

**Figure S5.** Overlap between top k-mers selected based on their SVM  $\alpha$ -values and by using the recursive feature elimination method (SVM-RFE).