R Notebook

Code

INTRODUCTION

Scientific Question: BRCA-1 is a tumor suppressor gene found in humans and many other species like dogs and cats. When they are mutated, breast cancer is very likely to develop. How similar is the normal gene across these different species and how does expression of this gene change after it is mutated during breast cancer?

Background: BRCA1 gene is a tumor suppresor gene found in many species including humans, dogs, and cats. When there is a mutation in this gene, cancer is more prone to develop.

Scientific Hypothesis: If there was a mutation in the BRCA1 gene of these three different species, then all of them will have decreased levels of expression of this gene due to it being a tumor suppressor gene. Although the levels of expression will decrease for all three, they will vary in how much the expression decreases after mutated.

In this project, I will perform multiple sequence alignment, RNA sequencing, heatmap, and sequence logos. I will use multiple sequence alignment to compare the BRCA1 gene in humans, dogs, and cats and see how similar they are. I will download the sequences from NCBI. I will then use sequence logos to visualize the the sequence alignment across these species. I will also use RNA seq to look at the levels of expression of these genes after being mutated. I will obtain this data from the scientific articles that I found and look at how much this gene was down regulated during breast cancer. I will then use a heatmap to show the comparison between the levels of expression of this gene across all three species. This will allow me to see which species had the greater decrease in expression of this gene and which species had the least decrease.

I will list and define the packages needed to run below:

1. msa
2. seqLogo
3. DESeq2
4. GEOquery
5. canvasXress
6. ggplot2
7. clinfun
8. GGally
9. factoextra
10. pheatmp

Hide

```
library(BiocManager)
library(msa)
#if (!require("BiocManager", quietly = TRUE))
    #install.packages("BiocManager")
#BiocManager::install("seqLogo")
library(seqLogo)

#BiocManager::install(c("DESeq2", "GEOquery", "canvasXpress", "ggplot2", "clinfun", "GGa
lly", "factoextra"))
library(DESeq2)
library(GEOquery)
library(canvasXpress)
library(ggplot2)
library(clinfun)
library(GGally)
library(factoextra)

library(pheatmap)
```

PERFORMING BIONFORMATICS ANALYSES

Below, this code is responsible for loading in the 3 different sequences of the BRCA1 gene of the human, dog, and cat and comparing them using multiple sequence alignment. This will scan the sequences and find spot where there are differences.

Hide

```
library(BiocManager)
library(msa)
#if (!require("BiocManager", quietly = TRUE))
    #install.packages("BiocManager")
#BiocManager::install("seqLogo")
library(seqLogo)

#BiocManager::install(c("DESeq2", "GEOquery", "canvasXpress", "ggplot2", "clinfun", "GGa
lly", "factoextra"))
library(DESeq2)
library(GEOquery)
library(canvasXpress)
library(ggplot2)
library(clinfun)
library(GGally)
library(factoextra)

library(pheatmap)
```

Below, this code is responsible for sequence. This allows me to visualize the differences in the sequences of the three species. It will give me a graphical representation of the sequence conservation of nucleotides and will show me the diversity that exists.

Hide

```
#Sequence Logo code below
```

Below, this code is responsible for finding the difference in expression of the BRCA1 gene after being mutated for all three species. I will obtain this data from the articles that I have found and plug these into the code to run the RNA sequence.

Hide

```
#Multiple Sequence Alignment code below
```

Below, this code is responsible for showing a visual representation of the difference in expression of the BRCA1 gene obtained by RNA sequencing. It involves a heatmap that, depending on the color, will show whether the gene was upregulated or downregulated after mutated and will be compared between the three species.

Hide

```
#Sequence Logo code below
```

ANALYSIS OF RESULTS

#Make sure to tie back to the hypothesis

#Separate results for the first section of comparing the BRCA1 gene, and then for comparing expression levels