Santhanakrishnan Ramani

SID: 105720585

**Bioinformatics & Genomics**

**Assignment – 5**

1. a) Sensitivity = True Positive / (True Positive + False Negative) = 125 / (125 + 375) = 0.25

b) Specificity = True Negative / (True Negative + False Positive) = 9475 / (9475 + 25) = 0.9974

c) Positive predictive value = True Positive / (True Positive + False Positive) = 125 / (125 + 25) = 0.8333

2. a)

b)

c)

3. a) Gro-seq and Ribo-seq give us information about genes getting transcribed and mRNAs being translated actively at

a certain instant of time respectively at their nascent state, whereas RNA-seq and proteomics gives us

information about RNA and protein respectively after they attain their steady state.

b) The FLOSS Ribo-seq scoring method algorithm is designed to distinguish true coding from non-coding sequences

based on the Ribosome protected fragment length distribution. The FLOSS algorithm provides a score based on

the comparison between the Ribosome protected fragment length distribution in a given small open reading

frame and the Ribosome protected fragment length distribution found in canonical protein-coding/non-coding

sequences. Protected fragments on nearly every individual lncRNA will have a FLOSS value very similar to that

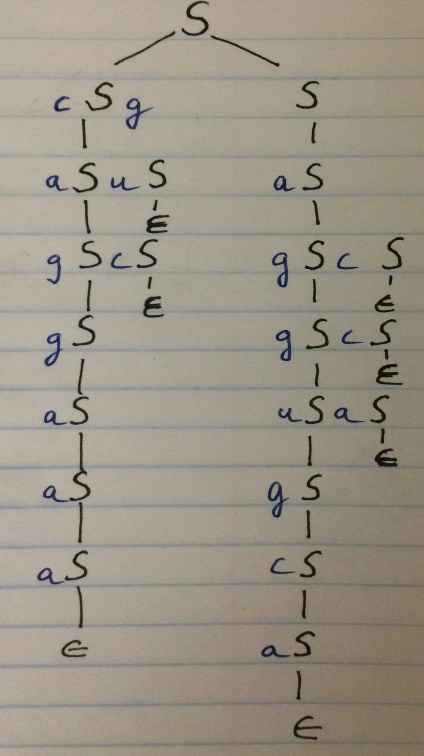
seen on coding sequences, in contrast to background from classical non-coding RNAs.

4. a)



b) 1 hit on B, 3 hits on C, 1 unknown mass. So the MADLI spectrum above represents Protein C.

c) The 1001.8 fragment doesn’t come from any of the protein and it will not have any effect on the previous answer.

5. Parse tree for the given structure

6. a) Supervised Learning:

* This method attempts to find instances of one or more pre-determined classes of elements.
* It is widely used in automatic gene finding methods and to recognize promoters, enhancers and microRNAs, based on known examples.
* Since they require a training set of known examples, they are incapable of discovering novel types of functional elements.

Unsupervised Learning:

* This methods seeks to simultaneously discover functional classes and annotate their instances de novo.
* Identifies candidate functional elements without the need for previously defined classes or known examples, thereby avoiding biases toward well-understood phenomena.

b) From fig 3, we can see that Phenotype-associated SNPs mostly reside in the Enh state.

c) Segway method is superior.

d) The number of states picked must be large enough to describe many interesting functional elements while still

being small enough for a biologist to interpret easily.