Online Supplemental Methods

Two multi-echo fMRI resting state datasets are examined in this paper. The first dataset is the 89-subject 4-echo cohort published in (1), referred to as the ME (multi-echo) dataset. The second dataset was a smaller 12-subject 3-echo cohort acquired at the National Institutes of Health, referred to as the NA dataset (so named because the scans were acquired as part of a task study on object naming; only the resting state fMRI scans are used). Each subject's data included a high-resolution T1-weighted (MP-RAGE) scan and one or more resting state T2*-weighted (BOLD-weighted) multi-echo scans. Pertinent details of these studies are provided in SI Appendix, Table S2.

Structural data: processing

All T1-weighted images underwent automated segmentation by FreeSurfer version 5.3. FreeSurfer segmentation yielded high-resolution (1 mm isotropic) masks of brain compartments. Lower resolution masks to match the fMRI data (3 mm isotropic) were created using nearest-neighbor resampling. Separate gray matter masks of the cortical ribbon, cerebellum, and subcortical nuclei were created. Masks of the white matter underwent 0-4 erosion cycles, and ventricle masks underwent 0-2 erosion cycles at 1 mm resolution prior to resampling to fMRI resolution.

To separate signals with varying proximity to gray matter, fMRI-resolution masks were subtracted from each other to identify voxels in superficial, deeper, and deepest nuisance masks (in white matter: ero0-ero2, ero2-ero4, and ero4 masks). In the ventricles, which are much smaller, only superficial and deeper masks (ero0-2 and ero2) were possible. In Figures, color darkness corresponds to mask depth (i.e., light green are superficial voxels, dark green are deepest voxels). These procedures are identical to those described in (2, 3).

Two whole-brain masks are used. The all-brain mask contains all voxels inside the brain based on FreeSurfer segmentation and is used to mask images prior to ME-ICA. Because the spatial coverage of the structural image is often larger than that of the fMRI data (e.g., fMRI acquisitions often crop out portions of the cerebellum), another whole-brain mask including only voxels with mean fMRI values over 20% of the modal value was combined with all the above-mentioned masks to exclude any non-scanned parts of the brain.

fMRI data: preprocessing

fMRI scans of all cohorts underwent identical processing (except as noted). Throughout processing the first 4 volumes from all fMRI scans were ignored to ensure magnetization equilibrium in used data.

To prepare scans for analysis, all scans first underwent slice time correction using AFNI's 3dTshift, followed by motion correction via AFNI's 3dvolreg. If multiple runs existed (NA cohort only), the 5^{th} volume of the first run was the target volume for motion correction, thus enforcing cross-run alignment. This 5^{th} volume was linearly registered to the subject's T1-weighted image using AFNI's 3dAllineate, and the T1-weighted image was linearly registered to the TT_N27 atlas space via AFNI's @auto_tlrc command. All transformations were concatenated to resample the slice-time-corrected fMRI images to atlas space at 3 mm isotropic resolution in a single step. Prior to examination of the temporal properties of the data, all images of all cohorts were visually inspected to ensure proper registration and adequate brain coverage.

Note that a 4-echo dataset exists as 4 volumes per run and that these volumes were processed independently of each other. However, the resulting independently-estimated transforms are nearly identical, reflecting both the stability of the registration algorithms and the fact that only ~12-16 ms elapse between each version of an image (see detailed discussion of motion in these data in (4)). These separate preprocessed volumes - slice-time corrected, motion-corrected, and in atlas space - serve as the substrates of analyses leveraging multi-echo signal decay. We also performed all multi-echo computations in scanner space followed by the appropriate transformations into atlas space, but these results were comparable to our main results and accordingly we do not show them.

fMRI data: processing to leverage multi-echo decay properties

To obtain ME-ICA denoised data, the preprocessed images were masked by the whole-brain mask, vertically concatenated, and then submitted using "default" settings to tedana.py as implemented in AFNI[1]. As described in (1, 5), this procedure yielded a single "optimally combined" image that merged the multiple volumes for various echo times according to each voxel's estimated T2*, a list of ICA components and estimates of the extent of R2* and S0 modulation observed in the components across echoes (summarized by kappa and rho values, which are statistics describing the fit of signals over echoes to ΔR2* and ΔS0 models, respectively, see (5) for full definitions and discussion), and the "multi-echo ICA denoised" image excluding components that show high S0 dependence but low R2* dependence as determined automatically by elbows in the kappa and rho distributions. The difference between the "optimally combined" and "multi-echo ICA denoised" images is the "discarded" image.

Multi-echo ICA leverages the covariance structure of tens or hundreds of thousands of voxel signals over time to identify major signals in the data, followed by classification of these signals as "BOLD-like" or "not-BOLD-like" according to their decay properties across echoes. This procedure has the advantage of finding structured signals, but inherits some constraints of ICA. Because signal separation by ICA is optimized for spatial independence of the signals, ICA tends to place spatially specific signals into single components but will not place widespread signals into single components. In other words, global or near-global signals will pervade many component timeseries.

We thus decided to also fit signal decay, at each voxel and timepoint, to the monoexponential decay, $S(t) = S_o e^{-t/T2^*}$, using Matlab's nlinfit with least squares optimization. This 2-parameter model requires at least 2 echoes to fit, and the accuracy of the fit will increase with increasing echoes. This technique is noisy in 3-echo data, and somewhat less noisy in 4-echo data, but it also has no spatial constraints on S0 and R2* estimation. Under this approach, fMRI timeseries become decomposed into S0-timeseries and R2*-timeseries (or, equivalently, T2*-timeseries). An example of signals and fits is shown in SI Appendix, Fig. S1. We use this approach not to advocate its use in the present data (FIT would perform better with more echoes), but instead to show that certain separations of S0 and R2* effects by ME-ICA are truly S0 and R2* effects, the spatial constraints of ICA aside.

In figures, fMRI timeseries are presented after "mode 1000" normalization (the modal value of all voxels at all times in a scan defines 1000 units). The scaling is held constant across images

[1] A forked, freestanding version of ME-ICA is available at: https://bitbucket.org/prantikk/me-ica

when possible (e.g., from “optimally combined” to “multi-echo ICA denoised”), but since multiple kinds of images are presented (R2*-timeseries vs. optimally combined timeseries, etc.) different scaling factors must sometimes be used.

Physiological traces

The 12 NA datasets includes pulse oximeter traces (in arbitrary units) and respiratory belt traces (in arbitrary units) for each scan, both sampled at 500Hz. The pulse oximeter traces underwent automated peak detection in Matlab to detect systole, and instantaneous heart rate was calculated from peak-to-peak intervals. Pulse oximeter traces contain many artifacts, and suspicious portions of signals were identified using methods such as sensible checks on changes in heart rate (e.g., >25 Δbpm in a single beat). All signals and detected peaks were visually checked in their entirety to exclude spurious peaks. The respiratory belt traces are presented without further processing and were submitted to RetroTS.m in AFNI to derive respiration volume per unit time (RVT) measures following (6). In Figure 2 and elsewhere, respiratory pattern variability is defined as the standard deviation of the envelope of the z-scored respiratory belt record (which is in arbitrary units). This exact measure to our knowledge has not been used in the literature but is a sensible way to gauge the variety of slow, fast, deep, and shallow breathing patterns that occur in a scan.

fMRI data: nuisance regressions

Several denoising strategies are examined to attempt to remove ventilation-related signals from data after ME-ICA or FIT, shown in Figures S5-7. The nuisance regression of mean compartment signals used the deepest white matter mask and deep ventricle mask in each subject, as well as the 6 realignment parameters and their first temporal derivatives, for a total of 14 regressors (Figures S6 and S7). In SI Appendix, Fig. S5, two models of respiratory variance were tested. First, the RVT measures at 5 lags (the default output of RetroTS.m) were used as regressors, as well as 5 variables created by convolving each of those regressors with a “respiratory response function” created in (3). Second, the RV (respiratory variance) measure was created from the standard deviation of the respiratory trace within a 6-second window (following (7)), and this trace was lagged back 3 seconds and forward 3 seconds to create 3 regressors, and each of these regressors was convolved with the RRF defined above for a total of 6 respiratory regressors. All regressors had mean and linear trend terms removed and were standardized prior to regressions.

fMRI data: approaches to separating widespread signals from sparse signals

After demonstrating in the NA cohort that widespread R2* signals are often linked to respiratory patterns, we demonstrate in detail two methods for removing widespread signals: GODEC and regression of mean gray matter signals. Results related to a third method, robust principal components analysis (PCA), are shown in supplemental figures, as are results related to a fourth method, CompCor.

All of these methods were applied to “multi-echo ICA denoised” images.

Mean signal regression entailed computing the mean signal of the cortical ribbon in a subject (this mean signal, as mentioned in the main text, correlates at nearly r = 1 with the average of all gray matter signals and with the average of all within-brain signals), removing mean and trend terms from the mean signal regressor and all voxel timeseries, and performing a univariate regression. We show the variance removed and retained (in the residuals) by this regression.

CompCor describes a family of approaches that identify nuisance signals in non-gray matter portions of fMRI scans and attempt to derive nuisance regressors from those signals by PCA (8). Our results related to CompCor take 3 forms, and leverage the signals in each of the 3 white matter nuisance compartment masks and 2 ventricle compartment masks mentioned above. In each of these masks and in each subject, the mean signal was calculated and then correlated with the global signal, and the values of these correlations in different masks are shown in SI Appendix, Fig. S11. In addition to the mean signal, we also decomposed the voxel signals in each compartment by PCA and SVD, selected the 5 signals that explained the most variance (as is commonly done in CompCor approaches), correlated these signals with the global (mean cortical) signal, and plotted the highest correlation for each subject within each mask space in SI Appendix, Fig. S11. The mean value across subjects within each compartment for each treatment of the nuisance signals is shown with a large red dot. For the purposes of Figures S12 and S13, only white matter signals in the deepest mask were used (for these signals are the most isolated from and distinct from those of the gray matter).

GODEC and robust PCA, in contrast to CompCor approaches, operate on gray matter timeseries and attempt to divide widespread signals from the other signals of the gray matter. GODEC is a multivariate technique related to robust PCA, and both techniques are part of a large family of techniques intended to separate background signals from sparse signals in large matrices. GODEC and robust PCA can achieve similar solutions, but GODEC is considerably more computationally efficient, which is why we focused on it in this paper. GODEC is formally described below, Python code for our implementation is found here[2], and the paper detailing the algorithm (9) can be found here[3] or at the website accompanying this paper[4]. Numerous implementations of robust PCA have been developed in the past two decades. Code for many approaches can be found here[5], and the specific form we applied (Augmented Lagrange Multiplier (ALM)) is described here[6] with Matlab code available at the just-mentioned website or the website accompanying this paper.

Robust PCA decomposes data into the sum of a low-rank space of signals and a sparse space of signals. The low-rank space includes a high percentage of variance in the data but is explained by a small number of uncorrelated signals. In contrast, the sparse space includes a low percentage of variance in the data but is explained by a large number of uncorrelated signals. GODEC differs from conventional robust PCA in accommodating a random noise component and in using different techniques to achieve the signal partitioning, described below. The low-rank space was considered to carry the global signal. The sparse space was considered to carry BOLD network activity. Our implementation of GODEC is Python-based (godec.py), and included a random sampling method to estimate the covariance matrix iteratively with a power method as described below. We also included steps of discrete wavelet transform before and after GODEC to conserve autocorrelation in the final solution, using the Daubechies wavelet. A rank-1 approximation was used, with 100 iterations.

---

[2] Code implementing GODEC is available with ME-ICA at: https://bitbucket.org/prantikk/me-ica
[3] https://sites.google.com/site/godecomposition/home
[4] www.jonathanpower.net/paper-multiecho.html
[5] http://perception.csl.illinois.edu/matrix-rank/sample_code.html
[6] https://arxiv.org/abs/1009.5055

As mentioned in the main text, all background separation algorithms have tunable parameters. We used parameters that returned low-rank spaces with rank approximately of 1-4 to minimize removal of signals associated with resting state networks. Inclusion of higher numbers of signals in the background for in these data (e.g., ~10-15) began to remove much of the on-block structure of the correlation matrices shown in SI Appendix, Fig. S14, consistent with incorporation of resting state networks in the background, which we deemed undesirable for our purposes of separating the largely additive global respiratory signals from other signals in the data.

In Figures 3, S9, and S12 we show signal separation by mean signal regression, by GODEC, by robust PCA, and by CompCor as gray plots. Online Movie 6 shows such separations for all ME subjects. In SI Appendix, Fig. S13 we plot the mean correlation matrix after all of these techniques, the histograms of these matrices (all are approximately zero-centered), and plot the matrices against one another.

Implementation of GODEC

Given an $m\ x\ n$ dense matrix X, we use GODEC to approximate X as $X = L + S + G$, where L is a low-rank space, S is a sparse space, and G is random noise. This approximation differs from the decomposition performed by robust PCA (RPCA) which solves $X = L + S$. GODEC's computations can be performed much more quickly than those of RPCA, which is the chief reason it has been chosen in this paper.

We use the bilateral random projection (BRP) method to estimate L as a rank-r approximation of X such that $L = Y_1(A_2^T Y_1)^{-1} Y_2^T$, where $Y_1 = XA_1$and $Y_2 = X^T A_2$. $A_1 \in \mathbb{R}^{n\ x\ r}$ and $A_2 \in \mathbb{R}^{m\ x\ r}$ are random matrices. Efficient estimation of L is accomplished via an iterative updating process, where we use the right random projection Y_1 to build a better left projection matrix A_2 and then use $Y_2$to build A_1. In particular after $Y_1 = XA_1$, we set $A_2 = Y_1$ and calculate $Y_2 = X^T A_2$, and then we set $A_1 = Y_2$ and calculate $Y_1 = XA_1$.

In cases where the singular values of X decay slowly, a power scheme modification can improve estimation: X is reformulated as $\tilde{X} = (XX^T)^q X$ such that $\lambda_i(\tilde{X}) = \lambda_i(\tilde{X})^{2q+1}$. The singular values of $\tilde{X}$ decay faster than those of X but the matrices share the same singular vectors. The BRP of $\tilde{X}$ is $\tilde{L} = Y_1(A_2^T Y_1)^{-1} Y_2^T$, where $Y_1 = \tilde{X}A_1$and $Y_2 = \tilde{X}^T A_2$. We use QR decomposition to calculate $Y_1 = Q_1 R_1$ and $Y_2 = Q_2 R_2$.
The low-rank approximation of X thus becomes $L = (\tilde{L})^{\frac{1}{2q+1}} = Q_1[R_1(A_2^T Y_1)^{-1} R_2^T]^{\frac{1}{2q+1}} Q_2^T$.

The error of the BRP procedure, including when paired with the power scheme, approaches the error of SVD (which is used in many robust PCA approaches) under mild conditions. However, whereas the computational cost of SVD is either $m^2 n$ or mn^2, the cost of GODEC is of order $r^2(m + 3n + 4r) + (4q + 4)mnr$, with r typically far less than m or n.

The overall goal of GODEC is to minimize the decomposition error: $\min_{L,S} \|X - L - S\|_F^2$, with rank $(L) \leq r$ and card$(S) \leq k$. GODEC accomplishes this task by iteratively solving two subproblems:

$$\begin{cases} L_t = \arg \min_{\text{rank}\,(L)\leq r} \|X - L - S_{t-1}\|_F^2; \\ S_t = \arg \min_{\text{card}\,(S)\leq k} \|X - L_t - S\|_F^2 . \end{cases}$$

These subproblems are solved by updating L_t by singular value hard thresholding of $X - S_{t-1}$ and by updating S_t by entry-wise hard thresholding of $X - L_t$.

Algorithmically, our implementation is as follows:
Input: X, r, k, ϵ, q
Output: L, S
Initialize: $L_0 = X;\ S_0 = 0;\ t = 0$
while $\|X - L_t - S_t\|_F^2 / \|X\|_F^2 \ > \ \epsilon$ do
$t = t + 1$
$\tilde{L} = [(X - S_{t-1})(X - S_{t-1})^T]^q (X - S_{t-1})$
$Y_1 = \tilde{L}A_1, A_2 = \ Y_1$
$Y_2 = \tilde{L}^T Y_1 = Q_2 R_2, Y_1 = \tilde{L} Y_2 = \ Q_1 R_1$
If $rank(A_2^T Y_1) < r$ then $r = rank(A_2^T Y_1)$, goto first step, end;
$L_t = Q_1 [R_1 (A_2^T Y_1)^{-1} R_2^T]^{\frac{1}{2q+1}} Q_2^T$
$S_t = \wp_\Omega (X - L_t)$, with Ω the nonzero subset of the first k entries of $|X - L_t|$
end while

Motion-related analyses

A variety of analyses have been developed to identify motion-related influences in functional connectivity data (reviewed in (10)). Here we utilize four techniques: 1) visualizing thousands of voxel timeseries at the single-subject level as a heatmap paired with motion traces, following (11); 2) "QC:RSFC" analyses, which document the linear dependence of resting state functional connectivity (RSFC) measures on quality control (QC) measures, following (12); 3) high-low motion differences, which document disparities between motion-binned groups of subjects; and 4) scrubbing analyses, which demonstrate how covariance changes when certain volumes of a scan are excluded from calculations.

Head position traces: the {X Y Z pitch roll yaw} estimates provided by 3dvolreg. The motion estimates plotted in the paper and used for QC:RSFC analyses are derived in unprocessed data since motion estimates on slice time corrected data can be misleadingly low (2, 4). The 5th volume was the reference volume.

Head motion traces: head motion was calculated as the sum of the absolute values of the derivatives of the head position traces, after converting rotational estimates to arc displacements at a radius of 50 mm (approximately the radius of the cortex). This measure is called Framewise Displacement (FD) (13). FD = 0 for the first volume of a scan by convention.

Heatmaps: voxel timeseries are shown as a gray-scale heatmap (time on x-axis, voxels on y-axis), with voxels organized by the FreeSurfer-derived brain compartments mentioned above, following procedures described in (2, 3). Colored bars at the side of the voxel timeseries denote the various brain compartments, and a bright green line separates gray matter timeseries from all other brain compartments. Mean and trend terms were removed from all voxel timeseries. Above the heatmaps, position estimates are shown in a rainbow of colors, and head motion (FD)

traces are shown in a thicker red trace. In some maps, heart rate (in beats per minute, derived from the pulse oximeter record) and breathing traces (from the respiratory belt) are shown as well. Scales of traces and heatmaps are held constant across subjects and figures.

QC:RSFC analyses: In Figure 4 the QC measure employed was mean FD, summarizing the amount of motion in a scan. The RSFC measure employed was the correlation between fMRI timeseries at various pairs of regions throughout the brain. 264 regions of interest (ROIs) defined in (14) were used, yielding ~35,000 pairwise correlations. The dependence of each of these correlations on motion was quantified by correlating the Fisher-z-transformed correlations across subjects with the mean FD across subjects to yield a 264 x 264 matrix. The values in this matrix are plotted as a function of the distance between the ROIs, thus providing an indication of how increasing motion modulates correlations between nearby versus distant ROIs, across subjects. By examining QC:RSFC plots across stages of analysis, the influence (and removal) of motion-linked focal and global influences can be recognized (see conceptual discussion in (10)). 10,000 permutations of mean FD were used to establish significance. In Figure 5, both mean FD and respiratory pattern variability are separately used for QC:RSFC analyses.

High-low motion differences: subjects were binned into high- and low-motion groups by median split of mean FD, and, using the same correlation matrices as above, the 264 x 264 mean matrix from the low-motion group was subtracted from the 264 x 264 mean matrix of the high-motion group. The resulting 264 x 264 matrix was plotted as a function of distance, as above. 10,000 permutations of mean FD were used to establish significance.

Scrubbing analyses: in each subject, volumes with FD > 0.2 mm were identified, and 264 x 264 correlation matrices were calculated for each subject either from signals at all volumes or only from signals at non-identified volumes (thus eliminating high-motion volumes from calculations). The difference between scrubbed and unscrubbed matrices was calculated in each subject, and the mean value of this difference across subjects was calculated. The resulting 264 x 264 matrix was plotted as a function of distance, as above. Subjects with no timepoints excluded were withheld from analysis, and subjects with more than 50% of timepoints identified were withheld from analysis. 10,000 permutations that randomly censored identical numbers of volumes were used to establish significance.

In plots of these analyses (Figures 4, 5, and S10), red points are the real data. White smoothing curves generated via sliding boxcars over 1000 data points summarize the red points, and black smoothing curves from 50 of the 10,000 permutation analyses are shown to visually represent random effects. The rank of the real versus permuted smoothing curves at 35 mm is listed as a percentile to index general dependence on motion, and the comparable rank for the difference between the smoothing curve at 35 and 100 mm is shown to index distance dependence (e.g., a rank of 99.97% indicates that the observed value is greater than 9,997 of the 10,000 permutations).

So that the same subjects could be studied in both Figure 4 and S10 (SI Appendix, Fig. S10 repeats the analyses of Figure 4 with only low-motion volumes, i.e., those not removed by scrubbing analyses), we excluded from the above calculations subjects who lost more than 50% of their volumes to scrubbing, and also excluded subjects who lost zero volumes to scrubbing (in order not to dilute scrubbing effects). Other selection criteria, such as using only subjects with mean FD < 0.3 mm or < 0.2 mm (so that a minority of badly corrupted scans with highly unusual

values do not dominate correlation calculations) yielded similar results. Detailed information on motion in the ME cohort can be found at (4), and SI Appendix, Fig. S15 shows the distribution of mean FD among the ME subjects. In Figure 5, the same criteria were used for the NA scans, except the FD value for scrubbing was set to 0.25 due to slightly higher "floors" in the FD values compared to ME data.

Online Supplemental Results

ME-ICA places global BOLD signals into multiple BOLD-like ICA components

ICA optimizes signal separation using spatial criteria. Signals found throughout the gray matter (the vertical black and white bands in Figure 1), since they are spatially non-specific, are unlikely to be categorized by ICA as distinct components and are instead likely to be placed in many component timeseries.

The presence of global signals in multiple components is easily seen. SI Appendix, Fig. S2 shows all component timeseries for two subjects. The global black and white bands are present in multiple BOLD-like component timeseries, and these timeseries correlate highly with one another and with the mean cortical signal of optimally combined data. Online Video 1[7] shows such plots for each ME subject, all showing the same phenomenon.

Quantitatively, 26% $\pm$ 16% of BOLD-like components (3.9 $\pm$ 2.7 in raw numbers) correlate at $r > 0.5$ with the mean optimally combined cortical signal, whereas 1% $\pm$ 1% (0.3 $\pm$ 0.7) of non-BOLD-like components display such correlations. If $r > 0.3$ is the criterion, these numbers are 60% $\pm$ 17% (9.3 $\pm$ 4.2) versus 3% $\pm$ 3% (1.6 $\pm$ 2.0). On average, retained BOLD-like component timeseries correlate with the optimally combined mean cortical signal at $r = 0.32 \pm 0.24$ ($p = 0$), whereas discarded non-BOLD-like components correlate at $r = 0.00 \pm 0.13$ ($p = 0.6$), a highly significant difference ($p = 0$). The mean cortical signal in optimally combined data correlates with the mean cortical signal of ME-ICA denoised data at $r = 0.95 \pm 0.08$, indicating that the global signal is largely preserved from undenoised data to ME-ICA denoised data.

Global BOLD signals are truly R2*-dependent signals

To ensure that the vertical black bands in the BOLD-like components are R2* signals and not brain-wide S0 signals that were inadequately separated from R2* signals due to the spatial criteria of ICA, all datasets also underwent monoexponential fitting of S0 and R2* at every voxel and timepoint. This procedure is labeled FIT and yields "S0-timeseries" and "R2*-timeseries". SI Appendix, Fig. S3 illustrates the separation of S0 vs R2* signals in a subject by both ME-ICA and the FIT procedure. Although the FIT procedure is noisier, it has no spatial constraints, and identifies similar kinds of distinctions as ME-ICA. Importantly, the same prominent widespread signals are identified as R2* signals by both approaches. Online Video 1 shows versions of SI Appendix, Fig. S3 for all 87 ME subjects that completed ME-ICA and FIT (1 dataset failed ME-ICA and 1 dataset failed FIT).

Across all subjects, the mean signal across cortical voxels is nearly the same in the "optimally combined", "multi-echo ICA denoised", and "R2*-timeseries" images: mean signals in the R2*-isolating images are highly correlated (red cells of SI Appendix, Table S1). Additionally, mean

[7] Videos are at: www.jonathanpower.net/paper-multiecho.html.

signals in the S0-isolating images are highly correlated (red cells). In contrast, correlations between mean signals of R2*-isolating and S0-isolating images are much lower (yellow cells, all p's < 10e-5 compared to the red cells). SI Appendix, Table S1 indicates that global signals are mostly R2* signals, and that global variance is partitioned similarly into R2* and S0 elements by ME-ICA and FIT procedures.

Methods to remove brain-wide signals

Two main options exist for removing respiratory BOLD signals: to externally measure respiration and model its consequences in BOLD signals, or to use data-driven methods to identify and remove respiratory BOLD signals. In this paper, we pursue the data-driven option because the 87 multi-echo-denoised ME datasets have no physiological records and only 19 NA scans with physiological records successfully underwent multi-echo denoising and we therefore would be able to make only limited statements about the efficacy of physiological models in removing respiratory signals. Additionally, extant model-driven approaches have been reported to be inadequate methods of removing respiratory variance (3).

The characteristics of respiratory signals make them unlikely to be removed by many common data-driven denoising approaches. As already stated, spatial ICA does not isolate global signals as distinct components because they are not spatially specific. Frequency filtering cannot eliminate global respiratory signals because they are within the typical pass-band of interest of resting state studies (common pass-bands are 10-100 seconds; global respiratory modulations span 20-40 seconds or more (7, 15)). And common nuisance regressions do not remove the global signals: Figures S5 and S6 show respiratory signals persisting in NA subjects despite regression of variance in respiratory models and regression of motion, mean white matter, and mean ventricle signals. SI Appendix, Fig. S7 shows global signals (which are likely respiratory) persisting in ME subjects after such regressions. Online Video 3 shows such plots for all 87 ME subjects. Global signals before and after such regression correlate at $r = 0.72 \pm 0.11$ across ME subjects.

None of the above approaches removed respiratory signals adequately from data. Several approaches can remove global respiratory signals along with all other global signals. Mean signal regression, by definition, is one such approach. Depending on how nuisance masks are defined, regression of white matter mean signals or white matter PCA signals will yield nuisance regressors either nearly identical to the global signal (if superficial voxels are used, SI Appendix, Fig. S11) or at very highly correlated to the global signal (if deeper signals are used, such as in Figures S12 and S13). Robust PCA and GODEC can both identify low-rank spaces that effectively contain most or all global signals. All of these approaches produce similar correlation matrices that are nearly zero-centered (SI Appendix, Fig. S13). All of these approaches largely eliminate group-level dependence of correlations on motion. These approaches achieve a common goal but by distinct means, and, though broadly the correlation matrices they produce are similar, and they also share similar motion-related properties, there are distinctions between the matrices, as mentioned in the main text. These distinctions deserve further attention but a full exploration of these differences is beyond the scope of this manuscript.

Online Supplemental Discussion

Multi-echo signals can be used in multiple ways for denoising purposes

The basic utility of multi-echo signals is in their decay properties, which in principle allow one to separate S0 from R2* influences in signals. These decay properties can be isolated or used in multiple ways, each with particular advantages and disadvantages.

One approach is to fit a monoexponential decay to the observed signal at each voxel, a procedure called FIT in this paper. The FIT method is attractive for its direct and unambiguous treatment of the signals. However, this method is limited by noise in the signals (smaller voxels with higher thermal noise will yield poorer fits) and by the number of echoes obtained (fit accuracy scales with the number of echoes). In both of these respects the ME data were more tractable to FIT than the NA data: the ME voxels were roughly 2.5 times larger than the NA voxels, and 4 echoes were available in the ME data compared to 3 in the NA data. Accordingly, fits were visibly better (less noisy) in the ME data compared to the NA data (see Online Videos 1 and 2; or see the "static" in SI Appendix, Fig. S3 in the FIT procedure compared to ME-ICA). One could reduce noise in fits by constraining fits by priors (e.g., temporal priors), but this step would essentially amount to applying a filter to the data and we chose not to do that in this report.

A related approach is the dual-echo procedure in which signal at a very early echo time is regressed from a typical echo time at each voxel in an attempt to correct the typical echo time for S0 changes (which should be well-captured at the early echo time). This approach has the advantage of being simple to implement with little alterations to sequences needed, but it also does not fully isolate S0 from R2* effects, and has only partial efficacy in removing motion artifact in data (16).

Another approach is to group signals across voxels in various ways prior to examining decay properties. Multi-echo ICA is an example of this approach (5). Such procedures benefit from suppressing thermal (or other) noise by averaging many signals prior to calculating signal decay properties, but they also inherit the limitations of the grouping procedure. For example, in this paper we demonstrated the tendency of spatial ICA to place global signals into multiple components. Another limitation of this approach is that dimensionality reduction is needed and does not succeed comparably across data with different properties. For example, the ME data (the data on which the ME-ICA procedure was developed) usually underwent ME-ICA with consistent separation of BOLD and non-BOLD signal in ways that were congruent with FIT's separations, whereas the NA datasets in several instances either failed ME-ICA completely or placed some BOLD-like signals into non-BOLD-like components. This behavior was easy to identify because global signals that were synced to changes in respiration appeared in non-BOLD-like components, whereas FIT identified the global signals as R2* signals. Adjustment of parameters in the algorithm can improve performance across datasets with various characteristics, but the need for such adjustment is a limitation of the approach.

We mention the limitations of FIT and ME-ICA in order to give a full account of our "user experience", not to dissuade readers from using the techniques. In fact, we are enthusiastic about both of these approaches. A principal attraction of these approaches over single-echo approaches to denoising is that multi-echo approaches leverage the physical properties of the signal, rather than assumed features of the data, to separate "signal" from "noise". For example, an ICA procedure that bins signals by decay properties is fundamentally different than an ICA procedure that bins signals by their spatial or temporal features – it is not expected that the

former features would differ across clinical populations (e.g., stroke patients versus typical subjects), whereas the latter features might differ.

References

1. Kundu P*, et al.* (2013) Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proc Natl Acad Sci U S A* 110(40):16187-16192.

2. Power JD (2017) A simple but useful way to assess fMRI scan qualities. *Neuroimage* 154:150-158.

3. Power JD, Plitt M, Laumann TO, & Martin A (2017) Sources and implications of whole-brain fMRI signals in humans. *Neuroimage* 146:609-625.

4. Power JD, Plitt M, Kundu P, Bandettini PA, & Martin A (2017) Temporal interpolation alters motion in fMRI scans: Magnitudes and consequences for artifact detection. *PLoS One* 12(9):e0182939.

5. Kundu P, Inati SJ, Evans JW, Luh WM, & Bandettini PA (2012) Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *Neuroimage* 60(3):1759-1770.

6. Birn RM, Diamond JB, Smith MA, & Bandettini PA (2006) Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *Neuroimage* 31(4):1536-1548.

7. Chang C & Glover GH (2009) Relationship between respiration, end-tidal CO2, and BOLD signals in resting-state fMRI. *Neuroimage* 47(4):1381-1393.

8. Behzadi Y, Restom K, Liau J, & Liu TT (2007) A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 37(1):90-101.

9. Zhou T TD (2011) GoDec: Randomized low-rank & sparse matrix decomposition in noisy case. *Proceedings of the 28th International Conference on Machine Learning*.

10. Power JD, Schlaggar BL, & Petersen SE (2015) Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage* 105:536-551.

11. Power JD*, et al.* (2014) Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* 84:320-341.

12. Satterthwaite TD*, et al.* (2013) An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* 64:240-256.

13. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, & Petersen SE (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59(3):2142-2154.

14. Power JD*, et al.* (2011) Functional network organization of the human brain. *Neuron* 72(4):665-678.

15. Birn RM, Smith MA, Jones TB, & Bandettini PA (2008) The respiration response function: the temporal dynamics of fMRI signal fluctuations related to changes in respiration. *Neuroimage* 40(2):644-654.

16. Bright MG & Murphy K (2013) Removing motion and physiological artifacts from intrinsic BOLD fluctuations using short echo data. *Neuroimage* 64:526-537.

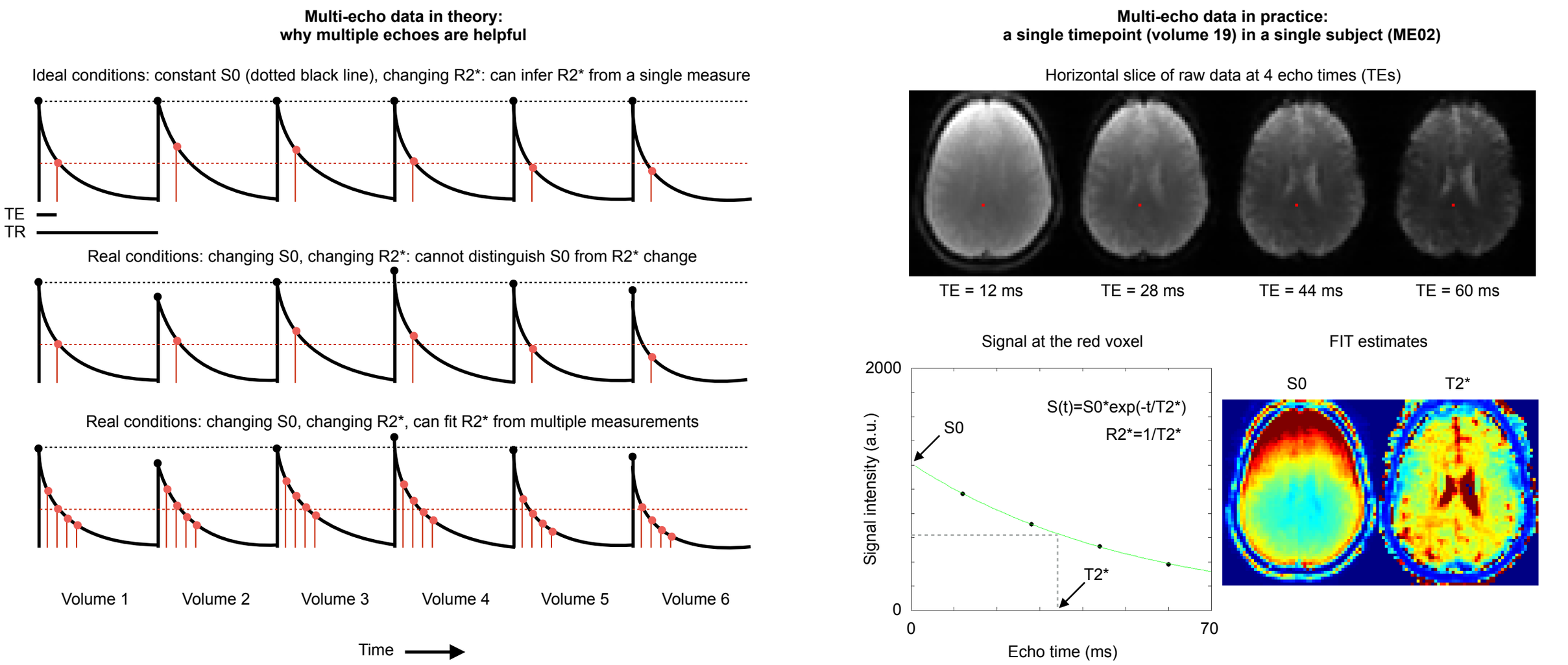


Figure S1: Using multi-echo data to separate S0 from R2* effects. At left, 3 schematic timeseries of signal intensity at a voxel over 6 volumes, showing how single-echo sequences cannot distinguish S0 from R2* effects in fMRI signals. In multi-echo sequences, S0 and R2* can be derived from the multiple readouts of the fMRI signal per excitation. At right, real data at 4 echo times are shown from a single timepoint (volume) of a single subject, illustrating how S0 and R2* are derived in practice. The green fitting curve and the colored heatmaps reflect the FIT procedure in operation.

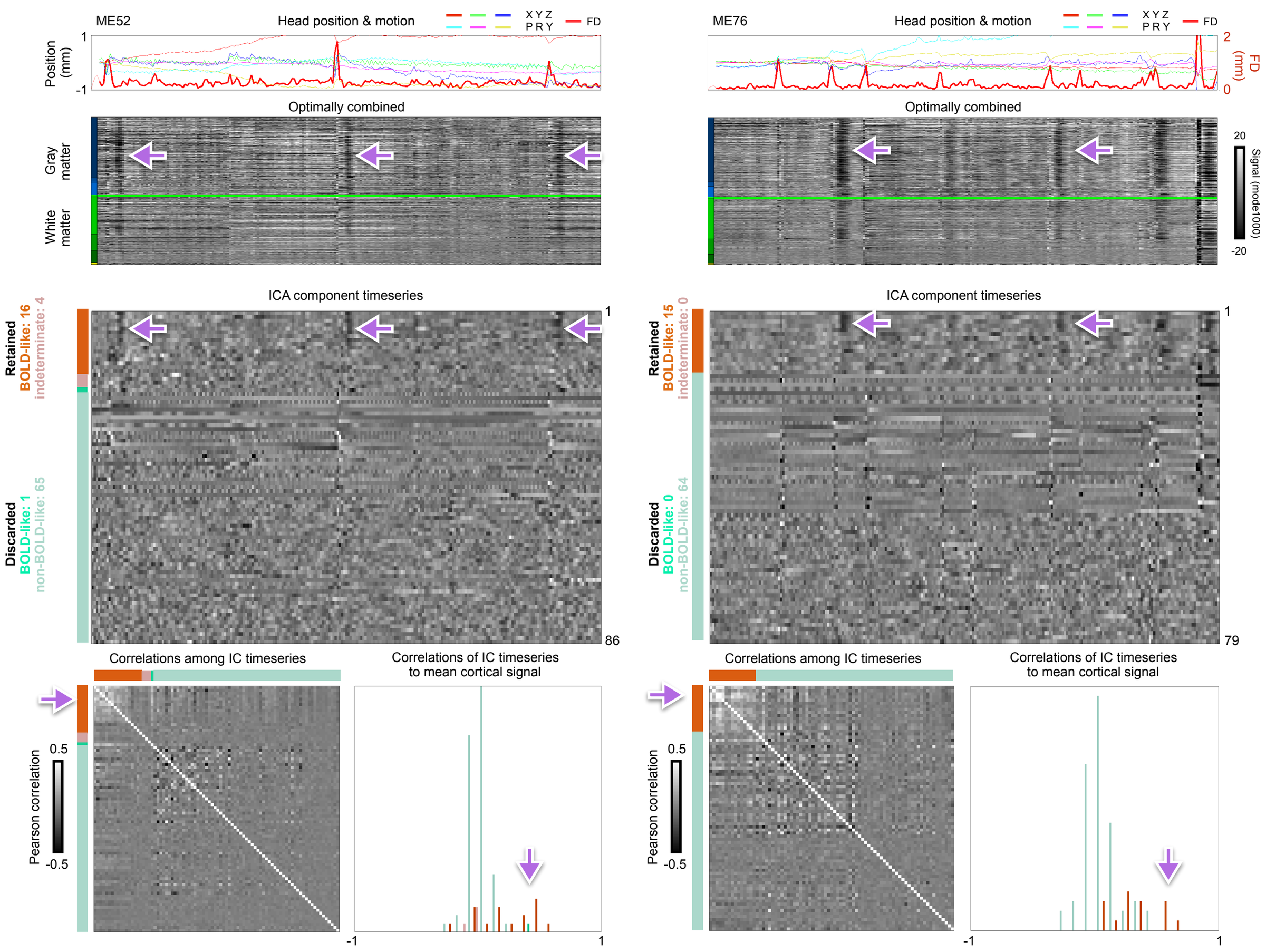


Figure S2: Brain-wide signals appear in multiple BOLD-like ICA component timeseries. Data for 2 ME subjects are shown. Top panels are as in Figure 1. In the middle panels, all ICA component timeseries are shown with their ME-ICA categorizations in the color bar. Multiple components BOLD-like timeseries include the global signal, see the purple arrows for examples. At bottom, Pearson correlations among the component timeseries, and a histogram of the correlations of component timeseries to the mean cortical signal are shown.

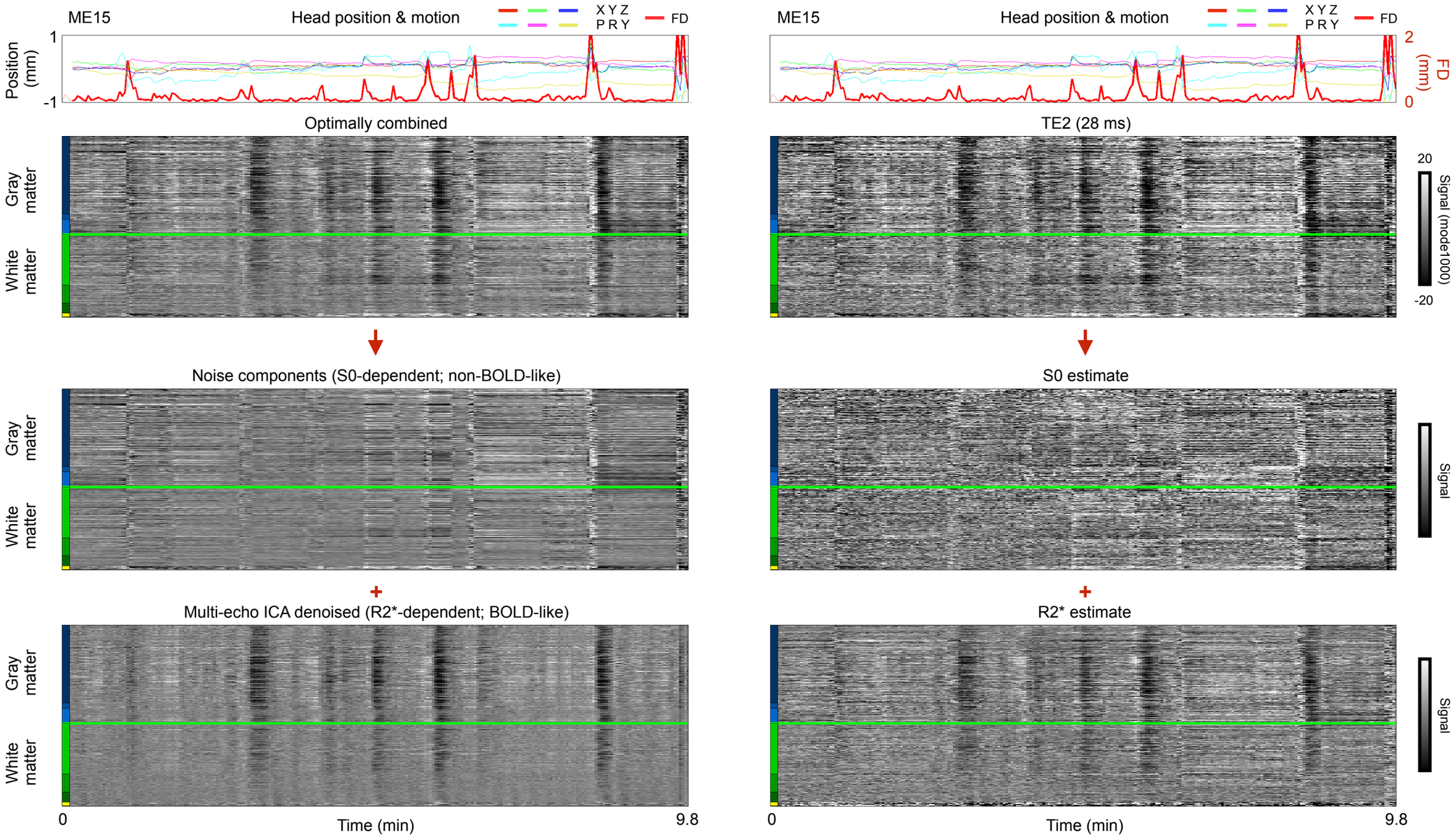
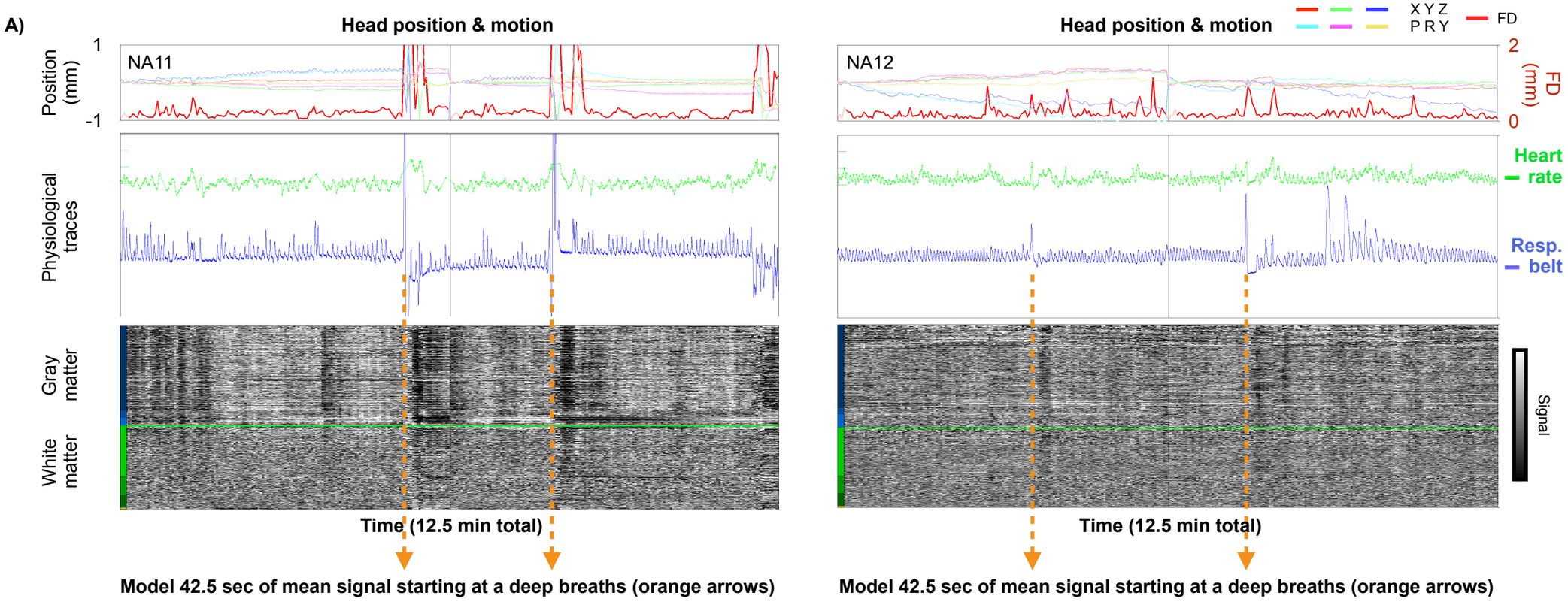


Figure S3: Multiple techniques indicate that global signals are BOLD-like signals. Both left and right columns represent the same ME subject. At left, ME-ICA denoising, at right, the FIT procedure.



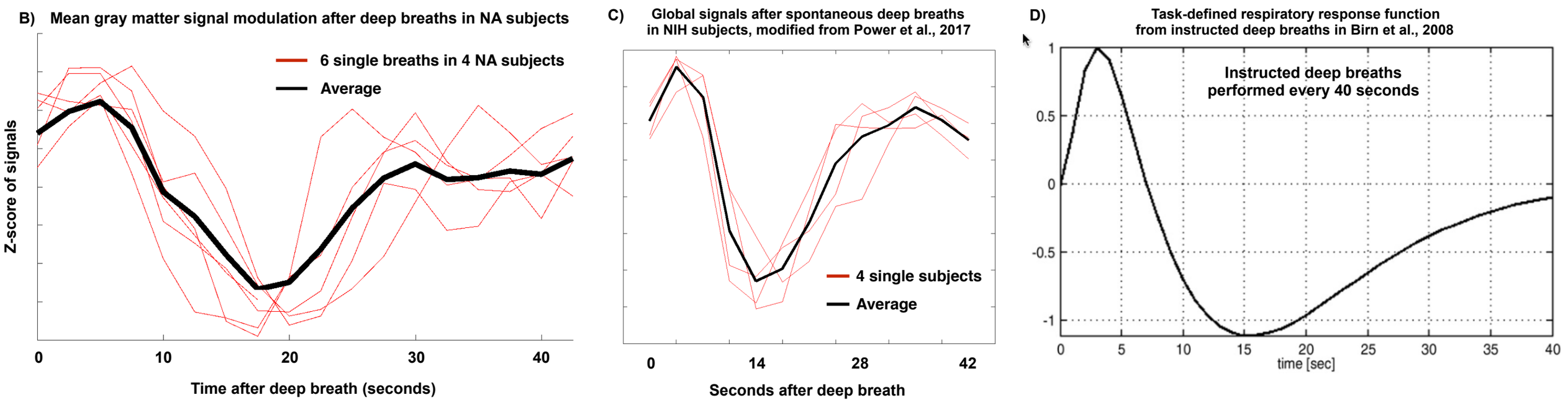


Figure S4: The timecourse of fMRI global signal modulation by deep breaths in NA data resembles respiratory response functions defined in other publications. A) 4 scans from 2 NA subjects are shown in the "gray plot" format. Orange arrows denote four large breaths. The timeseries in the heat maps are the T2* FIT timeseries (ME-ICA didn't run completely on some of these scans). **B)** Six instances of large, relatively isolated breaths are apparent in the NA data (four are shown in (A)). The z-scores of the mean gray matter signals in the 42.5 seconds following the breaths are shown in red and the average z-score in black. **C)** For comparison, post-deep-breath signals in a different dataset from a different publication are shown. The responses in (C) were pristine examples drawn from nearly 90 scans and represent well-isolated breaths with little head motion, whereas the responses in (B) are less pristine (only 24 scans existed, and thus breaths with motion or that had some respiratory modulation afterwards had to be included). Reprinted from Neuroimage, 146, Power et al., Sources and implications of whole-brain fMRI signals in humans, 609-625, 2017, with permission from Elsevier. **D)** A respiratory response elicited by instructed breaths. Reprinted from Neuroimage, 47, Chang and Glover, Relationship between respiration, end-tidal CO2, and BOLD signals in resting-state fMRI, 1381-1393, 2009, with permission from Elsevier.

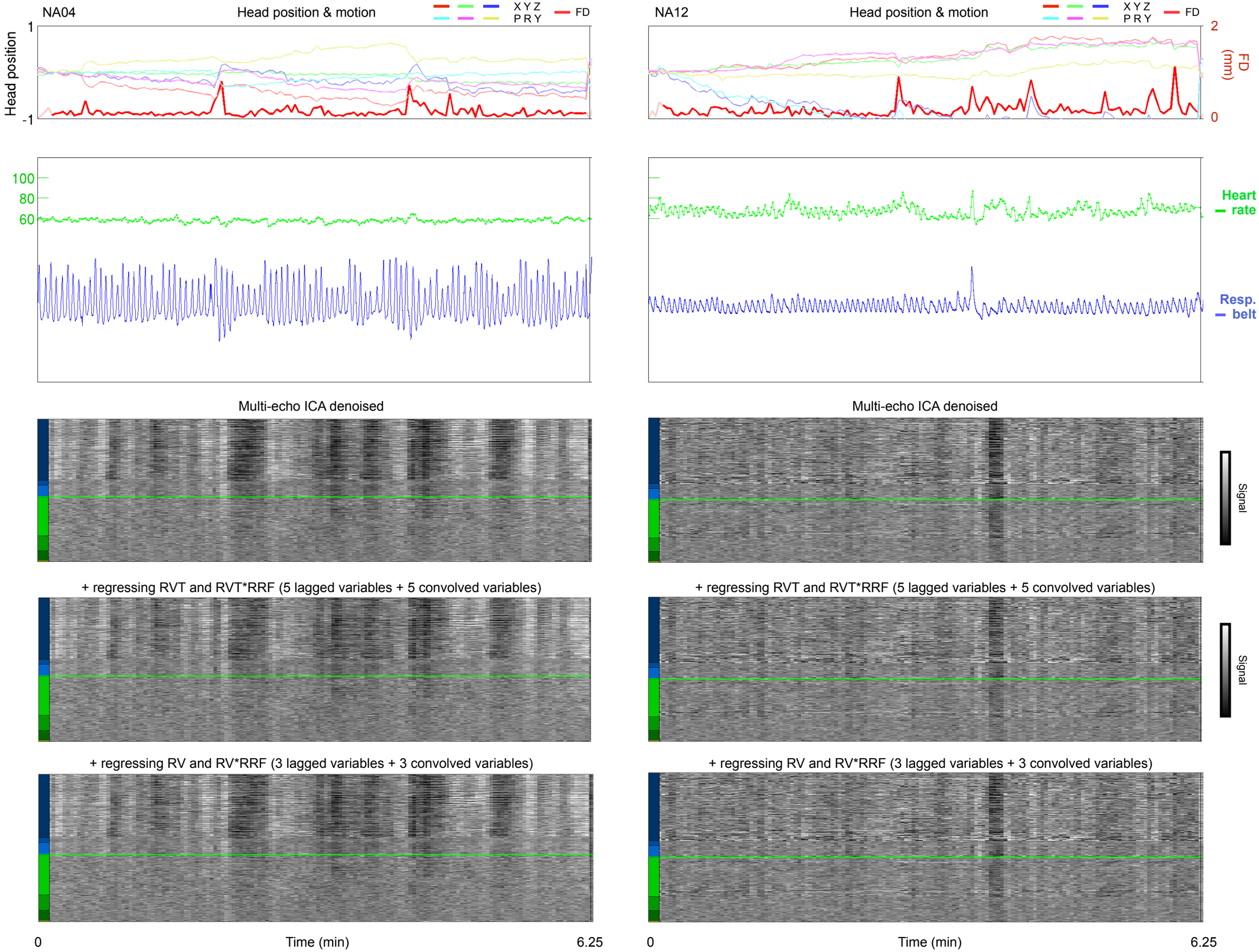


Figure S5: Respiratory fluctuations in NA data remain after attempts to model respiratory variance. Plots as in Figure 2 but for different NA subjects, using respiration volume per unit time (RVT) and respiratory variance (RV) models to remove respiratory variance after ME-ICA (following methods developed in [Birn, 2008; Chang, 2009]). The respiratory response function (RRF) used was obtained empirically from deep breaths of several subjects in [Power, 2017].

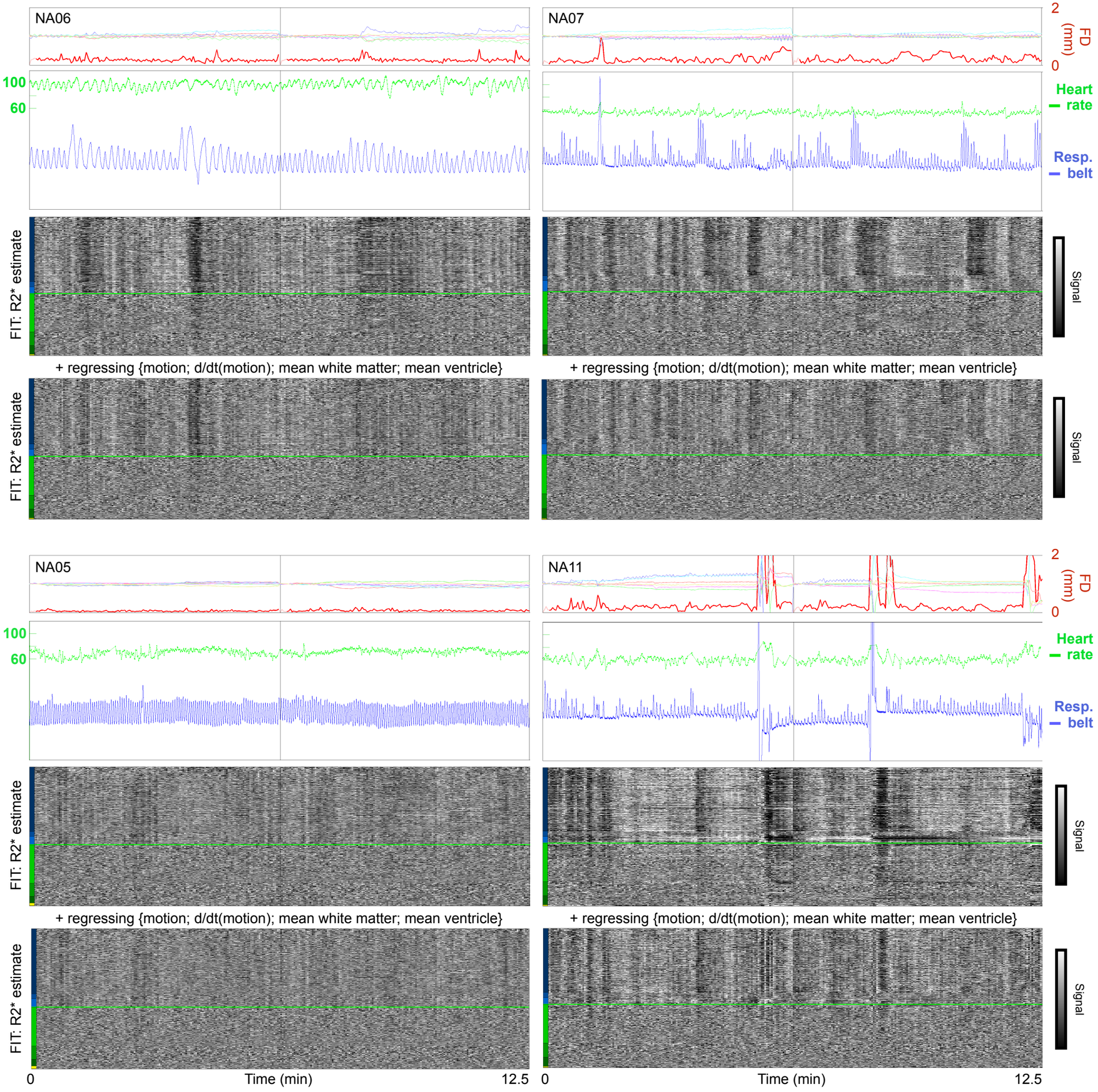


Figure S6: Respiratory fluctuations in NA data remain after common nuisance regressions. Plots as in Figure 2, but R2*-timeseries are also shown after removing 14 nuisance variables by multiple linear regression (12 motion, 1 mean white matter, and 1 mean ventricle signal).

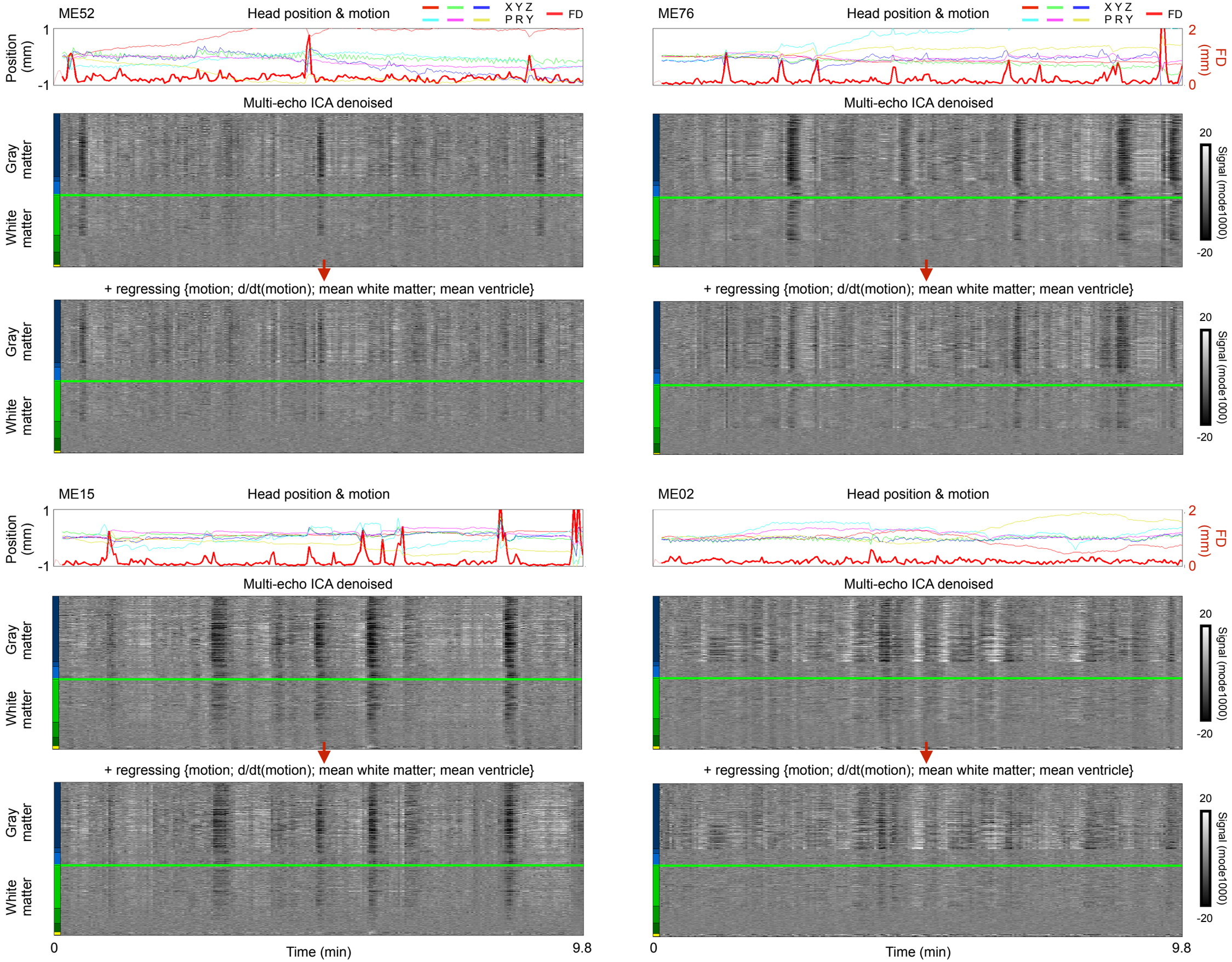


Figure S7: Global fluctuations in ME data remain after common nuisance regressions. Plots similar to those in Figure S6, but now for ME subjects after ME-ICA denoising.

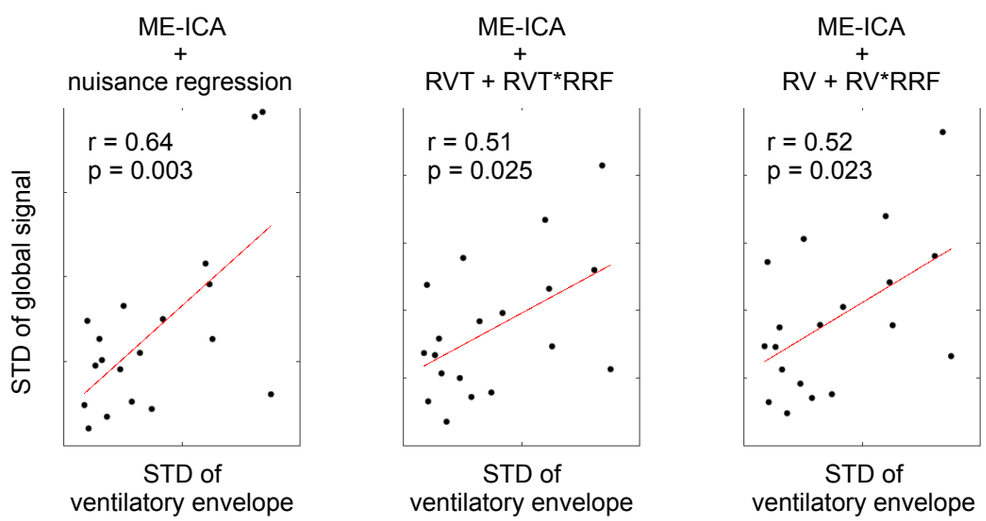


Figure S8: Global signal variance in NA data remains linked to variability in respiratory patterns after common nuisance regressions and attempts to model physiologic variance. Correlations are calculated as in Figure 2, but for processing shown in Figures S5 and S6.

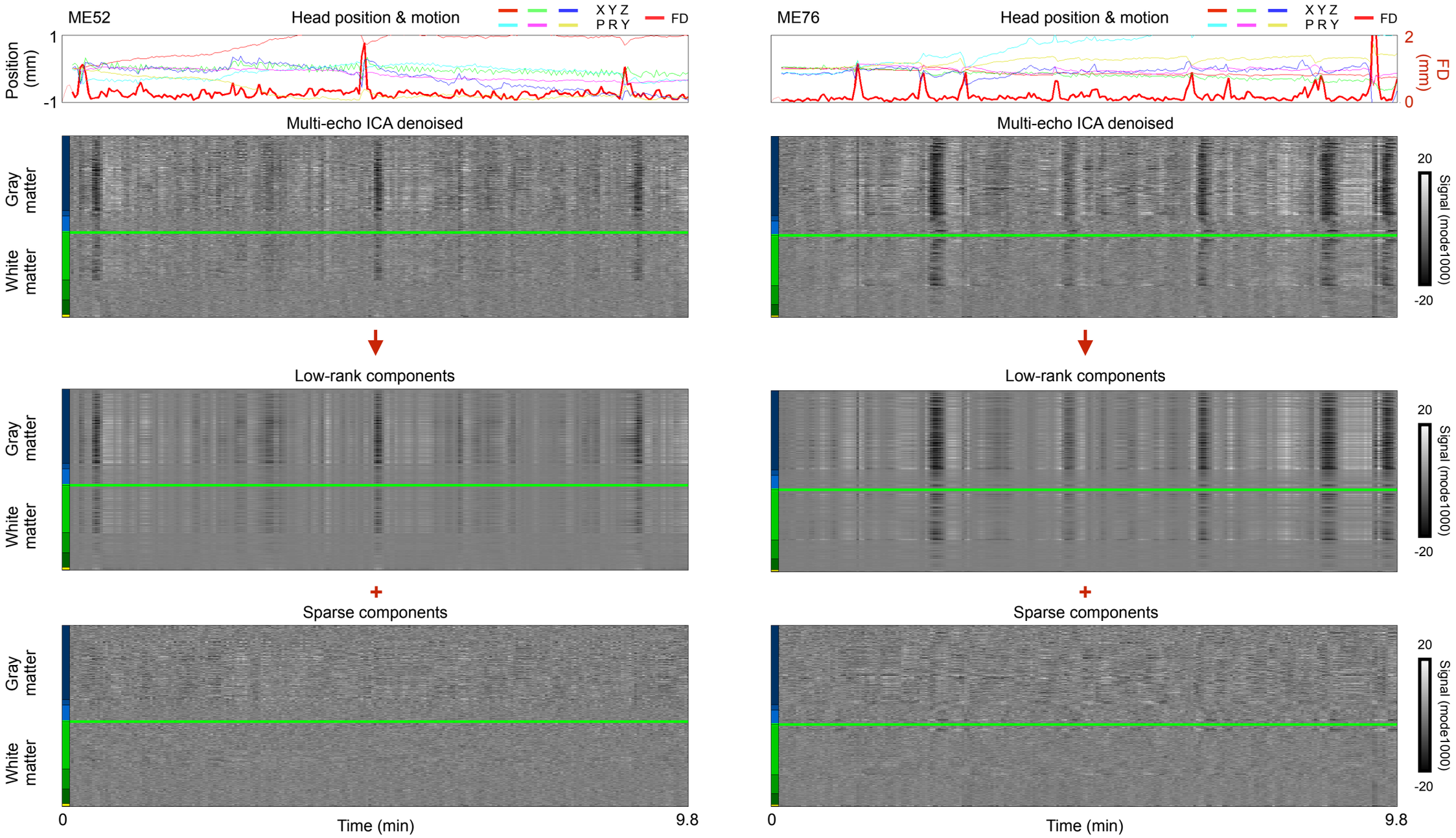


Figure S9: Further examples of GODEC. At top, the “multi-echo denoised” data from Figure 1. Middle and bottom show the splitting of these signals into low-rank and sparse elements by GODEC. The low-rank elements include the global signals likely due to to respiratory patterns.

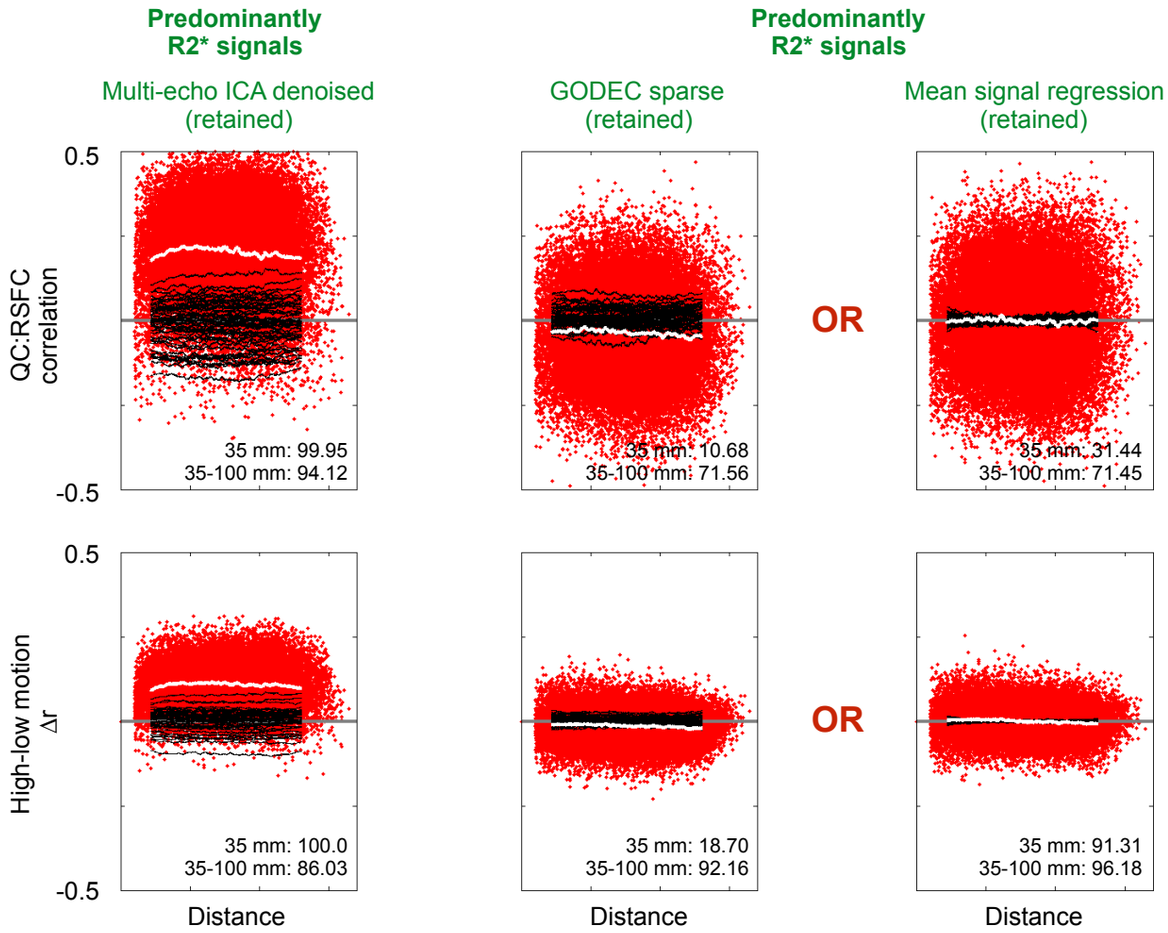


Figure S10: Censoring high-motion volumes can attenuate or eliminate significant distance dependence. These analyses mirror those in Figure 4, except now volumes with FD > 0.2 are withheld from calculations (the same volumes shown in the scrubbing analysis of Figure 4). Note the decreased permutation ranks for distance dependence compared to Figure 4.

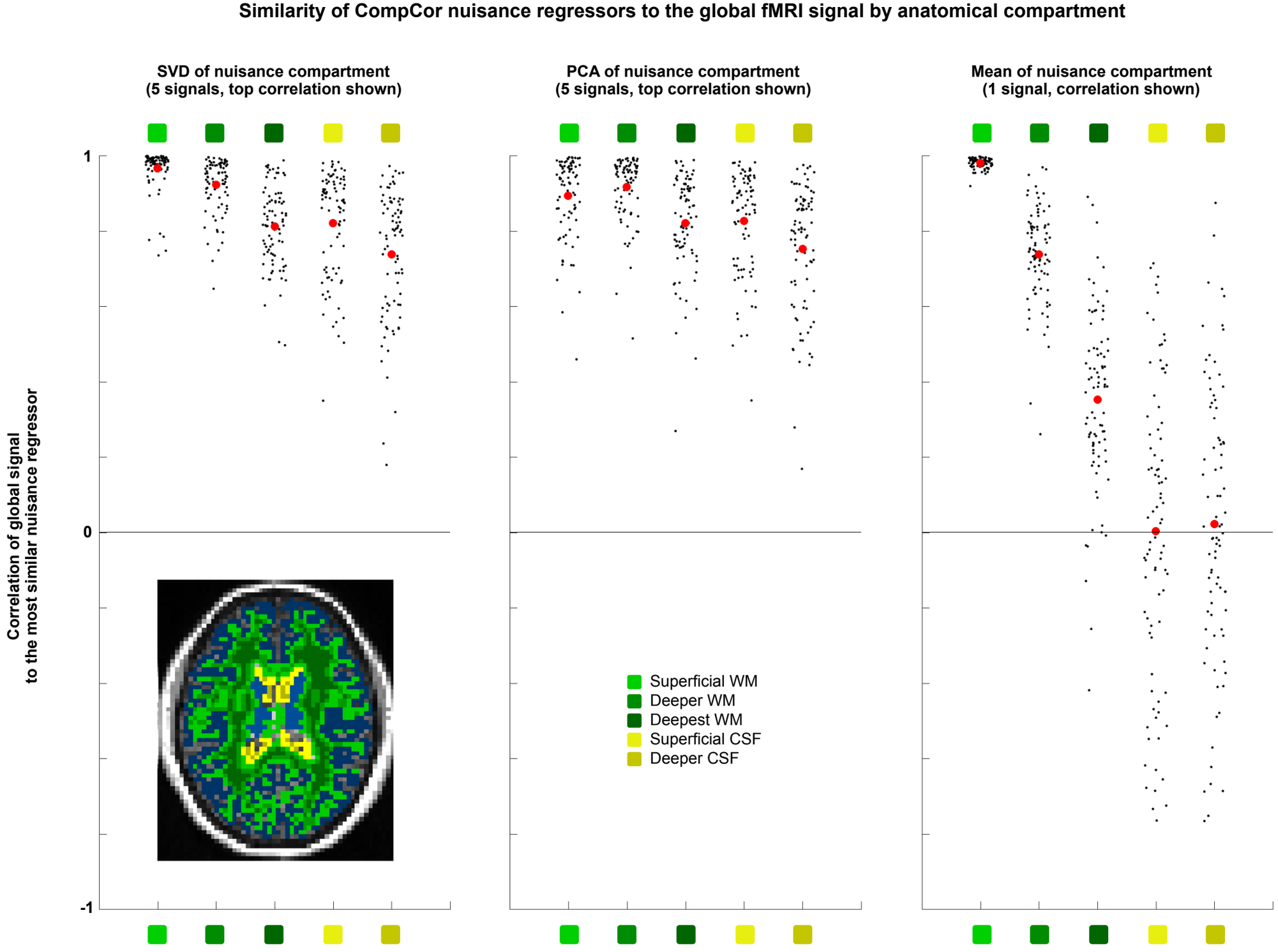


Figure S11: Global signals can be captured by CompCor. 5-signal SVD and PCA of signals within the indicated masks was performed, correlations of all signals with the global signal were calculated, and the top correlation is plotted in each subject of the ME cohort. Mean signals of the compartments were also calculated and their correlations to the global signal are presented. The red dot shows the mean value in a given analysis. Approximately, the superficial white matter contains 60% of the voxels in the white matter, the deeper mask 30%, and the deepest compartment 10%.

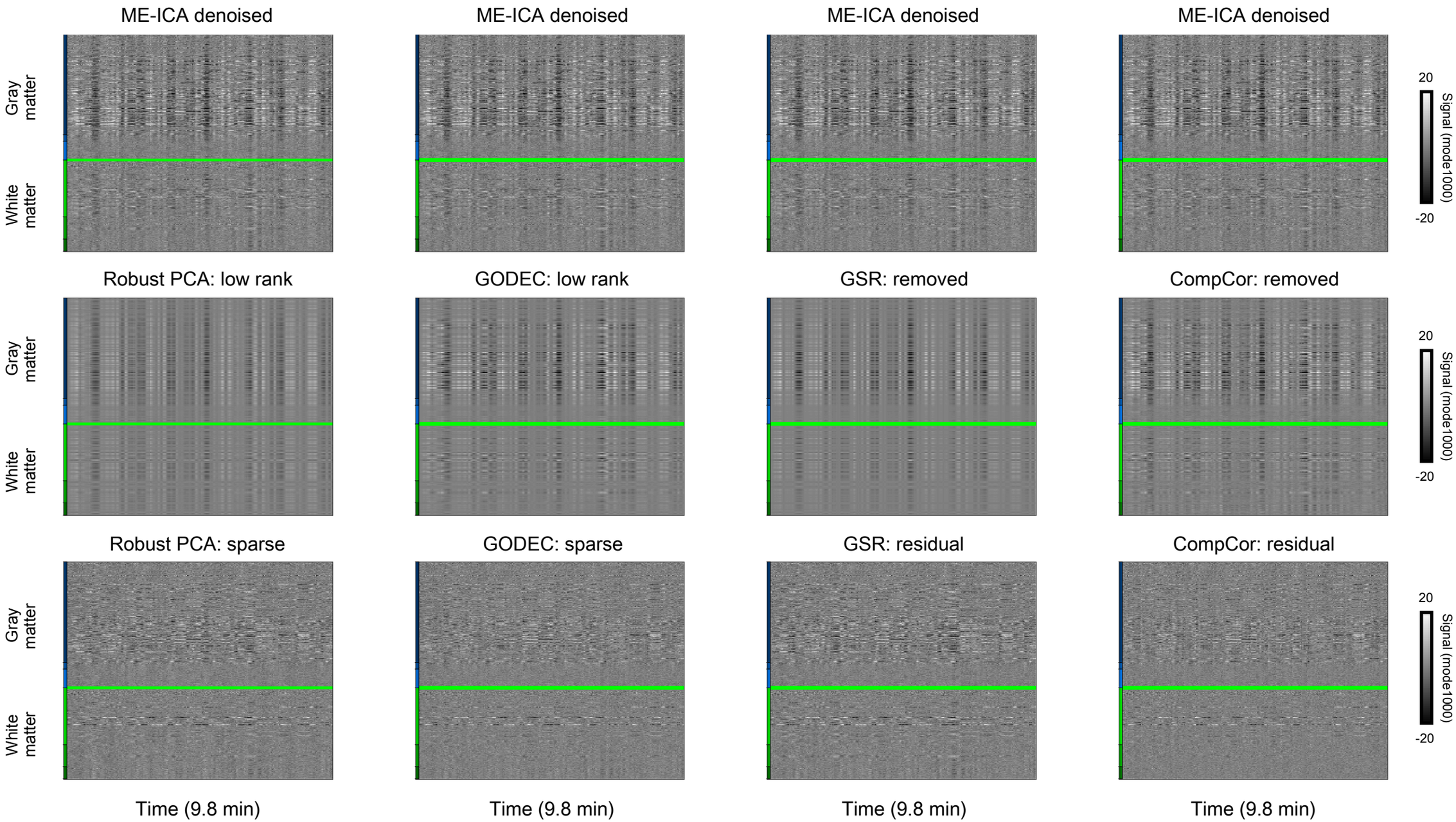


Figure S12: Illustration of global signal removal by multiple methods. For a single subject, ME-ICA denoised data are shown at top, the global signal removed by robust PCA, GODEC, mean signal regression, and CompCor are shown in the middle row, and the residual timeseries are shown at bottom. Similar plots for all ME subjects are found in Online Movie 6.

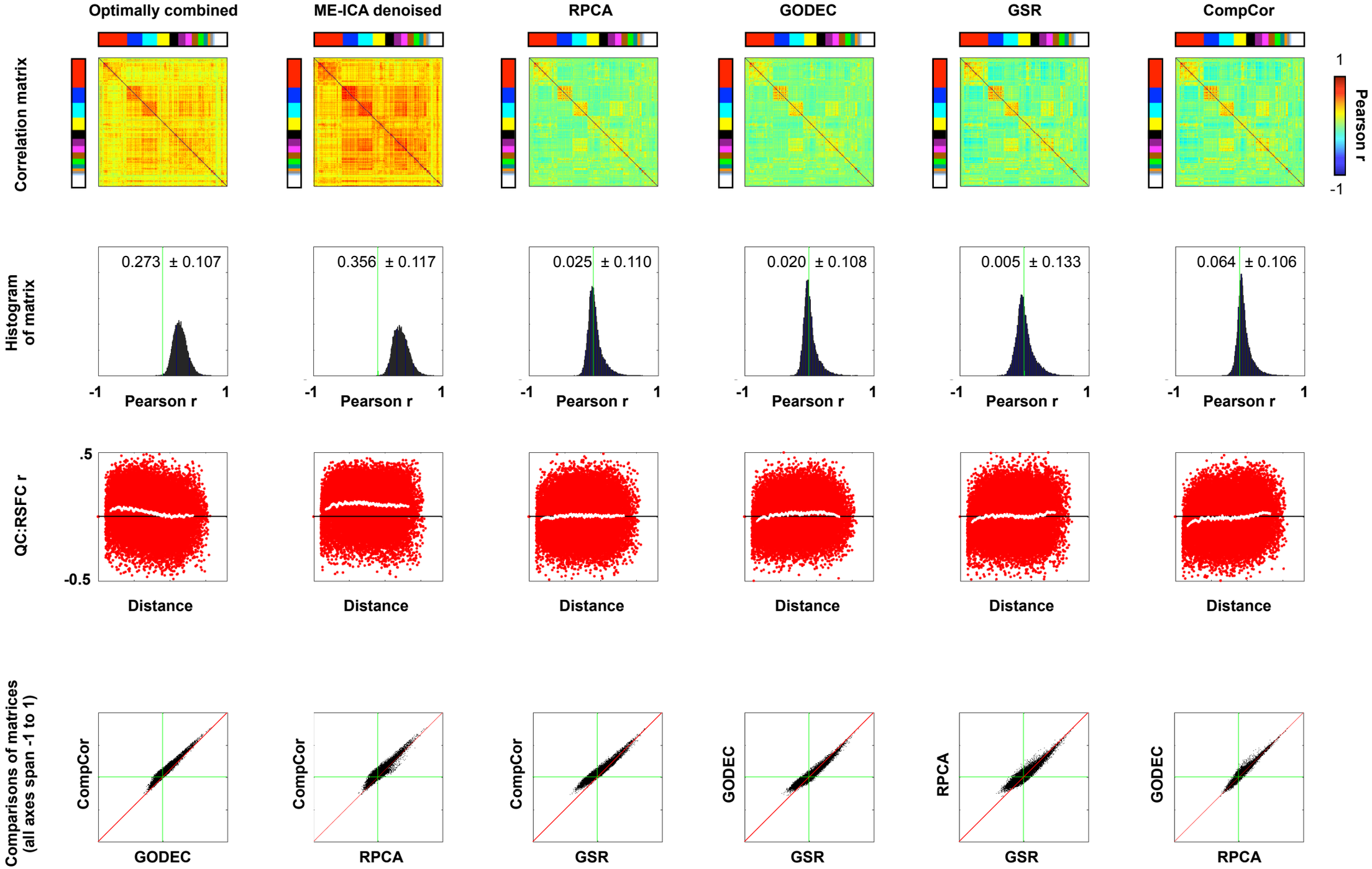


Figure S13: Summary of multiple methods to remove global signals. Top row shows mean correlation matrices (excluding subjects who failed a method). Next row shows histograms of the matrices. The third row shows QC:RSFC correlations produced from these data. The timeseries used are censored at FD > 0.2 mm as in Figure S10. At bottom, mean correlation matrices are plotted against one another. GODEC and RPCA produce similar results, CompCor yields generally higher correlations than all methods, and GSR yields higher high correlations than RPCA or GODEC and more negative negative correlations than the other methods.

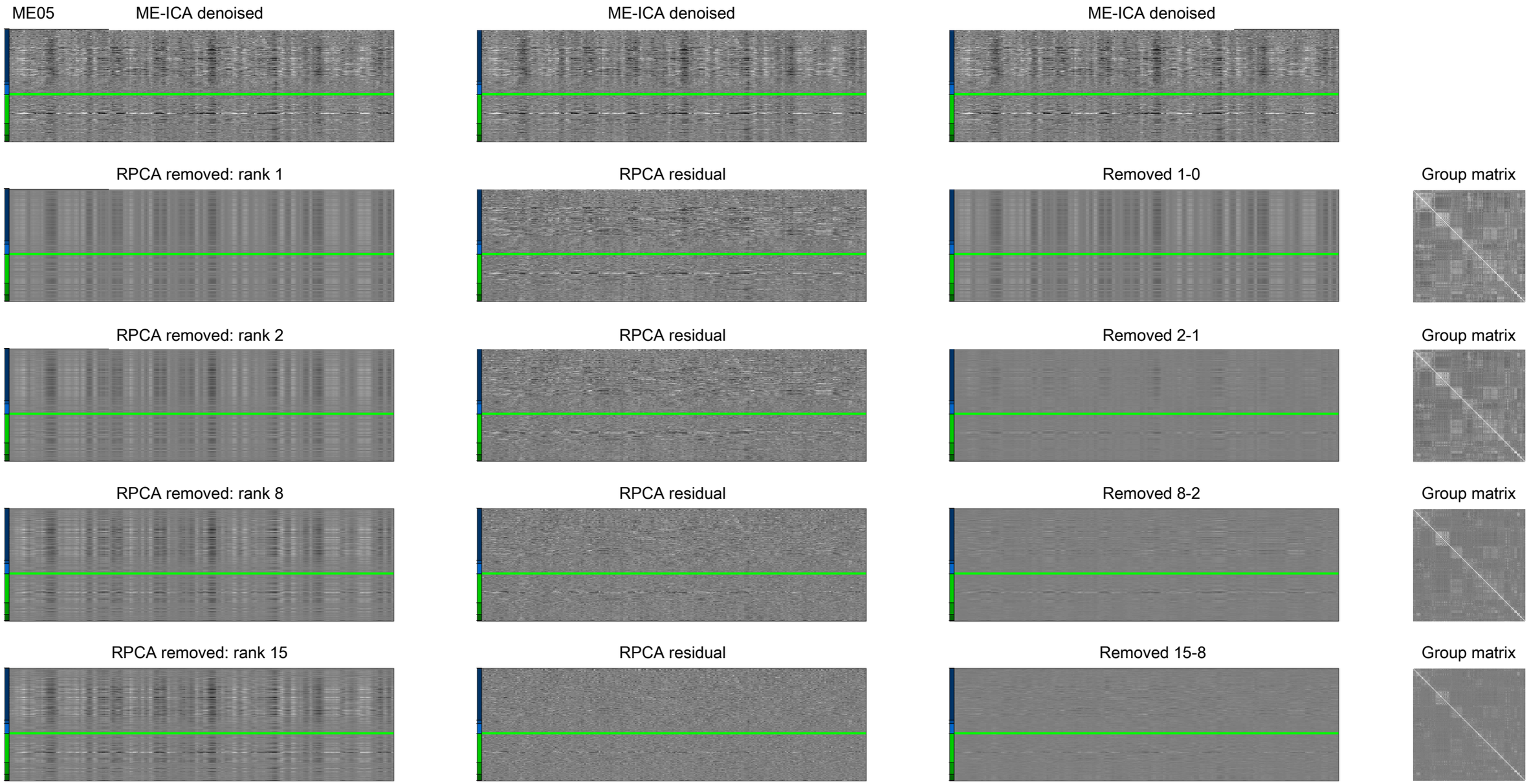


Figure S14: Robust PCA tuning illustration. Top row shows ME-ICA denoised data for reference. The next four rows illustrate the action of robust PCA in a single subject's gray plots and in group correlation matrices. From top to bottom, RPCA tuning parameters were chosen that yielded lower- to higher-rank low-rank partitions. At left, the variance removed by RPCA. At middle, the residual timeseries. At right, the differences between variance removed in various rows, illustrating the additional variance removed as the rank of the low-rank partition increases. The first few components remove variance time-locked to global signal changes, but as increasing components are included, the diagonal structure of the matrix is attenuated, indicating removal of sub-global signals (i.e., the "resting state network" signals that create the block diagonal structure).

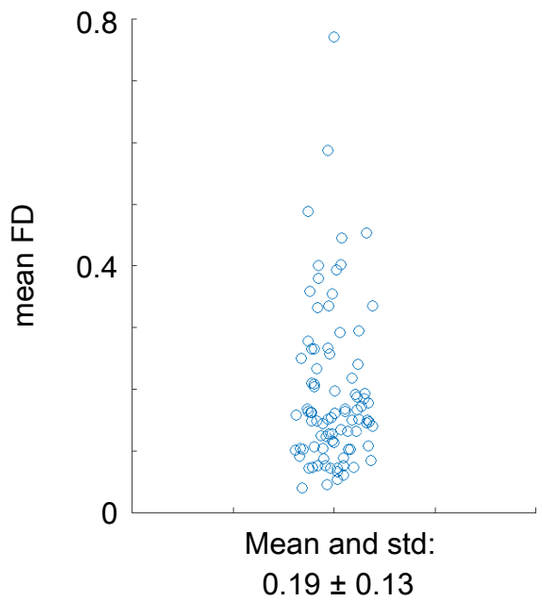


Figure S15: Scatter plot of mean FD values of the 89 ME subjects.

Table S1: mean correlations of various images across ME subjects					
	OC	R2*: MEICA	R2*: FIT	S0: MEICA	S0: FIT
Optimally combined (OC)	1.00	0.95 (0.08)	0.94 (0.10)	0.69 (0.16)	0.60 (0.17)
Multi-echo ICA denoised (R2*: MEICA)	0.95 (0.08)	1.00	0.96 (0.06)	0.39 (0.18)	0.38 (0.16)
FIT R2*-timeseries (R2*: FIT)	0.94 (0.10)	0.96 (0.06)	1.00	0.43 (0.21)	0.24 (0.21)
Multi-echo ICA discarded (S0: MEICA)	0.69 (0.16)	0.39 (0.18)	0.43 (0.21)	1.00	0.78 (0.22)
FIT: S0-timeseries (S0: FIT)	0.60 (0.17)	0.38 (0.16)	0.24 (0.21)	0.78 (0.22)	1.00

Table S2: cohort properties

Cohort	TE spacing (ms)	TR (s)	Run(s)	Acquisition voxel size	Scanner	Scan parameters	Comments
ME N=89	12, 28, 44, 60	2.47	1 x 9.84 min 239 TR/run	3.75 x 3.75 x 4.40 mm (62 mm^3)	Siemens Trio 3T 32-channel head coil Cambridge, England	Interleaved ascending (alt+z2 in AFNI)	Larger voxels means less thermal noise 4 data points for 2-parameter fit of S0 and R2* No physiological records
NA N=12	12, 24.5, 37	2.50	2 x 6.25 min 150 TR/run	3.00 x 3.00 x 3.00 mm (27 mm3)	Siemens Skyra 3T 20-channel head coil Bethesda, MD, USA	Interleaved ascending (alt+z2 in AFNI)	Smaller voxels means more thermal noise 3 data points for 2-parameter fit of S0 and R2* Have physiological records