Evaluation method

1) At the nucleotide level. We evaluated the prediction methods at the nucleotide level by using the benchmark dataset. That is, all evaluation indexes were calculated based on the number of overlapping nucleotides in both benchmark GIs and predicted GIs. In particular, true positives (TP) are the number of nucleotides found both in benchmark GI dataset and in predicted GIs. True negatives (TN) are the number of nucleotides found both in benchmark non-GI dataset and predicted non-GIs. False positives (FP) are the number of nucleotides found in predicted GIs but not in benchmark GI dataset. False negatives (FN) are the number of nucleotides found in benchmark GI dataset but not in predicted GIs. The recall, precision and classification accuracy are defined as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{25}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{26}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{27}$$

2) At the GI level. We also evaluated the prediction methods at GI level using the benchmark dataset, where all evaluation indexes were calculated based on the number of identified GIs, at least 50% of which is covered by the predicted GIs. TP are the number of the published GIs, at least 50% of which is covered by the predicted GIs. FN are the number of published GIs, less than 50% of which is covered by the predicted GIs. FP are the number of the predicted not included in TPs. False discovery rate (FDR) and F1 score are defined as follows:

$$\text{FDR} = \frac{\text{FP}}{\text{FP} + \text{TP}} \tag{28}$$

$$\text{F1} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{29}$$