

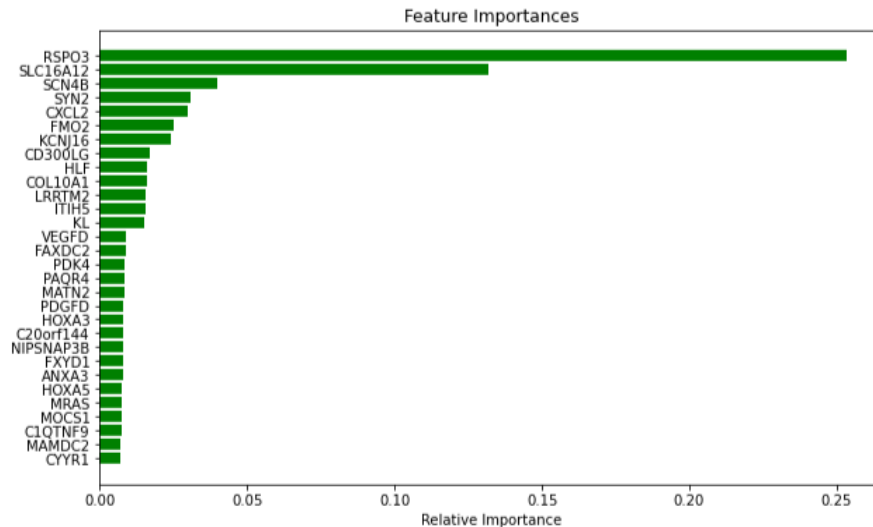
# Feature Analysis Project

Maral Karbaschi

13-August

## Data evaluation:

In this project, we found features that are associated with the "type". The dataset contains 122 samples and 19234 features and in the file of TCGA-BRCA.pheno we have 122 labels. Here, we had to subset and realign data which included in TCGA-BRCA.htseq\_fpkm-uq\_gene\_name and TCGA-BRCA.pheno. Firstly, we extracted mutual columns by columns intersection which were features transposed the feature data frame. Secondly, we relabeled data with integers which 1 represented "Tumor" and 0 denoted "Normal". Finally, it trained with Random Forest Regressor to discover relevant features by its feature\_importances\_ attribute. The top 30 features are demonstrated in the bar chart below:



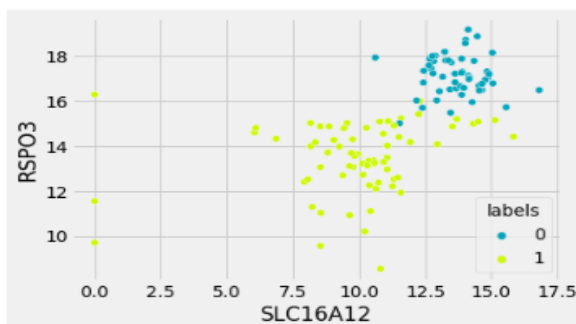
Moreover, we implemented Pearson Correlation to discover how correlated features were with the results. That is, absolute value of the correlation ratio was calculated and, the output list was sorted. The top 10 features were selected to demonstrate in the following heatmap figure.



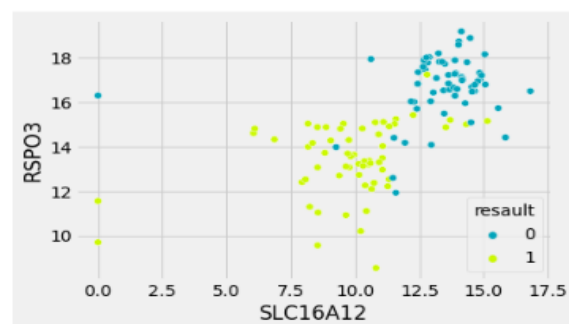
## Clustering:

Here, we used three model, KMeans, MiniBatchKMeans and, Spectral Clustering that two models, KMeans, MiniBatchKMeans utilized elbow method which is one of the most popular ways to find the optimal number of clusters.

To sum up, SLC16A12 feature was one of the mutual features among others, and it was referenced in [1] that is a highly-expressed protein in the kidney and has been reported to participate in the transport of creatine. Visualized data is demonstrated below. The right figure is the actual labeled data, and the left one is the KMeans clustering algorithm predicted result.



KMeans



Actual Data

## References:

1. Jie Mei, BS, Kehan Hu, BS, Xiafeng Peng, BS, Huiyu Wang, MD, PhD, and Chaoying Liu, MD. Monitoring Editor: Hisashi Oshiro. Decreased expression of SLC16A12 mRNA predicts poor prognosis of patients with clear cell renal cell carcinoma. 2019 Jul; 98(30): e16624.
2. Shujie Ruan, Jingping Shi, Ming Wang, and Zhechen Zhu, Analysis of Multiple Human Tumor Cases Reveals the Carcinogenic Effects of PKP3. 2021; 2021: 9391104.