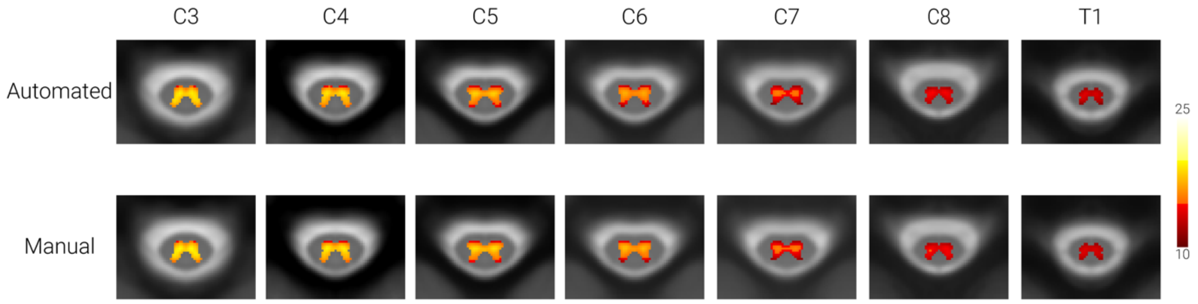

Figure S1. **tSNR maps**. Group-average gray matter tSNR maps of transversal slices at the middle of each segment are shown for the two different shim conditions (top row: Automated; bottom row: Manual). The maps are overlaid onto the PAM50 T2*-weighted template and depict a tSNR range from 10-25.

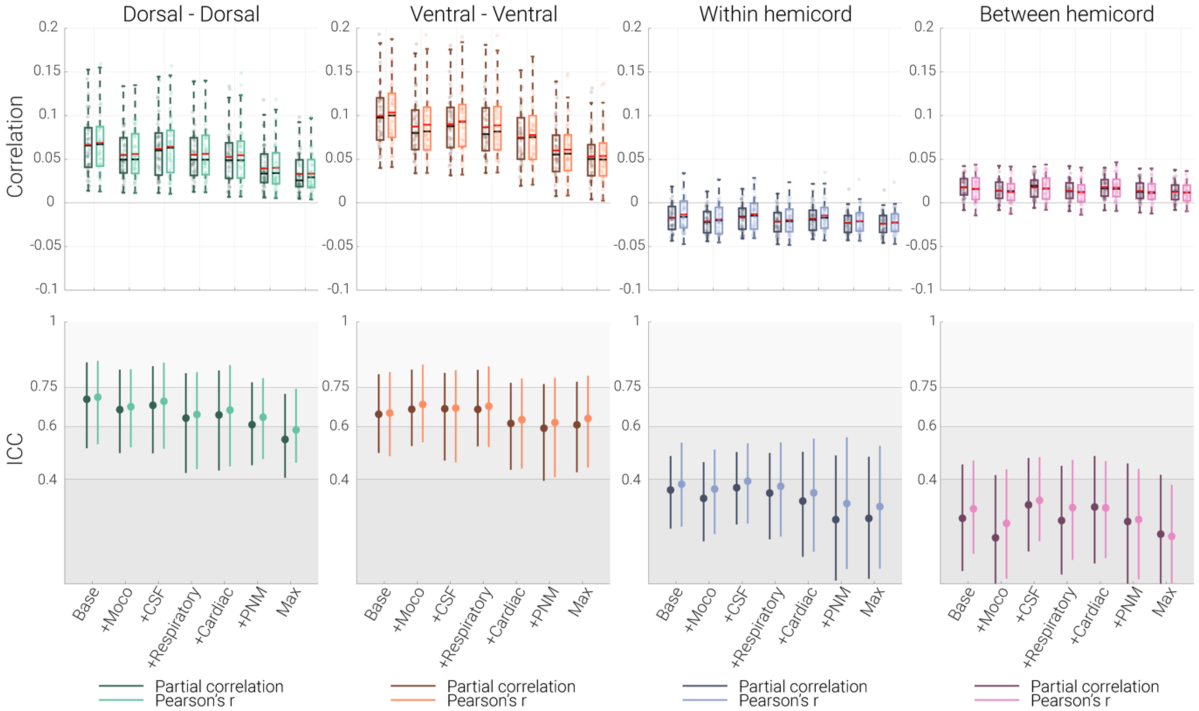

Figure S2. **Partial correlation vs Pearson correlation**. The top panels depict functional connectivity estimates between different ROIs calculated with either partial correlation or Pearson correlation (average across two sessions) using grouped box plots for the seven denoising pipelines. For the box plots, median and mean are denoted by the central black and red marks, respectively. The bottom and top edges of the boxes represent the 25th and 75th percentiles, respectively, with the whiskers encompassing ~99% of the data. The bottom panels depict ICC values for different denoising pipelines with dots and lines denote 95% confidence intervals. The gray scale background reflects the ICC ranges (as defined by Cicchetti & Sparrow (1981) and Hallgren (2012)): poor <0.4, fair 0.4–0.59, good 0.6–0.74, excellent ≥0.75.

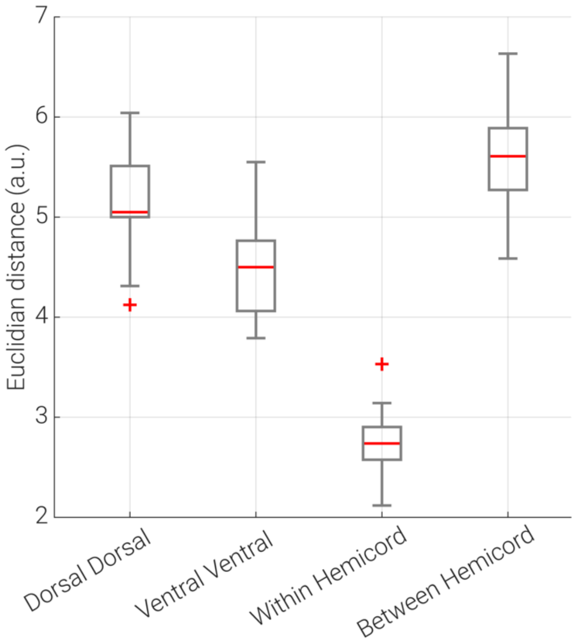

Figure S3. **Euclidian distance between ROIs**. Box plots show the median Euclidian distance between the closest voxels of different ROIs (within each slice) across slices and participants. The median is denoted by the central red line. The bottom and top edges of the boxes represent the 25th and 75th percentiles, respectively, with the whiskers encompassing ~99% of the data, and the outliers are denoted with the red crosses.

Table S1. *Functional connectivity and its reliability after addition of pre-whitening to the maximal denoising pipeline.*

	Dorsal Dorsal	**Ventral Ventral**	**Within Hemicord**	**Between Hemicord**
Maximal	r = 0.03 t = 9.5 p < 0.001 ICC (95% CI) = 0.59 (0.46 – 0.74)	r = 0.05 t = 11.6 p < 0.001 ICC (95% CI) = 0.63 (0.44 – 0.79)	r = -0.02 t = -10.7 p < 0.001 ICC (95% CI) = 0.30 (0.06 – 0.53)	r = 0.01 t = 6.7 p < 0.001 ICC (95% CI) = 0.18 (-0.03 – 0.38)
Maximal + Pre-whitening	r = 0.03 t = 9.6 p < 0.001 ICC (95% CI) = 0.59 (0.46 – 0.74)	r = 0.05 t = 11.7 p < 0.001 ICC (95% CI) = 0.65 (0.47 – 0.80)	r = -0.02 t = -10.6 p < 0.001 ICC (95% CI) = 0.29 (0.06 – 0.52)	r = 0.01 t = 6.7 p < 0.001 ICC (95% CI) = 0.17 (-0.05 – 0.37)

This table depicts functional connectivity and reliability results of each connection for two processing pipelines: the maximal pipeline and the maximal pipeline with the inclusion of FILM pre-whitening. r represents the mean Pearson correlation across participants, t and p represent the t-value and two-tailed family-wise-error corrected p-value from a permutation test (against 0), respectively. ICC (95% CI) represents ICC(2,1) values and 95% bootstrapped confidence intervals.

Table S2. *Functional connectivity and its reliability after addition of white matter regression to the baseline and maximal denoising pipelines.*

	Dorsal orsal	**Ventral Ventral**	**Within Hemicord**	**Between Hemicord**
Baseline	r = 0.07 t = 12.9 p < 0.001 ICC (95% CI) = 0.71 (0.53 - 0.85)	r = 0.10 t = 18.7 p < 0.001 ICC (95% CI) = 0.65 (0.49 – 0.81)	r = -0.01 t = -4.7 p < 0.001 ICC (95% CI) = 0.38 (0.22 – 0.54)	r = 0.02 t = 7.1 p < 0.001 ICC (95% CI) = 0.28 (0.11 – 0.47)
Baseline + WM	r = 0.06 t = 12.9 p < 0.001 ICC (95% CI) = 0.69 (0.51 0.84)	r = 0.10 t = 18.0 p < 0.001 ICC (95% CI) = 0.66 (0.47 – 0.81)	r = -0.01 t = -4.9 p < 0.001 ICC (95% CI) = 0.39 (0.20 – 0.54)	r = 0.02 t = 7.2 p < 0.001 ICC (95% CI) = 0.29 (0.10 – 0.47)
Maximal	r = 0.03 t = 9.5 p < 0.001 ICC (95% CI) = 0.59 (0.46 – 0.74)	r = 0.05 t = 11.6 p < 0.001 ICC (95% CI) = 0.63 (0.44 – 0.79)	r = -0.02 t = -10.7 p < 0.001 ICC (95% CI) = 0.30 (0.06 – 0.53)	r = 0.01 t = 6.7 p < 0.001 ICC (95% CI) = 0.18 (-0.03 – 0.38)
Maximal + WM	r = 0.03 t = 9.6 p < 0.001 ICC (95% CI) = 0.57 (0.44 – 0.73)	r = 0.05 t = 11.5 p < 0.001 ICC (95% CI) = 0.64 (0.46 – 0.80)	r = -0.02 t = -11.1 p < 0.001 ICC (95% CI) = 0.29 (0.05 – 0.54)	r = 0.01 t = 6.7 p < 0.001 ICC (95% CI) = 0.17 (-0.05 – 0.37)

This table depicts functional connectivity and reliability results of each connection for four processing pipelines: the baseline pipeline, the baseline pipeline with the inclusion of white matter regression, the maximal pipeline and the maximal pipeline with the inclusion of white matter regression. r represents the mean Pearson correlation across participants, t and p represent the t-value and two-tailed family-wise-error corrected p-value from a permutation test (against 0), respectively. ICC (95% CI) represents ICC(2,1) values and 95% bootstrapped confidence intervals.

Motivation for estimating correlations in a slice-wise manner

In this study, we calculated correlations for each slice and then averaged the obtained correlations across slices afterwards. While this approach has also been taken by previous studies (e.g. Barry et al., 2014; Conrad et al., 2018), there is also an alternative approach where time series for any given ROI are first averaged across slices and then correlations (between ROIs) are computed from these averaged time series (e.g. Eippert et al., 2017a; Vahdat et al., 2020). There are two reasons why we did not average the time series from all slices, but rather obtained time series correlations on a slice-wise level and then averaged the resulting correlation coefficients in this study.

The first reason is based on the fact that our field of view (composed of 24 slices) encompasses a large number of spinal segments (C3 to T1) and we assume that the segments represent rather separate functional units based on several lines of evidence. First, Weber et al. (2018) did not find strong evidence for between-segment functional connectivity during rest and describe the observed connectivity patterns as more lattice-like, suggesting that different segments might contain separate networks. Second, Eippert et al. (2017) observed that the similarity of within-segment connectivity patterns decreases with segmental distance. Third, Kong et al. (2014) decomposed resting-state data using spatial ICA and observed segment-like components, whose time series were either uncorrelated or even showed slightly negative correlations. While it might thus be reasonable to assume that time-series in neighboring slices might be similar (especially if they belong to one segment), there is clear evidence to suggest this is not the case across all 24 slices.

The second point we want to raise is best explained by a toy example. Let us assume that we have two regions of interest (e.g. left and right dorsal horns) and acquired time series from these regions in two slices, leading to four time series overall. Let us further assume that the correlation between dorsal horns for each slice is 0.3, so that averaging correlations for each slice would give an overall correlation of 0.3 Let us now assume that the time series from both slices are also correlated with each other, say a correlation of 0.1 from each region of one slice to any region of the other slice. If we create time series that have the described properties and average them before calculating the correlation, we obtain a correlation of 0.36 between the dorsal horns. Thus, depending on the overall correlational structure of the network, the correlation of averaged time series can deviate quite substantially from the average of correlations, as the former is capturing the overall network properties instead of pairwise correlations: importantly, depending on the specific network properties, it can over- or underestimate the slice-wise value. In order to explore this issue in more detail, we refer the reader to a markdown file containing simulations with code and explanations (see associated GitHub repo: https://github.com/eippertlab/restingstate-reliability-spinalcord).