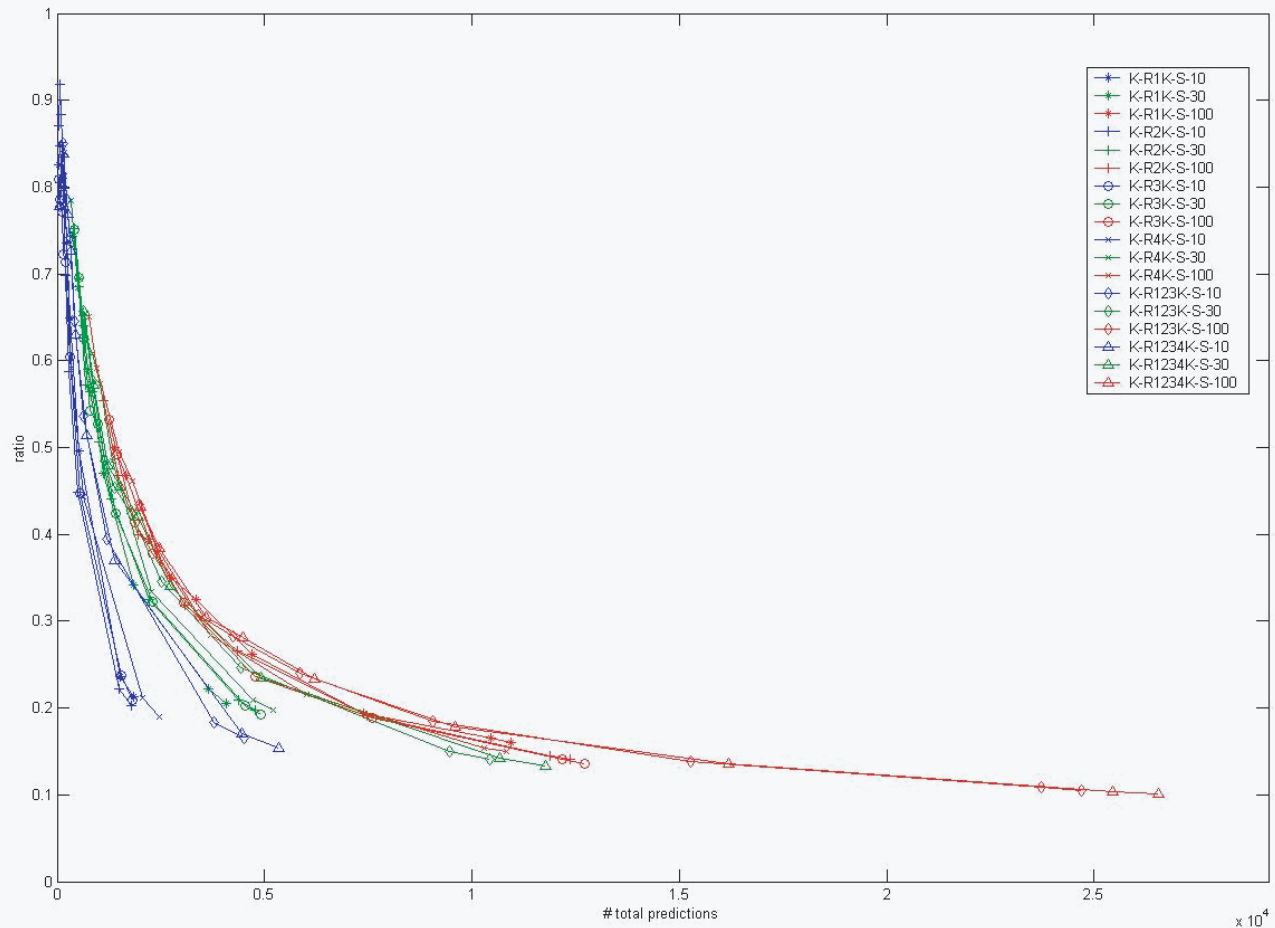




Supplemental Fig. C

Selecting cluster size cutoff for K-means clusters for the augmented data set. Performance of size cutoffs was compared over several runs of K-means (indicated as R1, R2, R3, R4) and combinations of runs (indicated as R123 and R1234). Size cutoffs are represented by color (size cutoffs 10 (blue line), 30 (green line), 100 (red line)). P values were modulated between 0 and 1 to form curves. Size cutoff of 100 (red line) offered the highest coverage at a fixed percent accuracy and was used for subsequent analysis.