Supplementary Figures

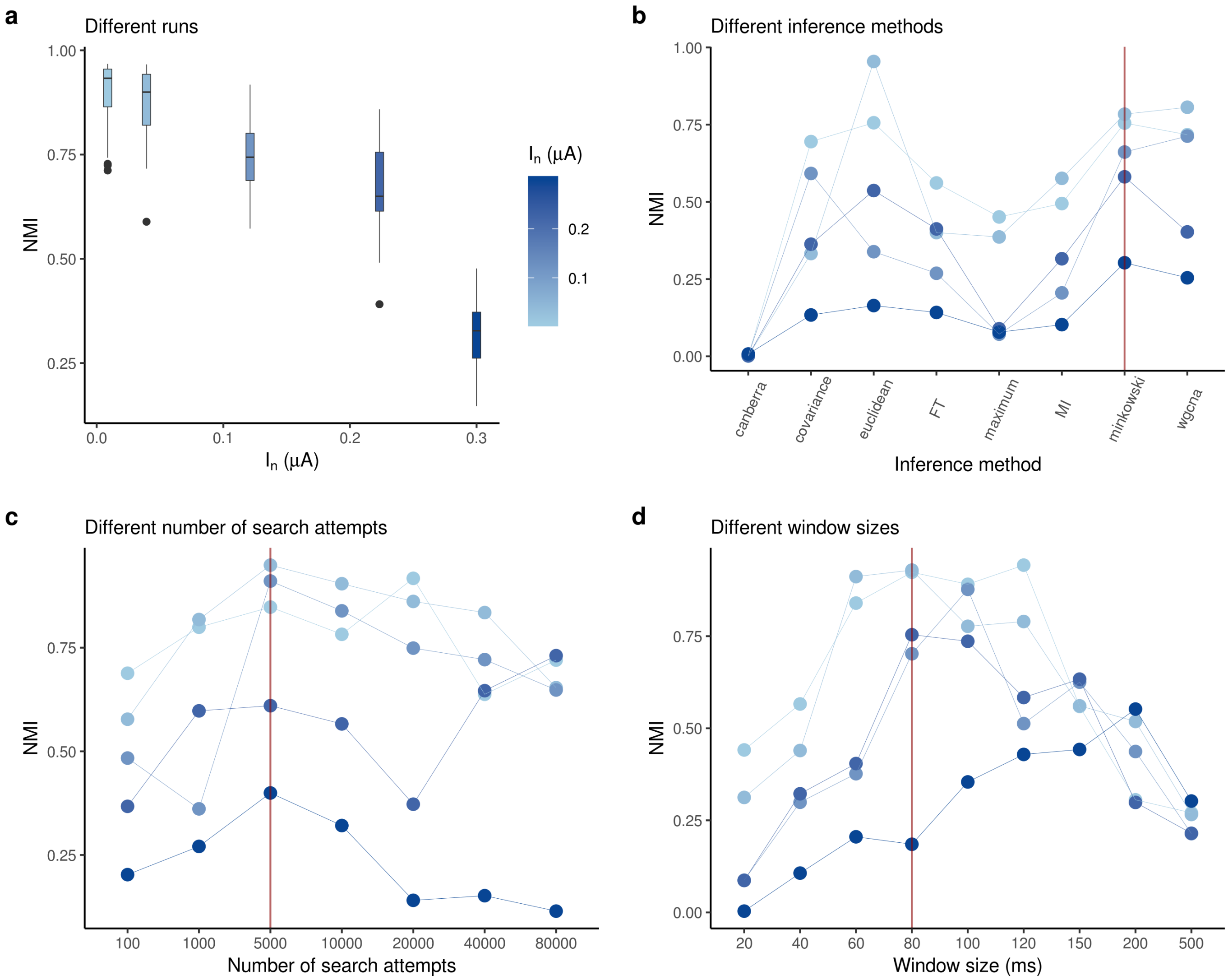

Figure S1: Robustness of NetSAP analysis. **a-d**) Evaluation of five different simulations with different amplitudes of excitatory current noise, in the range 0.01-0.3 μA (indicated in the legend from light to dark blue). These simulations are highlighted by gray circles in Fig. 8b, and have coherence levels of 0.67, 0.39, 0.17, 0.10, and 0.06, respectively. **a**) Variability estimation of the whole analysis pipeline. For each data set we run NetSAP 35 times changing only the random seed for the short spanning tree construction (5000 search attempts) and the automatic selection of barriers. The median is shown as an horizontal segment while the lower and upper hinges of the box represent the 25th and 75th percentiles. Outliers are indicated as black dots. **b-d**) Evaluation of parameters. The red vertical line indicates the values that are used as standards unless specified otherwise. **b**) Evaluation of different methods for network inference. We decided to use the Minkowski similarity because the resulting NMI scores appear to be more consistent across different noise levels, *i.e.*, they have less variability and higher scores in average. In this plot, MI stands for mutual information estimated using maximum likelihood, FT for Fourier transformation of the windowed data, while WGCNA[122] is the squared Pearson correlation. We also tried to infer the networks using the covariance matrix of each window. Other methods were classical metrics like Euclidean and Minkowski distances that we used in the form of similarities. **c**) NMI scores for different number of search attempts for the construction of the short spanning tree. **d**) Estimation of the best window size for the network inference procedure.

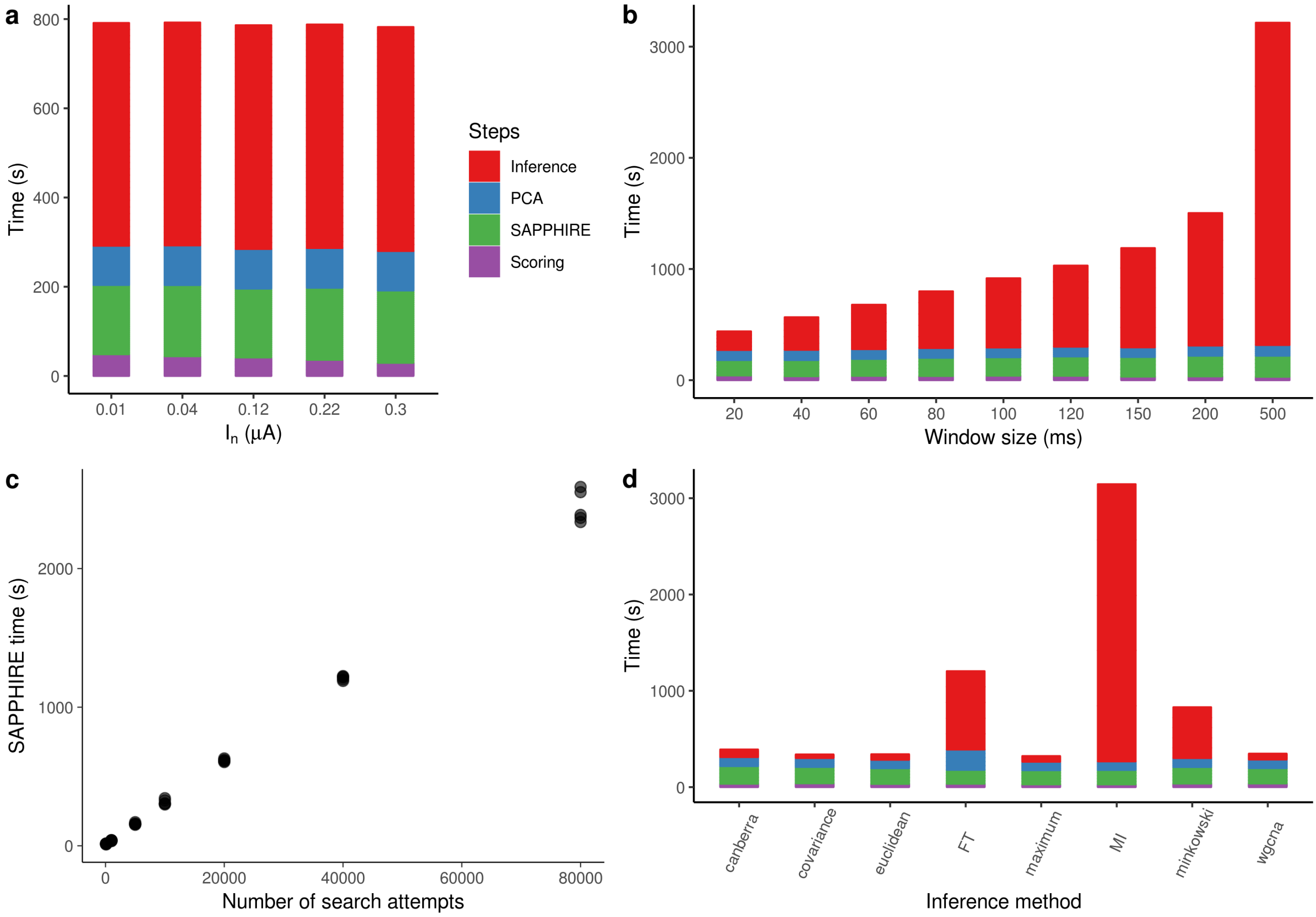

Figure S2: Computational time in seconds used for different analysis parameters. **a**) Time needed for the the four main steps of the analysis during variability estimation (indicated in the legend). Only the average results are shown while the standard deviation is omitted because it is usually less than 18 seconds. As for the other panels in this figure, the parameter choices were the standard ones from S1, *i.e.*, a Minkowski similarity metric with window size of 80 snapshots and 5000 search attempts. **b**) Evaluation of the computational time needed for different windows used to infer the networks for each time point. **c**) Time needed for different number of search attempts. **d**) Different inference methods can have significantly different computational times. In particular, maximum likelihood estimation of mutual information, MI_ML, is roughly five times slower than the Minkowski similarity.

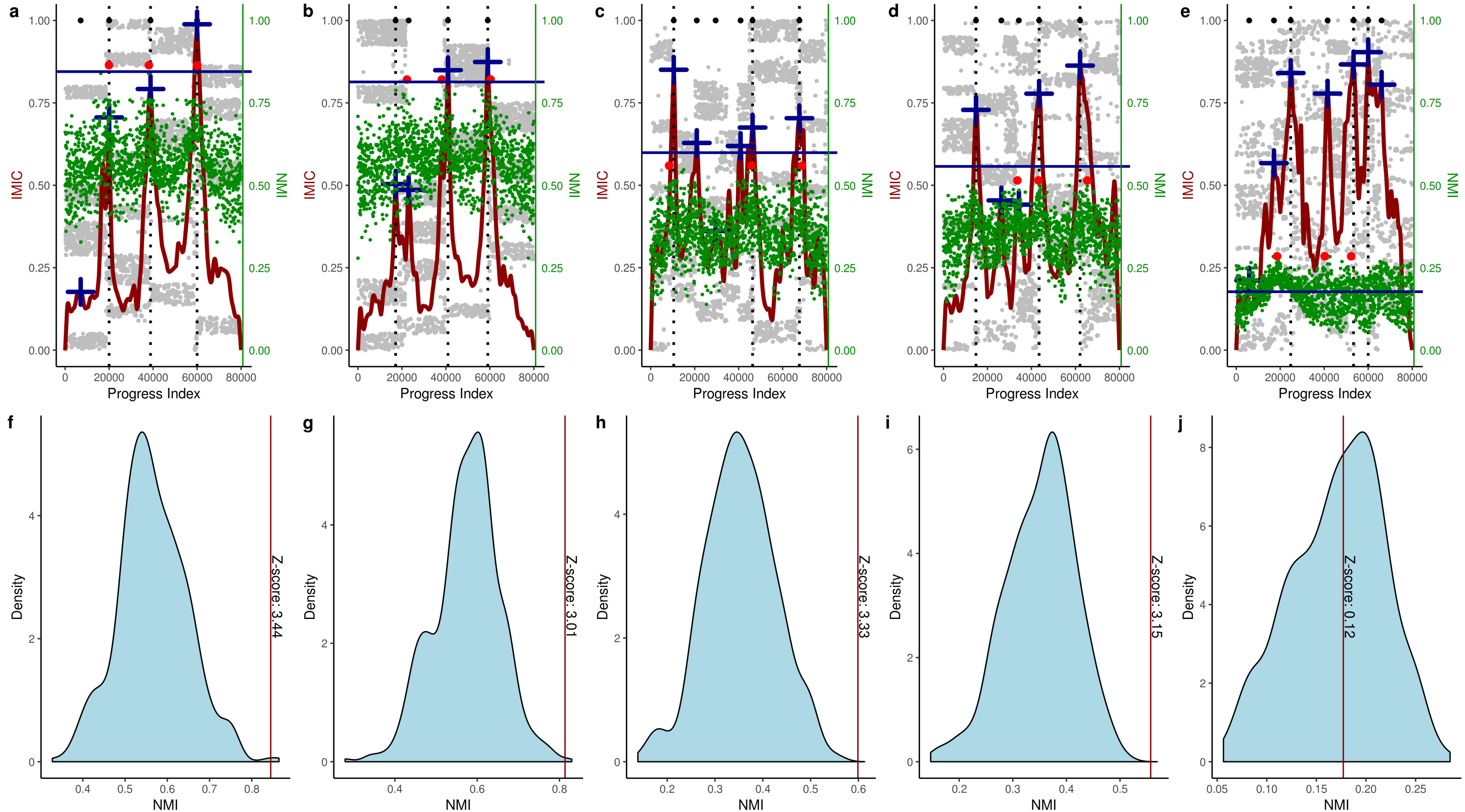

Figure S3: Robustness of barrier selection procedure for different noise levels of the simulation. **a-e**) The temporal annotation is depicted by gray points in the background. They are plotted as a function of their position in the progress index (x-axis) and the experimental times (y-axis). The dark red line shows the IMIC (inverted maximal information coefficient), which was interpolated to approximate a continuous line along the progress index. The blue crosses indicate the 'barrier candidates' and their respective height on the IMIC curve. The final selection of divisors (dotted vertical lines) is done by ranking and choosing the highest peaks found (black dots at the top). Using the first three divisors with the highest IMIC values, we computed the NMI score (shown as the horizontal dark blue line) between the inferred and true labels. As negative controls, we picked barriers randomly from a uniform distribution 500 times, and we use the same annotation to calculate the NMI (green dots, position of the barriers on the x-axis and score on the y-axis). The mutual distances between separators were restricted to be larger than a minimum number of snapshots of 2500. Red dots represent the set of those random divisors with the maximum NMI score. The five simulations in (a-e) refer to the ones selected in Fig. 8. **a**) Almost perfect simulation. Here the noise amplitude is 0.01 μA. **b-d**) Intermediate cases with about 0.15 μA of noise current. **e**) Worst case scenario with 0.3 μA. Notably, in this case the chosen barriers do not perform better than a random pick in terms of NMI. **f-j**) Density distributions of NMI values found for a-e are represented by light blue filled curves when calculated using the barriers picked randomly. The NMI score found with the IMIC optimization procedure is indicated by the dark red vertical line along with its relative Z-score. Only the simulations with noise < 0.3 μA (*i.e.*, panels (f-i)) are significant (p-value $< 5\%$).

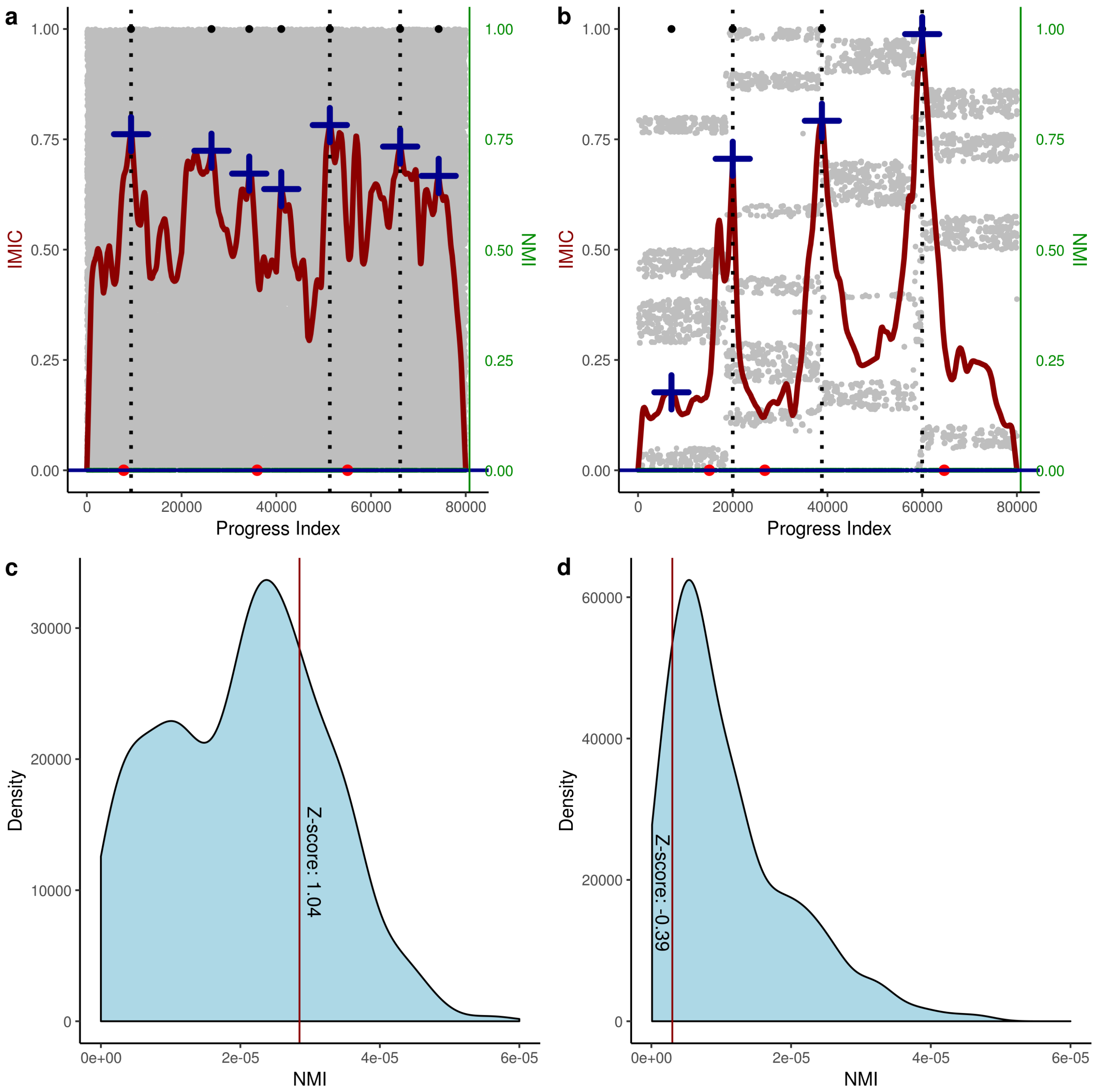

Figure S4: Randomization controls. **a**) The progress index from the simulation used in Fig. S3a is shuffled randomly on the temporal axis, producing a random distribution. The IMIC scores are all high because the sampling density is not high enough to make the time histograms sufficiently similar to each other for the chosen number of bins. The resulting NMI scores, both for the chosen divisors and the random picks are around 0 (the dark blue line and red dots, respectively). **b**) Similarly to (a), we shuffled the true annotation obtaining NMI scores around 0. **c-d**) The distribution of scores found randomly is close to 0.

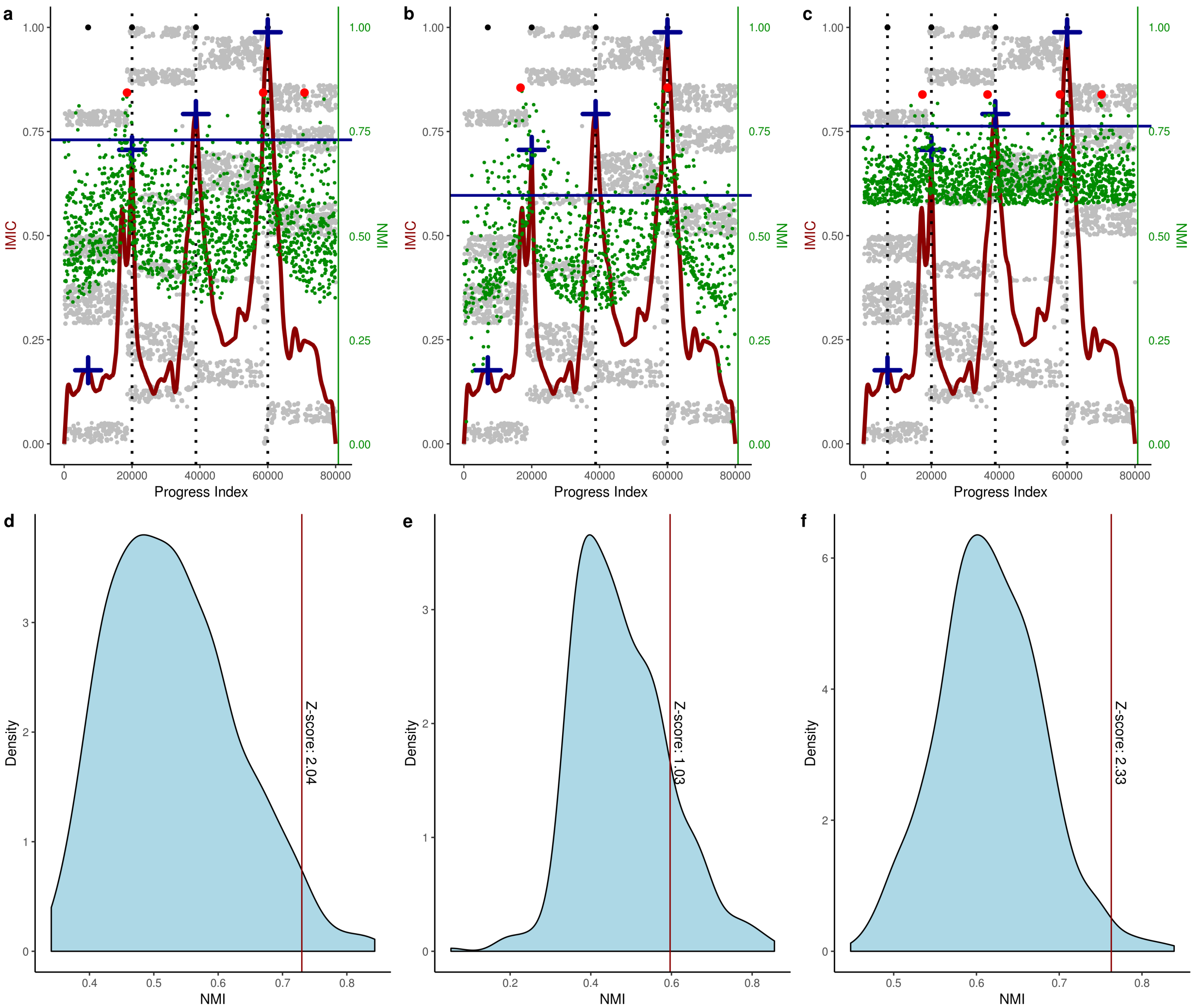

Figure S5: Robustness with respect to artificial variations in the number of clusters. **a**) Using the analysis results shown in Fig. S3a, we artificially modified the number of clusters in the annotation. Here, we kept the number of clusters searched at four but merged two clusters in the true label vector. **b**) The same annotation of (a) is used to find three clusters instead of four. **c**) In this case, we modified the true label vector by dividing the last basin of the progress index evenly into two labels (4 and 5). This artificial augmentation of the number of clusters lowers the resulting NMI for the automatic procedure (compare horizontal dark blue line at NMI = 0.77 in (c) with the one at NMI = 0.84 in Fig. S3) but not for an optimal random selection of barriers. Here, we used 1500 random picks to overcome the enhanced difficulties in finding the best random splits. The NMI values are similar to the original analysis in (a). **d-f**) Distributions of scores for the tests in panel (a-c). **f**) Here, the p-value (0.01) is significant and reflects how the barrier selection procedure selects the most important barriers first in order to have the maximum number of correct cluster assignments.