**Target**

Create a model to predict if new proteins can be involved into HBC or HCC (two types of cancers)

**How**

1. Over 2000 protein sequences involved into HBC or HCC
2. Transformation of protein sequences into numerical properties / molecular descriptors
3. Dataset {protein descriptors (input features) & protein class / output variable (HBC/HCC)}
4. Using ML from Weka find a mathematical relationship = model between input features and output class
5. The best classifier could be used to predict the class for new protein sequences