**Project 1 AI**

**What to do?**

Given:

hsa\_hg19 Rybak2015.bed – stores the genomic loci of circRNAs

hg19.fa – stores the genome (https://en.wikipedia.org/wiki/FASTA\_format)

The genes of circRNAs (loci in hsa hg19 Rybak2015.bed and genome to use in)

All exons – all exons that are not overlapped with circRNA can be considered negative examples

Find out if a certain gene (given by its start and end position in the genome) is a circRNA.

**How to solve?**

In Python

Try to adapt the predcirc.py to the project

**Get the training data**

**Positive**

Positives should be one gene that is a circRNA

hsa\_hg19 Rybak2015.bed stores the circRNA loci

**Negative**

We need to find exons that are not a circRNA that is we take the exons and remove the circRNA

* Create two files with the respective data

**Feature Extraction**

Use predcircrna lib to extract the features from any dna sequence

**Implement AI**

Training an svm

**Write report**

**Questions?**

**Sources Used**

**Report**

What to do?

“Please provide experimental results through tables or figures in your report, instead of asking TAs to run your codes.”

Assumptions:

A positive does only consist of one gene and not multiple genes.

The domain of inputs is limited to valid DNA sequences

The domain of inputs includes sequences of many sizes(, but only in one exon)