDD patients and healthy controls (Greenberg et al., 2015; Trivedi et al., 2016; Webb et al., 2016). Future studies will focus on whether the ComBat-harmonized fMRI data preserve functional brain networks (Gong & He, 2015; Kaiser, Andrews-Hanna, Wager, & Pizzagalli, 2015) associated with depression, and whether the abnormal network attributes in MDD after ComBat harmonization are associated with patients' symptoms (Otte et al., 2016; Sheline et al., 2009; Sheline, Price, Yan, & Mintun, 2010; Williams, 2016).

## 5 | CONCLUSION

ComBat harmonization is a powerful technique for removing site effects in functional connectivity matrices, network connectivity, and efficiency measures. In addition, it preserves or strengthens the power to detect age-related anti-correlations in network connectivity and efficiency measures. In the current multi-site fMRI study, the optimal performance of ComBat harmonization was obtained by using wavelet coherence to extract functional connectivity from the Power atlas segmentation of functional brain images.

## ACKNOWLEDGMENTS

The authors would like to thank Dr Danielle Bassett, Dr Richard Betzel, and Dr Prejaas Tewarie for their helpful suggestions about the estimation of functional connectivity. The authors are grateful to Dr Arjan Hillebrand for providing MNI-space centroids of the AAL atlas and his MATLAB code of cortical surface plots in Supporting Information, Figure S8. The authors thank Dr Romain Duprat and Jared Zimmerman for many helpful discussions and suggestions regarding presenting results. They also thank Rastko Ciric for providing and helping them with the XCP Engine and Irem Aselcioglu for helping them with using the ANTS Cortical Thickness pipeline and providing valuable support and discussion. The authors thank Maria Prociuk for her assistance with the preparation and submission of the manuscript. They also thank all patients and controls for their participation. RTS would like to acknowledge support from the National Institute of Neurological Disorders and Stroke (R01 NS085211 & R01 NS060910), the National Institute of Mental Health (R01 MH112847), and the National Multiple Sclerosis Society (RG-1707-28586). The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

## FINANCIAL DISCLOSURES

Except for federal grants, all authors report no biomedical financial interests or potential conflicts of interest.

## ORCID

Yvette I. Sheline  <http://orcid.org/0000-0002-6929-9659>

## REFERENCES

- Abraham, A., Milham, M. P., Di Martino, A., Cameron Craddock, R., Samaras, D., Thirion, B., & Varoquaux, G. (2017). Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. *NeuroImage*, *147*, 736–745.
- Achard, S., Salvador, R., Whitcher, B., Suckling, J., & Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, *26*(1), 63–72.
- Achard, S., & Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. *PLoS Computational Biology*, *3*(2), e17.
- Ajilore, O., Lamar, M., & Kumar, A. (2014). Association of brain network efficiency with aging, depression, and cognition. *American Journal of Geriatric Psychiatry*, *22*(2), 102–110.
- Avants, B. B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., & Gee, J. C. (2010). The optimal template effect in hippocampus studies of diseased populations. *NeuroImage*, *49*(3), 2457–2466.
- 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*(3), 2033–2044.
- Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A., & Gee, J. C. (2011). An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics*, *9*(4), 381–400.

- Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M., & Grafton, S. T. (2011). Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences of the United States of America*, 108(18), 7641–7646.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 57, 289–300.
- Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34(4), 537–541.
- Biswal, B. B., Mennes, M., Zuo, X. N., Gohel, S., Kelly, C., Smith, S. M., ... Milham, M. P. (2010). Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10), 4734–4739.
- Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: Emerging methods and principles. *Trends in Cognitive Sciences*, 14(6), 277–290.
- Brodman, K. (1909). *Localization in the cerebral cortex*, translated by Garey L.J. New York, NY: Springer.
- Brown, G. G., Mathalon, D. H., Stern, H., Ford, J., Mueller, B., Greve, D. N., ... Potkin, S. G. Function Biomedical Informatics Research Network (2011). Multisite reliability of cognitive BOLD data. *NeuroImage*, 54(3), 2163–2175.
- Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews. Neuroscience*, 10(3), 186–198.
- Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nature Reviews. Neuroscience*, 13(5), 336–349.
- Chavez, S., Viviano, J., Zamyadi, M., Kingsley, P. B., Kochunov, P., Strother, S., & Voineskos, A. (2018). A novel DTI-QA tool: Automated metric extraction exploiting the sphericity of an agar filled phantom. *Magnetic Resonance Imaging*, 46, 28–39.
- Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., ... 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.
- Cohen, J. R., & D'Esposito, M. (2016). The segregation and integration of distinct brain networks and their relationship to cognition. *Journal of Neuroscience*, 36(48), 12083–12094.
- Damoiseaux, J. S., Beckmann, C. F., Arigita, E. J., Barkhof, F., Scheltens, P., Stam, C. J., ... Rombouts, S. A. (2008). Reduced resting-state brain activity in the "default network" in normal aging. *Cerebral Cortex (New York, N.Y. : 1991)*, 18(8), 1856–1864.
- Damoiseaux, J. S. (2017). Effects of aging on functional and structural brain connectivity. *NeuroImage*, 160, 32–40.
- Dansereau, C., Benhajali, Y., Risterucci, C., Pich, E. M., Orban, P., Arnold, D., & Bellec, P. (2017). Statistical power and prediction accuracy in multisite resting-state fMRI connectivity. *NeuroImage*, 149, 220–232.
- Das, S. R., Avants, B. B., Grossman, M., & Gee, J. C. (2009). Registration based cortical thickness measurement. *NeuroImage*, 45(3), 867–879.
- Delaparte, L., Yeh, F. C., Adams, P., Malchow, A., Trivedi, M. H., Oquendo, M. A., ... DeLorenzo, C. (2017). A comparison of structural connectivity in anxious depression versus non-anxious depression. *Journal of Psychiatric Research*, 89, 38–47.
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3), 968.
- Di Martino, A., Yan, C. G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., ... Milham, M. P. (2014). The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, 19(6), 659–667.
- Dijkstra, E. W. (1959). A note on two problems in connection with graphs. *Numerische Mathematik*, 1(1), 269–271.
- Doucet, G. E., Bassett, D. S., Yao, N., Glahn, D. C., & Frangou, S. (2017). The role of intrinsic brain functional connectivity in vulnerability and resilience to bipolar disorder. *American Journal of Psychiatry*, 174(12), 1214–1222.
- Feis, R. A., Smith, S. M., Filippini, N., Douaud, G., Dopper, E. G., Heise, V., ... Mackay, C. E. (2015). ICA-based artifact removal diminishes scan site differences in multi-center resting-state fMRI. *Frontiers in Neuroscience*, 9, 395.
- Ferreira, L. K., & Busatto, G. F. (2013). Resting-state functional connectivity in normal brain aging. *Neuroscience and Biobehavioral Reviews*, 37(3), 384–400.
- First, M. B., Spitzer, R. L., & Gibbon, M. (2002). *Structured clinical interview for DSM-IV-TR axis I disorders, research version, patient edition (SCID-I/P)*. New York, NY: Biometrics Research, New York State Psychiatric Institute.
- Fornito, A., Zalesky, A., & Breakspear, M. (2015). The connectomics of brain disorders. *Nature Reviews. Neuroscience*, 16(3), 159–172.
- Fornito, A., Zalesky, A., & Bullmore, E. (2016). *Fundamentals of brain network analysis*. Cambridge: Academic Press.
- Forsyth, J. K., McEwen, S. C., Gee, D. G., Bearden, C. E., Addington, J., Goodyear, B., ... Cannon, T. D. (2014). Reliability of functional magnetic resonance imaging activation during working memory in a multi-site study: Analysis from the North American Prodrome Longitudinal Study. *NeuroImage*, 97, 41–52.
- Fortin, J. P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., ... Shinohara, R. T. (2017). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, 161, 149–170.
- Fortin, J. P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., ... Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, 167, 104–120.
- Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews. Neuroscience*, 8(9), 700–711.
- Friedman, L., & Glover, G. H. Fbirm Consortium (2006). Reducing inter-scanner variability of activation in a multicenter fMRI study: Controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. *NeuroImage*, 33(2), 471–481.
- Friedman, L., Stern, H., Brown, G. G., Mathalon, D. H., Turner, J., Glover, G. H., ... Potkin, S. G. (2008). Test-retest and between-site reliability in a multicenter fMRI study. *Human Brain Mapping*, 29(8), 958–972.
- Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615), 171–178.
- Glover, G. H., Mueller, B. A., Turner, J. A., van Erp, T. G., Liu, T. T., Greve, D. N., ... Potkin, S. G. (2012). Function biomedical informatics research network recommendations for prospective multicenter functional MRI studies. *Journal of Magnetic Resonance Imaging*, 36(1), 39–54.
- Gong, Q., & He, Y. (2015). Depression, neuroimaging and connectomics: A selective overview. *Biological Psychiatry*, 77(3), 223–235.
- Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., & Petersen, S. E. (2016). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral Cortex*, 26(1), 288–303.

- Gountouna, V. E., Job, D. E., McIntosh, A. M., Moorhead, T. W., Lymer, G. K., Whalley, H. C., ... Lawrie, S. M. (2010). Functional magnetic resonance imaging (fMRI) reproducibility and variance components across visits and scanning sites with a finger tapping task. *NeuroImage*, 49(1), 552–560.
- Gradin, V., Gountouna, V. E., Waiter, G., Ahearn, T. S., Brennan, D., Condon, B., ... Steele, J. D. (2010). Between- and within-scanner variability in the CaliBrain study n-back cognitive task. *Psychiatry Research*, 184(2), 86–95.
- Grady, C. L., Protzner, A. B., Kovacevic, N., Strother, S. C., Afshin-Pour, B., Wojtowicz, M., ... McIntosh, A. R. (2010). A multivariate analysis of age-related differences in default mode and task-positive networks across multiple cognitive domains. *Cerebral Cortex (New York, N.Y. : 1991)*, 20(6), 1432–1447.
- Greenberg, T., Chase, H. W., Almeida, J. R., Stiffler, R., Zevallos, C. R., Aslam, H. A., ... Phillips, M. L. (2015). Moderation of the relationship between reward expectancy and prediction error-related ventral striatal reactivity by anhedonia in unmedicated major depressive disorder: Findings from the EMBARC study. *American Journal of Psychiatry*, 172(9), 881–891.
- Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1), 63–72.
- Hallquist, M. N., Hwang, K., & Luna, B. (2013). The nuisance of nuisance regression: Spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures functional connectivity. *NeuroImage*, (2013). 82, 208–225.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 23(1), 56–62.
- Huber, P. J. (2004). *Robust statistics*. Wiley.
- 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(2), 825–841.
- Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics (Oxford, England)*, 8(1), 118–127.
- Jovicich, J., Minati, L., Marizzoni, M., Marchitelli, R., Sala-Llonch, R., Bartrés-Faz, D., ... Frisoni, G. B. PharmaCog Consortium (2016). Longitudinal reproducibility of default-mode network connectivity in healthy elderly participants: A multicentric resting-state fMRI study. *NeuroImage*, 124(Pt A), 442–454.
- Kaiser, R. H., Andrews-Hanna, J. R., Wager, T. D., & Pizzagalli, D. A. (2015). Large-scale network dysfunction in major depressive disorder: A meta-analysis of resting-state functional connectivity. *JAMA Psychiatry*, 72(6), 603–611.
- Keshavan, A., Paul, F., Beyer, M. K., Zhu, A. H., Papinutto, N., Shinohara, R. T., ... Henry, R. G. (2016). Power estimation for non-standardized multisite studies. *NeuroImage*, 134, 281–294.
- Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., ... Parsey, R. V. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3), 786–802.
- Koch, W., Teipel, S., Mueller, S., Buerger, K., Bokde, A. L., Hampel, H., ... Meindl, T. (2010). Effects of aging on default mode network activity in resting state fMRI: Does the method of analysis matter? *NeuroImage*, 51(1), 280–287.
- Kochunov, P., Dickie, E. W., Viviano, J. D., Turner, J., Kingsley, P. B., Jahanshad, N., ... Voineskos, A. N. (2018). Integration of routine QA data into mega-analysis may improve quality and sensitivity of multi-site diffusion tensor imaging studies. *Human Brain Mapping*, 39(2), 1015–1023.
- Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, 87(19), 198701.
- McGonigle, D. J. (2012). Test-retest reliability in fMRI: Or how I learned to stop worrying and love the variability. *NeuroImage*, 62(2), 1116–1120.
- Noble, S., Scheinost, D., Finn, E. S., Shen, X., Papademetris, X., McEwen, S. C., ... Constable, R. T. (2017). Multisite reliability of MR-based functional connectivity. *NeuroImage*, 146, 959–970.
- Oh, J., Bakshi, R., Calabresi, P. A., Crainiceanu, C., Henry, R. G., Nair, G., ... Sicotte, N. L. NAIMS Cooperative Steering Committee (2017). The NAIMS cooperative pilot project: Design, implementation and future directions. *Multiple Sclerosis Journal*, 135245851773999.
- Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., ... Schatzberg, A. F. (2016). Major depressive disorder. *Nature Reviews. Disease Primers*, 2, 16065.
- Percival, D. B., & Walden, A. T. (2000). *Wavelet methods for time series analysis*. Cambridge University Press.
- Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., ... Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, 72(4), 665–678.
- Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience*, 38, 433–447.
- Rath, J., Wurnig, M., Fischmeister, F., Klinger, N., Höllinger, I., Geißler, A., ... Beisteiner, R. (2016). Between- and within-site variability of fMRI localizations. *Human Brain Mapping*, 37(6), 2151–2160.
- Rosazza, C., & Minati, L. (2011). Resting-state brain networks: literature review and clinical applications. *Neurol Sci*, 32(5):773–85.
- Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3), 1059–1069.
- Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., ... Keller, M. B. (2003). The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, 54(5), 573–583.
- Satterthwaite, T. D., Ciric, R., Roalf, D. R., Davatzikos, C., Bassett, D. S., & Wolf, D. H. (2017). Motion artifact in studies of functional connectivity: Characteristics and mitigation strategies. *Human Brain Mapping*. <https://doi.org/10.1002/hbm.23665>. [Epub ahead of print]
- Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X. N., Holmes, A. J., ... Yeo, B. T. T. (2017). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb Cortex* 18. *Cerebral Cortex (New York, N.Y. : 1991)*, 1–20.
- Sheline, Y. I., Barch, D. M., Price, J. L., Rundle, M. M., Vaishnavi, S. N., Snyder, A. Z., ... Raichle, M. E. (2009). The default mode network and self-referential processes in depression. *Proceedings of the National Academy of Sciences of the United States of America*, 106(6), 1942–1947.
- Sheline, Y. I., Price, J. L., Yan, Z., & Mintun, M. A. (2010). Resting-state functional MRI in depression unmasks increased connectivity between networks via the dorsal nexus. *Proceedings of the National Academy of Sciences of the United States of America*, 107(24), 11020–11025.
- Shinohara, R. T., Oh, J., Nair, G., Calabresi, P. A., Davatzikos, C., Doshi, J., ... Bakshi, R. NAIMS Cooperative (2017). Volumetric analysis from a harmonized multisite brain MRI study of a single subject with multiple sclerosis. *American Journal of Neuroradiology*, 38(8), 1501–1509.
- Stam, C. J. (2014). Modern network science of neurological disorders. *Nature Reviews. Neuroscience*, 15(10), 683–695.
- Suckling, J., Ohlssen, D., Andrew, C., Johnson, G., Williams, S. C., Graves, M., ... Bullmore, E. (2008). Components of variance in a multicentre functional MRI study and implications for calculation of statistical power. *Human Brain Mapping*, 29(10), 1111–1122.

- Suckling, J., Barnes, A., Job, D., Brenan, D., Lymer, K., Dazzan, P., . . . Deakin, B. (2010). Power calculations for multicenter imaging studies controlled by the false discovery rate. *Human Brain Mapping, 31*(8), 1183–1195.
- Tomasi, D., & Volkow, N. D. (2012). Aging and functional brain networks. *Molecular Psychiatry, 17*(5), 471–458.
- Trivedi, M. H., McGrath, P. J., Fava, M., Parsey, R. V., Kurian, B. T., Phillips, M. L., . . . Weissman, M. M. (2016). Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. *Journal of Psychiatric Research, 78*, 11–23.
- Turner, J. A., Damaraju, E., van Erp, T. G., Mathalon, D. H., Ford, J. M., Voyvodic, J., . . . Calhoun, V. D. (2013). A multi-site resting state fMRI study on the amplitude of low frequency fluctuations in schizophrenia. *Frontiers in Neuroscience, 7*, 137.
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging, 29*(6), 1310–1320.
- Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., . . . Avants, B. B. (2014). Large-scale evaluation of ants and freesurfer cortical thickness measurements. *NeuroImage, 99*, 166–179.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., . . . Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage, 15*(1), 273–289.
- Van Horn, J. D., & Toga, A. W. (2009). Multi-site neuroimaging trials. *Current Opinion in Neurology, 22*(4), 370–378.
- Watson, D., & Clark, L. A. (1991). *The mood and anxiety symptom questionnaire*. Iowa City, IA: University of Iowa.
- Webb, C. A., Dillon, D. G., Pechtel, P., Goer, F. K., Murray, L., Huys, Q. J., . . . Pizzagalli, D. A. (2016). Neural correlates of three promising endophenotypes of depression: Evidence from the EMBARC study. *Neuropsychopharmacology, 41*(2), 454–463.
- Wig, G. S., Laumann, T. O., & Petersen, S. E. (2014). An approach for parcellating human cortical areas using resting-state correlations. *NeuroImage, 93*, 276–291.
- Williams, L. M. (2016). Precision psychiatry: A neural circuit taxonomy for depression and anxiety. *Lancet Psychiatry, 3*(5), 472–480. May
- Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A network visualization tool for human brain connectomics. *PLoS One, 8*(7), e68910.
- Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., . . . Polimeni, J. R. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology, 106*(3), 1125–1165.
- Yu, M., Gouw, A. A., Hillebrand, A., Tijms, B. M., Stam, C. J., van Straaten, E. C. W., & Pijnenburg, Y. A. L. (2016). Different functional connectivity and network topology in behavioral variant of frontotemporal dementia and Alzheimer's disease: An EEG study. *Neurobiology of Aging, 42*, 150–162.
- Yu, M., Engels, M. M. A., Hillebrand, A., van Straaten, E. C. W., Gouw, A. A., Teunissen, C., . . . Stam, C. J. (2017). Selective impairment of hippocampus and posterior hub areas in Alzheimer's disease: An MEG-based multiplex network study. *Brain, 140*(5), 1466–1485.
- Zhang, Z., Telesford, Q. K., Giusti, C., Lim, K. O., & Bassett, D. S. (2016). Choosing wavelet methods, filters, and lengths for functional brain network construction. *PLoS One, 11*(6), e0157243.

#### SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Yu M, Linn KA, Cook PA, et al. Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. *Hum Brain Mapp.* 2018;00:1–15. <https://doi.org/10.1002/hbm.24241>