

**Universidade Federal do Rio Grande do Sul**  
**Instituto de Informática Aplicada**  
**Biologia Computacional - Lista 9**  
**Prof. Dr. Márcio Dorn**

Nome: Vicente Merlo  
Matrícula: 244950

Manti o mesmo algoritmo da lista anterior, apenas adicionei uma seleção na lista de *amostras* e um log para o arquivo *usos.txt*, para salvar qual a lista de genes que escolhi para o KMeans. O melhor resultado, que garantiu uma melhor separação, foi:

*Cluster 0:*

*ALL: 1.0 - 2.0%*

*AML: 24.0 - 96.0%*

*Cluster 1:*

*ALL: 46.0 - 98.0%*

*AML: 1.0 - 4.0%*

O conjunto de genes usado para chegar nesse resultad está no arquivo *usos.txt*.

Para selecionar, gero uma lista com o mesmo número de elementos que o número de genes em uma amostra, e preencho ela aleatoriamente com True ou False. Para cada amostra carregada do .csv, descarto o gene que corresponde ao False na lista gerada anteriormente. Assim, consigo 3572 genes aleatórios, e os passo para o KMeans, descrito no relatório anterior.