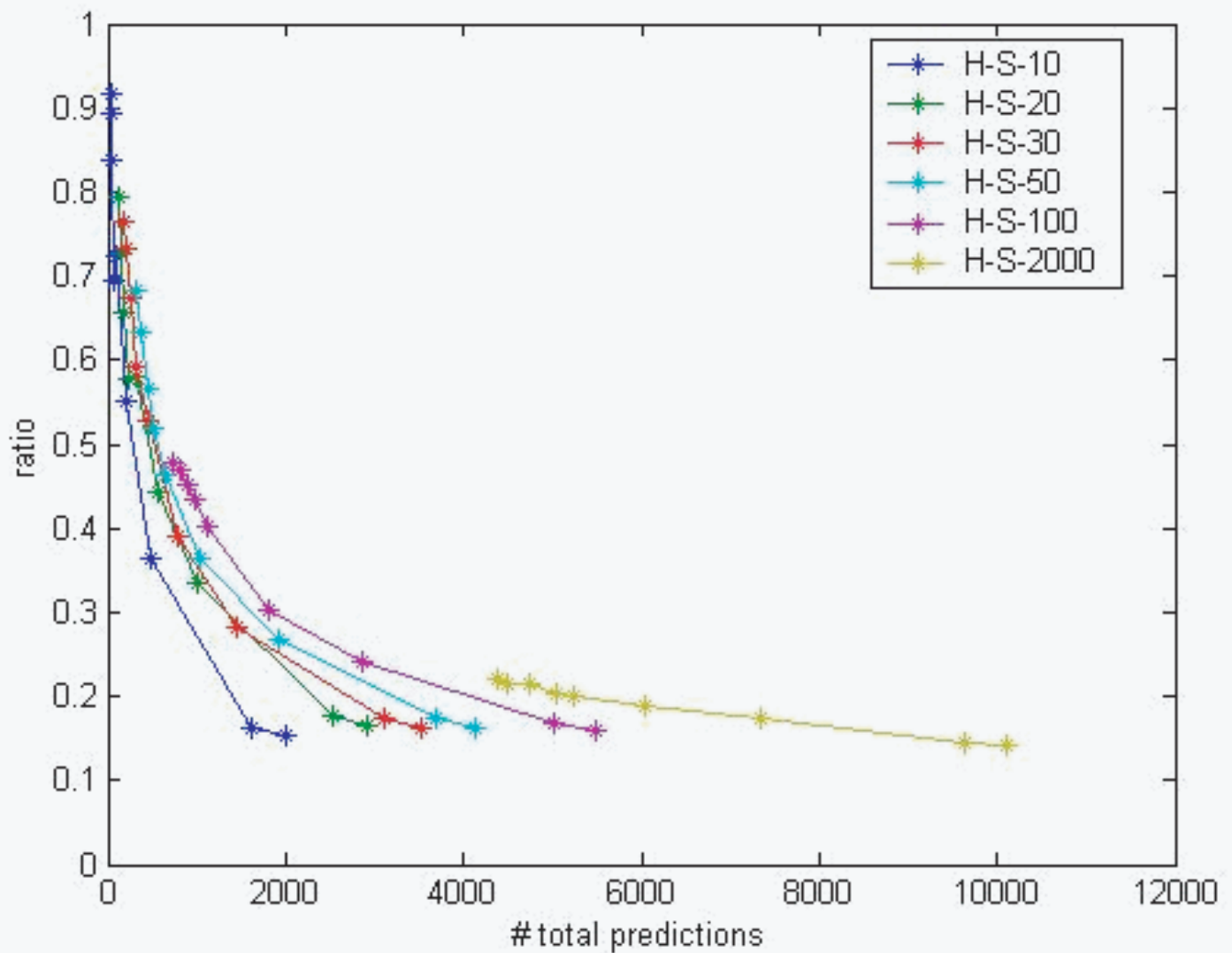




Supplemental Fig. E

Selecting cluster size cutoff for hierarchical clusters for the augmented data set. Runs of hierarchical clusters are compared by size cut (size cutoffs 10 (blue line), 20 (green line), 30 (red line), 50 (light blue line), 100 (purple line), 2000 (yellow line)). P values were modulated between 0 and 1 to form curves. Size cutoff of 100 (purple line) offered the highest coverage at a fixed percent accuracy and was used in subsequent analysis.