



**Figure S16** Uncorrected p-value distributions for the BIC-based CMST tests with data simulated from model  $G$  in Figure 5. Results based on 1,286,243 tests. For these simulations, the  $M_3$  call is the correct one, hence the skewed distribution towards small p-values at the  $M_3$  panels. The skewness towards larger p-values for the  $M_1$ ,  $M_2$ , and  $M_4$  calls follows from the fact that whenever a p-value for one model is smaller than  $\alpha$ , then the p-values for the other three models are greater than  $1 - \alpha$ . Note the larger frequency of small  $M_3$  p-values for the non-parametric CMST test (bottom left panel - the discrete nature of the histogram is a consequence of the test statistic being discrete for the non-parametric test).