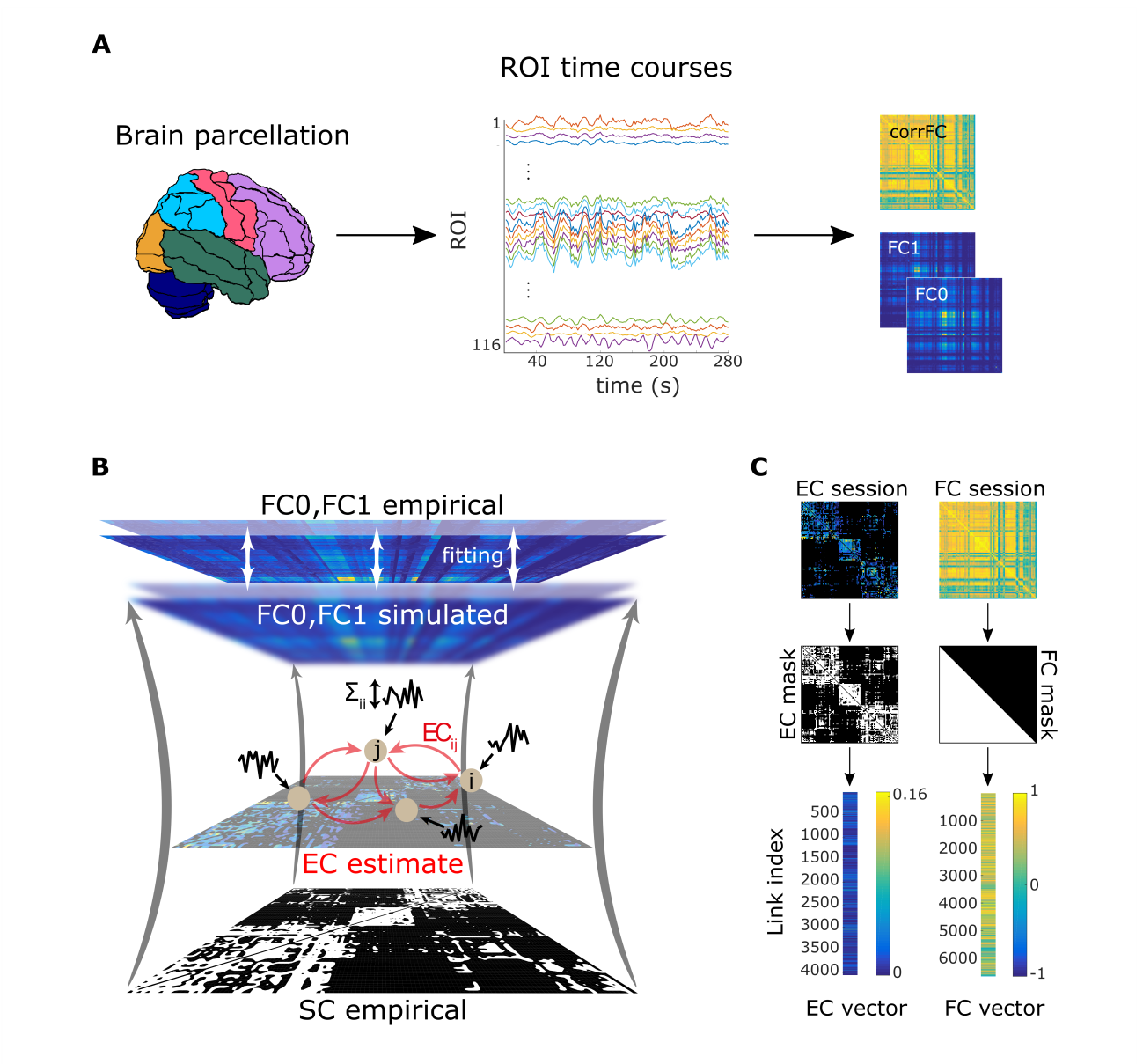
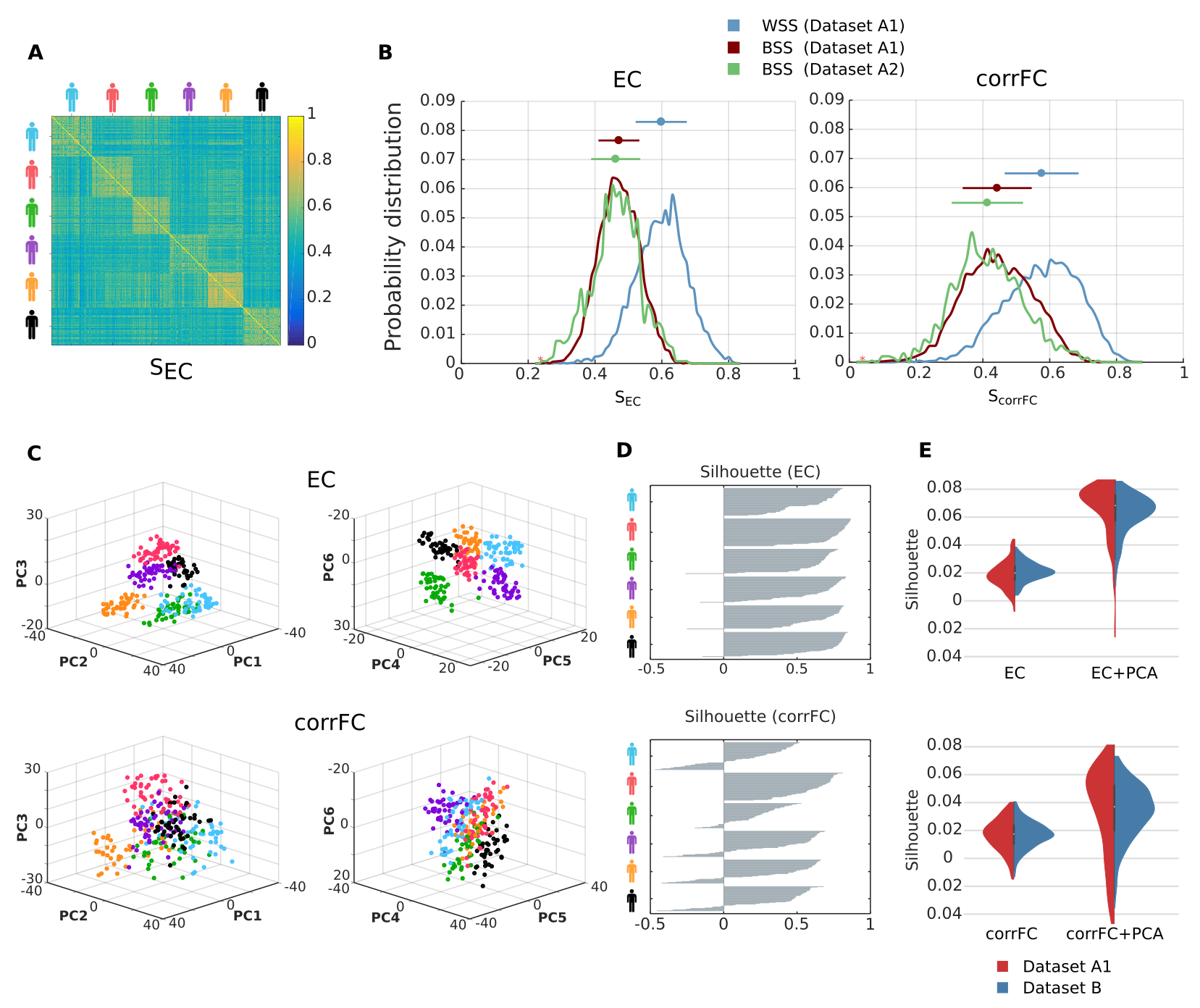
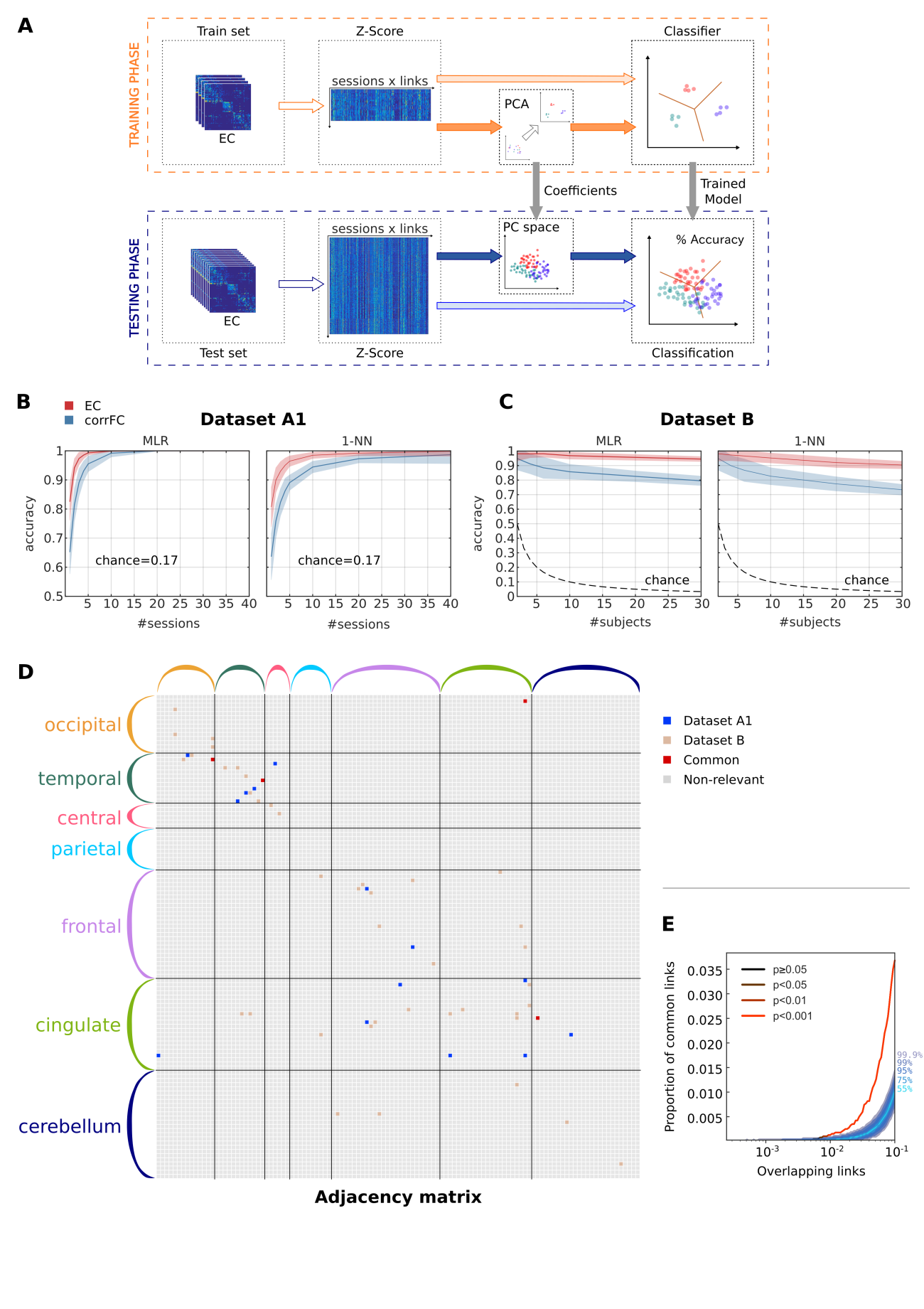
****

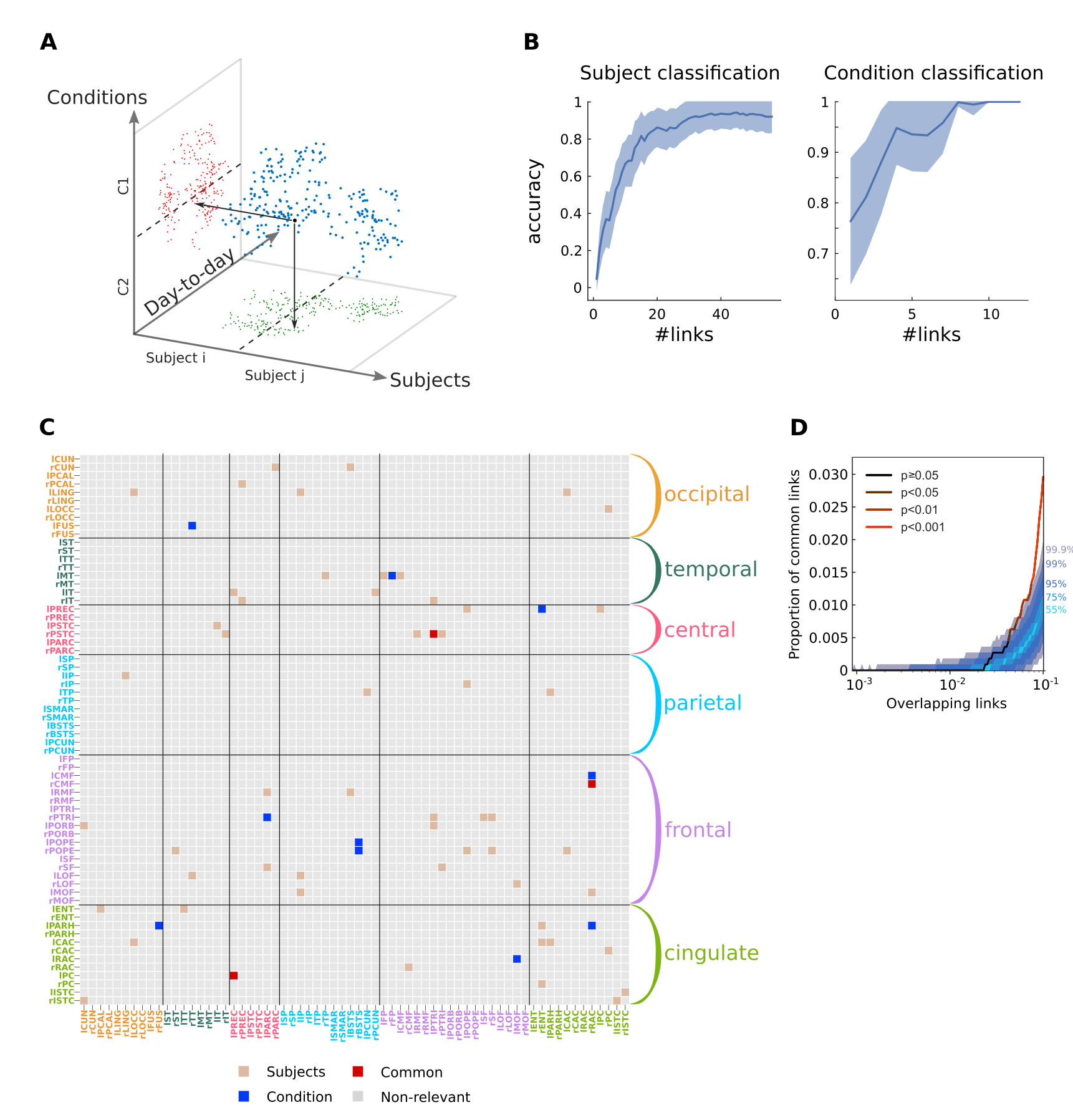
**Figure 1: Workflow for the calculation of the connectivity measures from fMRI measurements. A)** After a standard pre-processing pipeline, a parcellation covering the whole-brain is applied to extract BOLD time series: 116 ROIs for the AAL parcellation used here and 66 ROIs for the Hagmann parcellation, with each color representing an anatomical subsystem of several ROIs. Here we consider two versions of functional connectivity: the classical corrFC corresponding to the Pearson correlation coefficient (PCC) between pairs of time series; the spatiotemporal FC embodied by the two covariance matrices FC0 and FC1 without and with time shift, respectively. Details are provided in Methods, see Eq. (1) and (2). **B)** Whole-brain network model to extract effective connectivity (EC) from fMRI measurements. The local fluctuating activity (where Σii is the variance of the input to each region i) propagates via the recurrent EC to generate the correlation patterns at the network level. Structural connectivity (SC, bottom) obtained using DTI determines the skeleton of EC. The fitting procedure iteratively tunes EC and Σ such that the model best reproduces the empirical FC0 and FC1. **C)** Each corrFC matrix is symmetric and has all diagonal elements equal to 1, so that only 6670 independent links are retained for identification/classification (lower triangle). Likewise, the EC matrix has 4056 non-zero elements that are used in the classification (density of 30%).

****

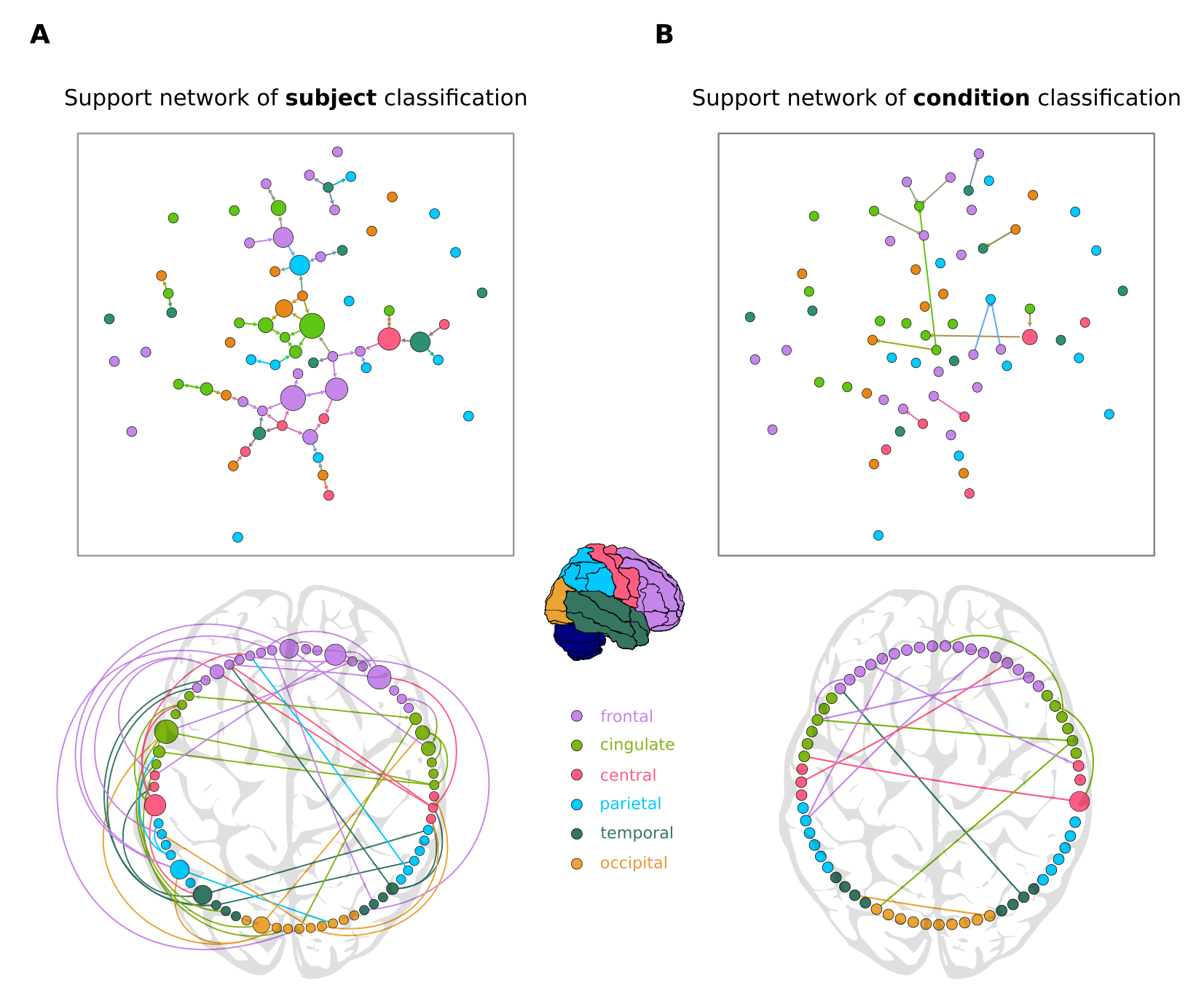
**Figure 2: Within- and between-subject similarity (WSS and BSS, respectively) for EC and corrFC. A)** Matrix of similarity values for EC between all pairs of sessions from Dataset A1. ECs from two sessions are transformed into two vectors (as illustrated in Figure 1C), from which the PCC is calculated to obtain SEC (see Eq. 13 in Methods). The sessions are grouped by subjects, as indicated by the colored symbols. **B)** The left panel shows that distributions of WSS (blue) and BSS (red) values for Datasets A1 – corresponding to diagonal and off-diagonal blocks in panel A, respectively – and of BSS (green) for Dataset A2. The right panel shows the corresponding distributions for corrFC. The above error bars indicate the means and standard deviations, indicating a smaller overlap between WSS and BSS for EC. **C)** Visualization of the sessions of Dataset A1 in the space of the first 6 PCs (split into the left and right panels) obtained from PCA for EC (top row) and corrFC (bottom row). Each point corresponds to a session and each color to one of the 6 subjects, as in panel A. **D)** Silhouette coefficients of each session in panel C (see main text and Eq. 15 in Methods for further details). **E)** Distribution of the silhouette coefficients for EC (top panel) and corrFC (bottom panel): comparison between the original link space (left) and the PCA space (right, corresponding to panel D). Both Datasets A1 (6 subjects with 6 PCs, in red) and B (30 subjects with 30 PCs, in blue) are represented by the violin plots; see also Figure S3 about the choice for the number of PCs. Note the larger silhouette coefficients for EC than for corrFC.



**Figure 3: Subject identification using EC and FC. A)** Classification pipeline used to assess the generalization of performance. The full set of connectivity measures (here EC) over all fMRI sessions was split into two groups: a train set and a test set. We use z-scores calculated over the elements of each session matrix (see Eq. 16 in Methods). We trained the classifier – with or without previously applying PCA – and evaluated the classification accuracy on the test set. **B)** Performance of multinomial logistic regression (MLR, left panel) and 1-nearest-neighbor (1NN, right panel) classifiers when increasing the number of sessions per subject used as training set with Dataset A1. The mean (solid curve) and standard deviation (colored area) were calculated for 100 repetitions with cross-validation. **C)** Same as B when varying the number of subjects using Dataset B, using a single training session per subject (leaving 9 sessions per subject as test test). **D)** Extracted links that contribute to the classification with both datasets, obtained using recursive feature elimination (RFE). The ROIs are grouped in anatomical pools, as detailed in Supplementary Table S1. **E)** Overlap between the two signatures for Datasets A1 and B as a function of selected links. The curve represents the amount of common links in the data. Shaded areas represent different quantiles of the surrogate distribution of common links under the null-hypothesis of random rankings. The color of the curve indicates the probability of the corresponding amount of common links under the null-hypothesis (here p-value < 0.001 when considering more than 1% of the total links, namely 40 links).

****

**Figure 4: Twofold discrimination between subjects and conditions using EC.   
A)** Idealized scheme of the twofold classification where each session (blue dots) is “projected” onto two subspaces, one for subjects (green) and one for conditions (red). In each subspace, classification can be performed efficiently. Depending on the orthogonality of the subspaces, the two signatures have more or less overlap. **B)** Performance of the classification for 19 subjects and 2 conditions using Dataset C as a function of number of links. Note the distinct scales for the y-axis, because the subject identification is a harder problem. **C)** Signatures of the most discriminative EC links (estimated with RFE, see text for details) for the twofold classification in B: 54 links for subject classification in brown, 10 for condition classification in blue, 3 common links in red. The ROIs are grouped in anatomical pools, as detailed in Supplementary Table S2. **D)** Proportion of common links between the subject and condition signatures as a function of selected links (in the order of the RFE ranking). Color coding is the same as in Figure 3E: the two signatures are significantly different, i.e., with a number of common links corresponding to the null hypothesis with p-value ≥ 0.05 (cf. legend) with up to 4% of the total links.



**Figure 5: Support networks of subject and condition classification. A)** The top graph plot represents the 57 most discriminative EC links supporting the classification of subjects (same as in Figure 3C). The size of each node represents its betweenness centrality in the extracted network. The most central regions are located mainly in the frontal and cingulate cortices. The bottom circular plot shows the asymmetry and lateralization of the network, with more links located in the left hemisphere. Links that are inside the circle correspond to contralateral connections, while links outside the circle correspond to ipsilateral connections. **B)** Similar graph and circular plots as A for the 13 links supporting the classification between the two conditions (resting versus movie viewing). Fewer links are required to reach high accuracy in the condition discrimination: they form a network with many disjoint components and are mainly contralateral, in comparison to the subject classification support network.