Table 6. Robustness of EMCMODULE with respect to the starting set of motifs.

Prior	% Common sites	MaxD	Sens (sd)	Spec (sd)
5	76	0.014	0.31 (0.02)	0.43 (0.03)
10	81	0.017	0.30 (0.01)	0.41 (0.02)
50	73	0.022	0.25 (0.06)	0.43 (0.10)
100	78	0.023	0.29 (0.00)	0.38 (0.02)

For the skeletal muscle regulatory sequence data (*Human and Mouse Skeletal Muscle Regulatory Sequences*), we used three different starting motif sets, based on 25, 45, and 65 matrices (total available) from the JASPAR database (all containing the "correct" motif matrices) and tested the consistency of EMCMODULE in finding the correct motif set in each case. For each starting motif set, runs were made for four prior settings corresponding to the prior mean intersite distance of $5, 10, 50,$ and 100, respectively, as in the previous example. EMCMODULE found the same overall sets of motifs, with some small variation in the sites found. "% Common sites" denotes the percentage of sites found that are common to all three data sets. The disparity in the actual motifs found is summarized through the correlation between motif column frequencies. The correlation $C_{ij}^{(k)}$ between motifs in set i and j, corresponding to the same consensus k, is considered the average correlation over all column pairs. *MaxD* denotes the maximum difference between all motifs over all three starting sets at each setting, where *MaxD* $= 1 - \min_k \min_{(i,j)\in\{(1,2),(1,3),(2,3)\}} C_{ij}^{(k)}$. As a benchmark for comparison, for two unrelated motif matrices in this data set (e.g. not sharing the same consensus), $MaxD$ is in the range $(0.48, 0.99)$ for more than 98% cases. Sensitivity and specificity are defined as in the Table 5 legend.