PREDICTING BREAST CANCER SURVIVAL BASED ON GENE EXPRESSION AND CLINICAL VARIALBES

Jitesh Patel

Supervisor – Dr. Youcef Derbal

Ryerson University

Data Science & Analytics Graduate Studies

Objective

Explore the potential correlation between breast cancer survival and the mRNA expression of a gene set selected based on prior knowledge. The results of the analysis may be used to validate the gene set as prognostic signature of breast cancer aggressiveness.

Background

Breast cancer is one of the most common cancer types in women. In 2015, an estimated 234,190 new cases will be diagnosed, and 40,730 deaths from breast cancer will occur.

Survival of breast cancer patients is irregular. It is crucial to identify which patients are at risk of developing a more fatal type of breast cancer and have much lower survival rate. This is particularly crucial for more aggressive HER2+ and triple negative subtypes of breast cancer.

Several gene sets are directly or indirectly involved in breast cancer. I explored whether the combination of mRNA expression of such sets may improve the prediction of breast cancer survival in triple-negative class.

Any correlation between expression of selected genes and survival rate can be very helpful to define probable markers for ill defined cancer subclasses such as Triple Negative and HER2+.

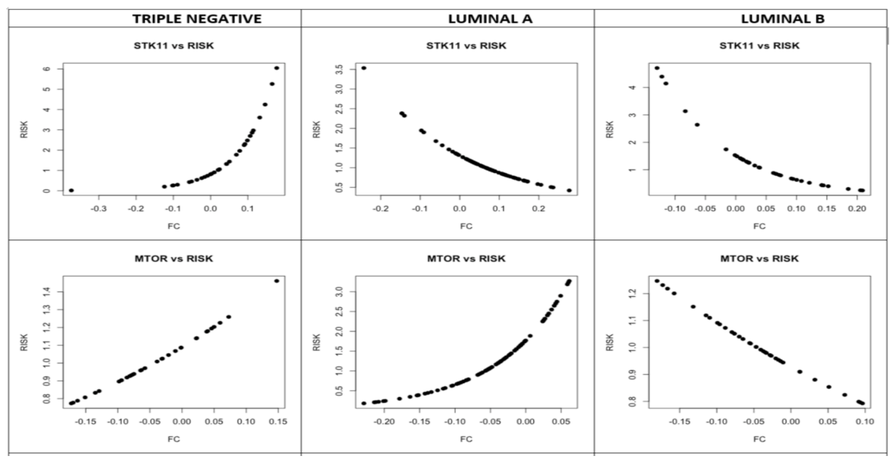
This correlation can also be helpful to speed up research on target therapies for these kind of cancer subtypes.

Methodology

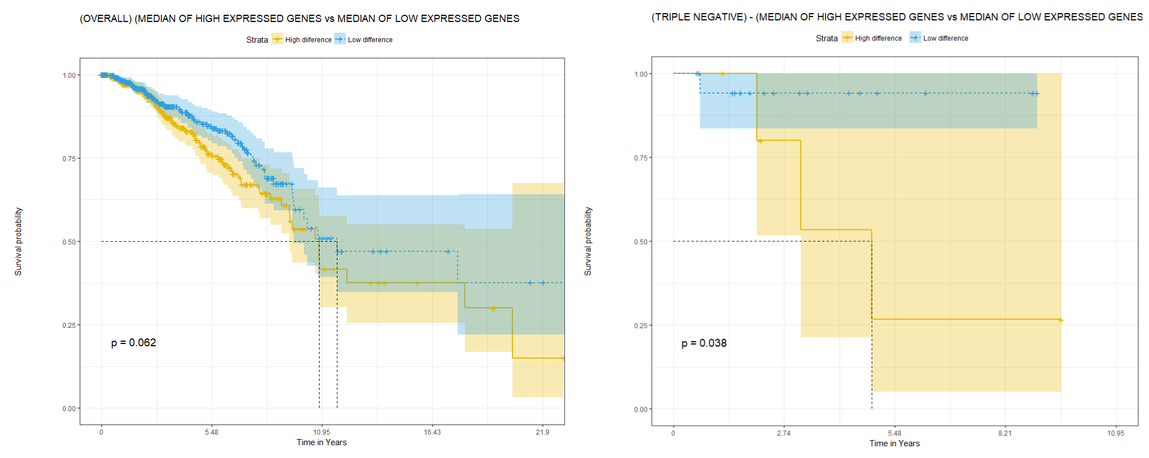
- Selected 19 genes based on prior knowledge.
- F**old Change** was calculated for respective gene expression in normal tissue and tumor tissue to measure change.
- Classify data downloaded from the TCGA into three groups (Triple Negative, Luminal A and Luminal B) based on Estrogen Receptor(ER), HER 2 Receptor (HER2) and Progesterone Receptor (PG) level for available data.
- Risk score predicted using **Cox proportional hazards regression analysis** based on fold change of each gene individually.
- Classified 19 genes into two groups (up regulated & down regulated) based on individual effect of their fold change on Risk score.
- To find out **combine effect** of above two sets of genes on survival, **MEDIAN** of up regulated genes (PRKAA1, PRKAB1, PRKAB2, MTOR, STK11, AKT2, AKT3, MYC, MYCL1, MYCN, CAMKK2, PTEN, TP53) was divided by **MEDIAN** of low regulated genes (RB1, AKT1, PIK3CA, PRKAA2, CHEK2) and classify that as high or low.
- Survival rate was predicted using **Kaplan-Meier survival analysis** based on categorical variable of **combine effect** for overall data, triple negative and luminal class of cancer.

Results

- 19 genes involved in cell cycle were selected for this study.
- 25 records were classify as Triple Negative class, 116 as Luminal class of breast cancer from 1092 records based on Estrogen, progesterone and Her2 receptor percentage level in clinical data.
- I have classified 19 genes of interest into two different sets (up-regulated and down-regulated) based on effect of **Fold Change** of gene on risk for Triple Negative subclass of breast cancer. Up-regulated gene set1 (PRKAA1, PRKAB1, PRKAB2, MTOR, STK11, AKT2, AKT3, MYC, MYCL1, MYCN, CAMKK2, PTEN, TP53) and down-regulated gene set2 (RB1, AKT1, PIK3CA, PRKAA2, CHEK2) leads to increase of risk hazard.



- Predicted survival probability using **combine effect** of set1 and set2 for Triple Negative class reflects that high difference in expression of these two sets of gene leads to low survival rate and years with p value 0.038 and survival rate for low difference in expression of these two gene sets is higher.



Conclusions

My conclusion presents the score of survival based on combine effect of up-regulated and down-regulated gene sets, which is sturdily correlated with shortened survival times in breast cancer, and the score of the model is consistently high in aggressive breast cancer types Triple Negative.

Therefore, we recommend the consideration of combine effect of these up regulated and down regulated genes as new prognostic markers and we expect that these findings can be adapted to research on target therapies for aggressive breast cancer types such as triple negative subclass of breast cancer.