

## **Supplementary Material**

### **Tables**

**Table 1. Overview of the BRCA-TCGA dataset used in this study**

	Tumour stage			Age at diagnosis		
	Early	Late	NAs	Young	Medium	Old
<b>Female</b>	169	60	4	27	131	75

**Table 1.** The table is showing the number of patients according to its tumour stage and age range. Notice that most of the patients were diagnosed with an initial tumour stage, plus they were between medium and old ages. Moreover, remark that all the patients are females. Note. Early stage of tumour is composed by the stages I and II of breast cancer and late stage is composed by the stages III and IV. The established age ranges are young, medium, and old being ages less than 45 years for young, ages between 45 years and 65 years for medium and ages above 65 years old.

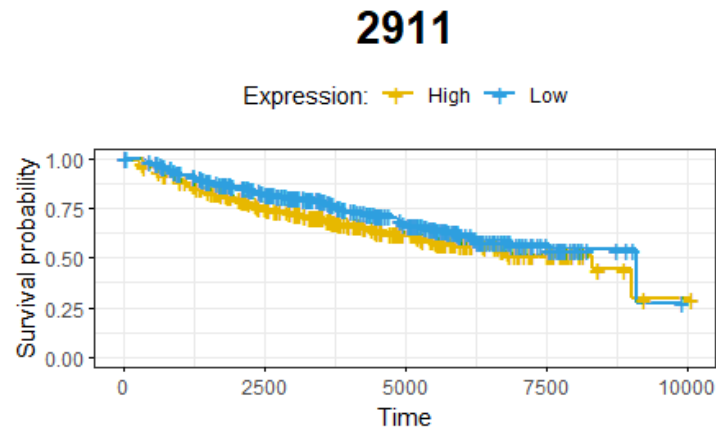
**Table 2. Overview of the METABRIC dataset used in this study**

	Tumour stage			Age at diagnosis		
	Early	Late	NAs	Young	Medium	Old
<b>Female</b>	508	28	164	319	323	58

**Table 2.** The table is showing the number of patients according to its tumour stage and age range. Notice that most of the patients were diagnosed with an early tumour stage, plus they were between young and medium ages. Moreover, remark that all the patients are females. Note. Early stage of tumour is composed by the stages I and II of breast cancer and late stage is composed by the stages III and IV. The established age ranges are young, medium, and old being ages less than 45 years for young, ages between 45 years and 65 years for medium and ages above 65 years old.

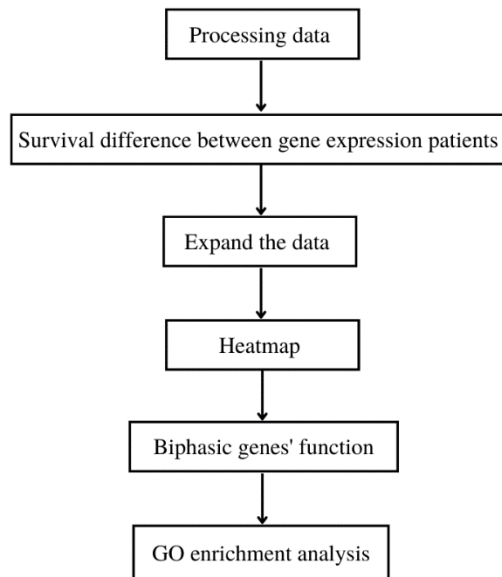
## Figures

**Figure 1. KM plot of 2911 gene**



**Figure 1.** KM plot showing a non-crossing survival curves that has been always expected to be found in the gene expression data. This KM plot is from a gene with a Entrez Gene ID 2911 (GRM1) from the METABRIC dataset.

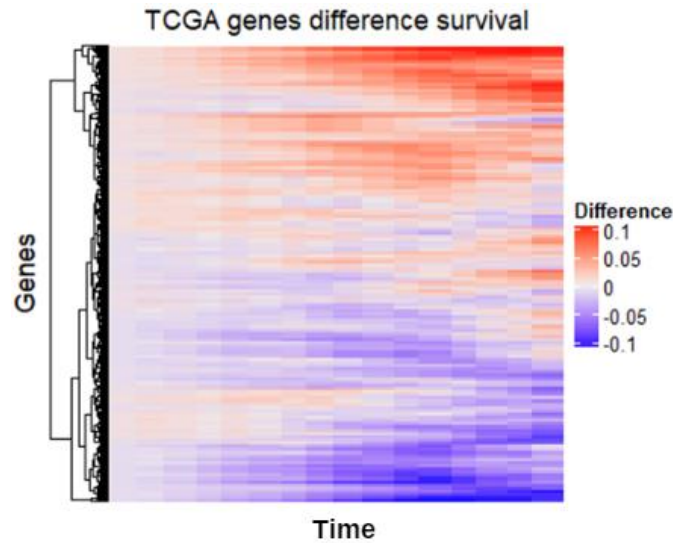
**Figure 2. Proposed bioinformatics pipeline**



**Figure 2.** Workflow of the full steps, using all functions and algorithms created, to identify genes whose expression shows a dual association with cancer risk/outcome, followed by a GO enrichment analysis.

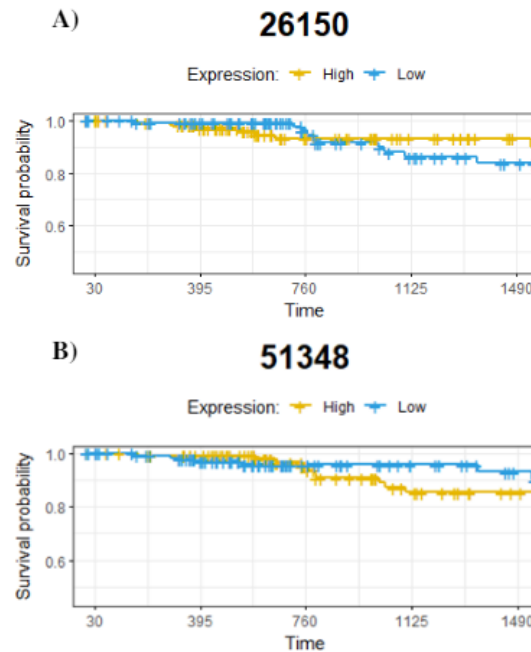
## 5.1 TCGA dataset

**Figure 3.** Heatmap of the survival differences between high and low expressions of each gene



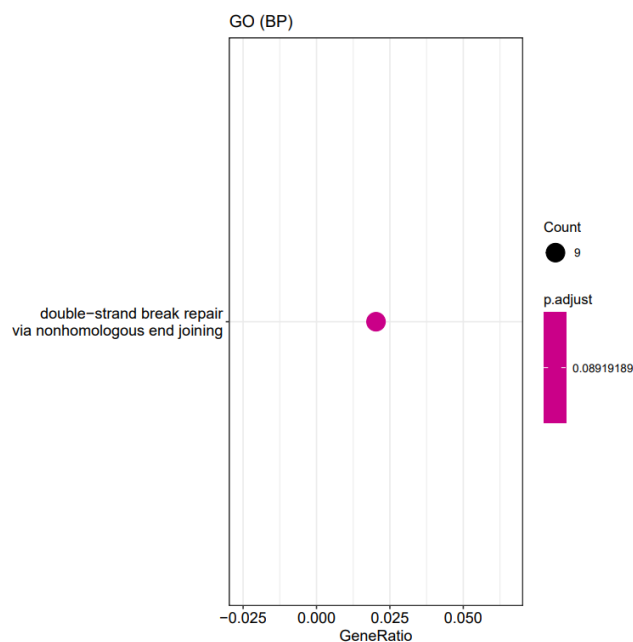
**Figure 3.** Notice that the blue colours are for the negative differences and the red colours are for the positive ones. The x-axis is the time, and the y-axis are all the genes in the dataset.

**Figure 4.** KM plots genes showing crossed survival curves of genes found with the *BIPHASIC\_FUNCTION*



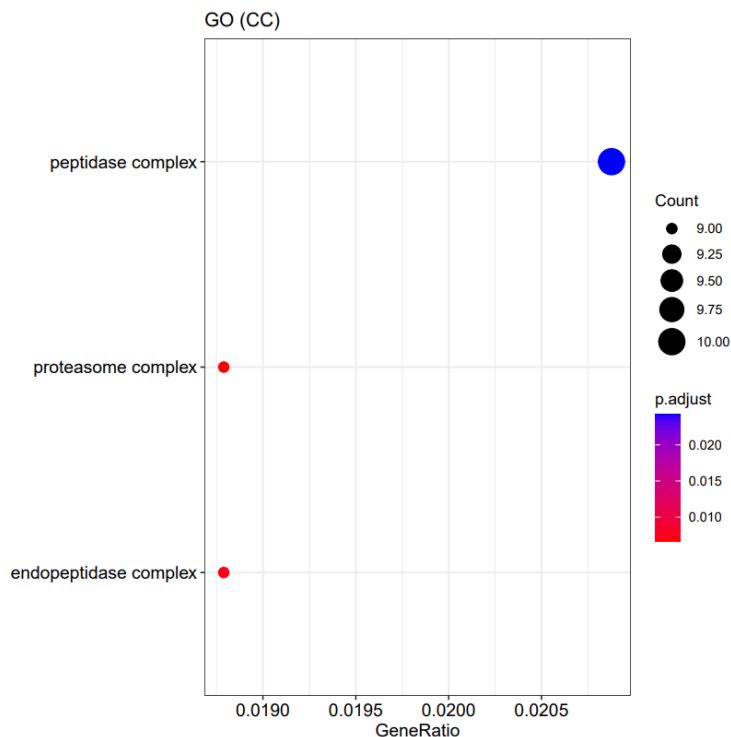
**Figure 4.** A) KM plot of gene 26150. B) KM plot of gene 51348. The blue curves are showing the survival probability of patients with low expression of the gene among time. The yellow curves are showing the survival probability of patients with high expression of the gene among time. The x-axis is the survival probability of the patients, and the y-axis is the time in days.

**Figure 5. GO enrichment dotplot of BP**



**Figure 5.** GO enrichment dotplot of BP for TCGA dataset, for the Biphasic Gens that act as a oncogene in early stages of cancer and as TDG in late stages of cancer. The plot shows an enrichment with double-strand break repair via nonhomologous end joining.

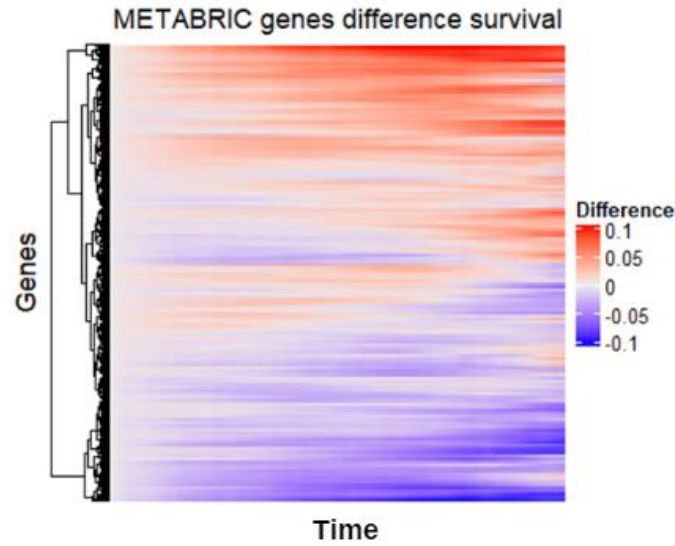
**Figure 6. GO enrichment dotplot of CC**



**Figure 6.** GO enrichment dotplot of CC for TCGA dataset, for the Biphasic Gens that act as a oncogene in early stages of cancer and as TDG in late stages of cancer. The plot shows an enrichment with peptidase, proteasome, and endopeptidase complexes.

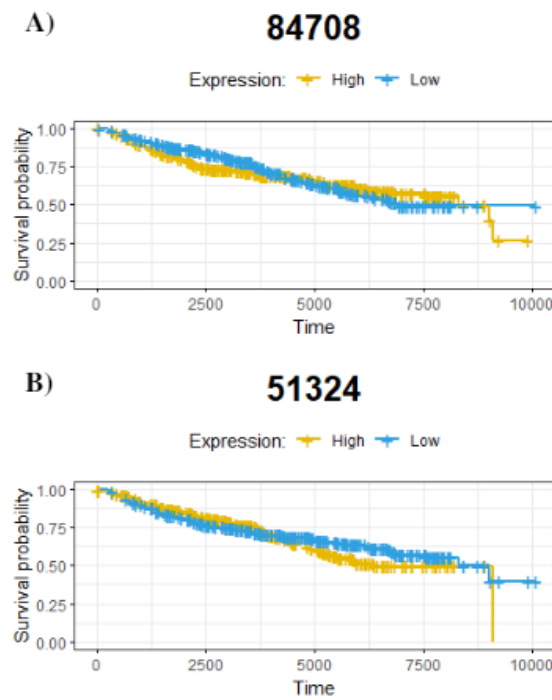
## 5.2 METABRIC dataset

**Figure 7.** Heatmap of the survival differences between high and low expressions of each gene



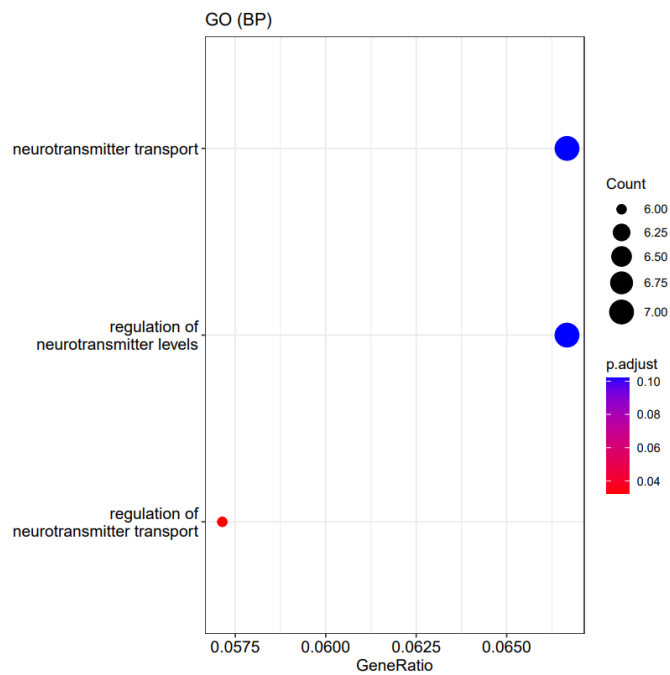
**Figure 7.** Notice that the blue colours are for the negative differences and the red colours are for the positive ones. The x-axis is the time, and the y-axis are all the genes in the dataset.

**Figure 8.** KM plots genes showing crossed survival curves of genes found with the *BIPHASIC\_FUNCTION*



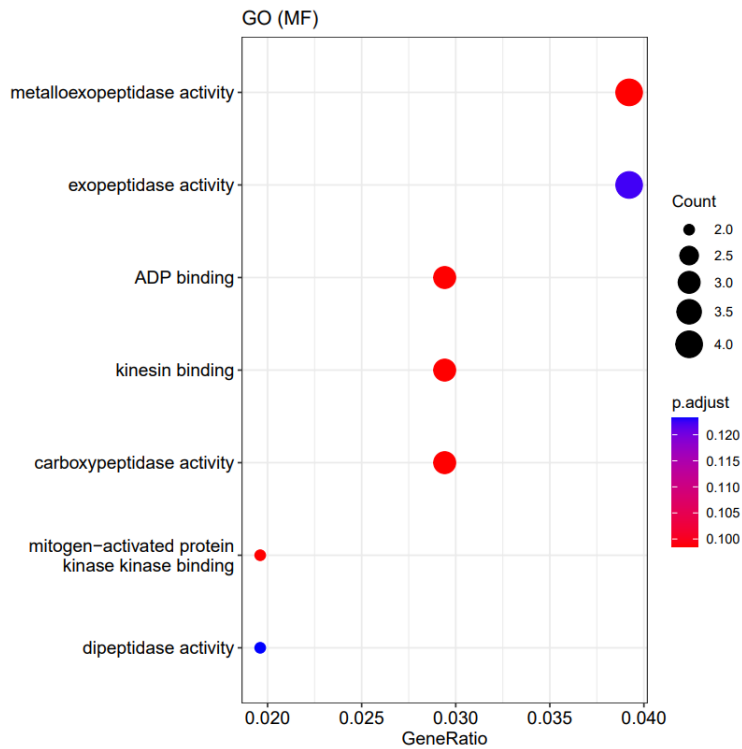
**Figure 8.** A) KM plot of gene 84708. B) KM plot of gene 51324. The blue curves are showing the survival probability of patients with low expression of the gene among time. The yellow curves are showing the survival probability of patients with high expression of the gene among time. The x-axis represents the survival probability of the patients, and the y-axis is the time in days.

**Figure 9. GO enrichment dotplot of BP**



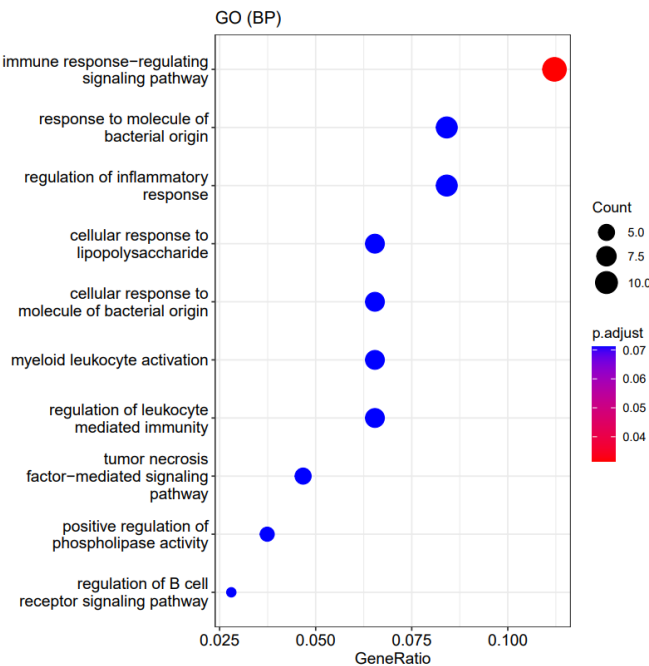
**Figure 9.** GO enrichment dotplot of BP for METABRIC dataset, for the Biphasic Gens that act as a oncogene in early stages of cancer and as TDG in late stages of cancer. The plot shows an enrichment link to neurotransmitters.

**Figure 10. GO enrichment dotplot of MF**



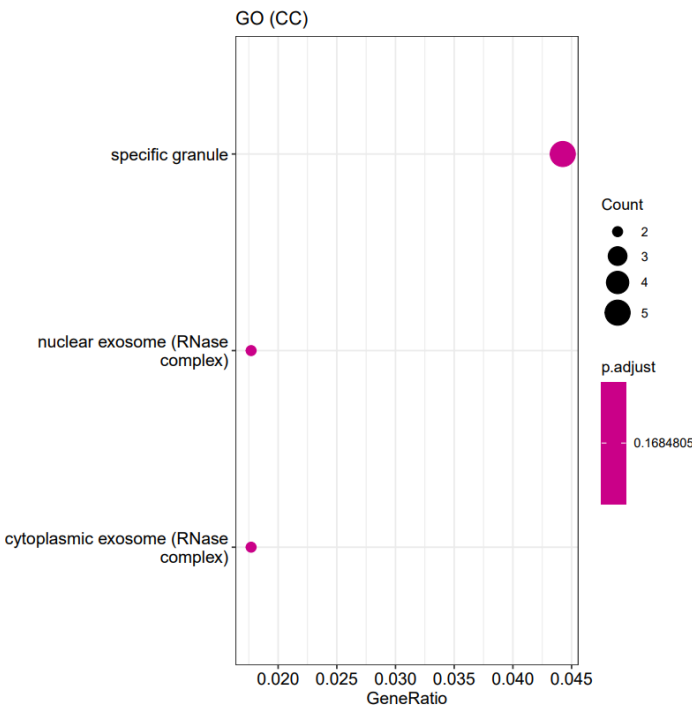
**Figure 10.** GO enrichment dotplot of MF for METABRIC dataset, for the Biphasic Gens that act as a oncogene in early stages of cancer and as TSG in late stages of cancer. The plot shows an enrichment link to peptides activities.

**Figure 11. GO enrichment dotplot of BP**



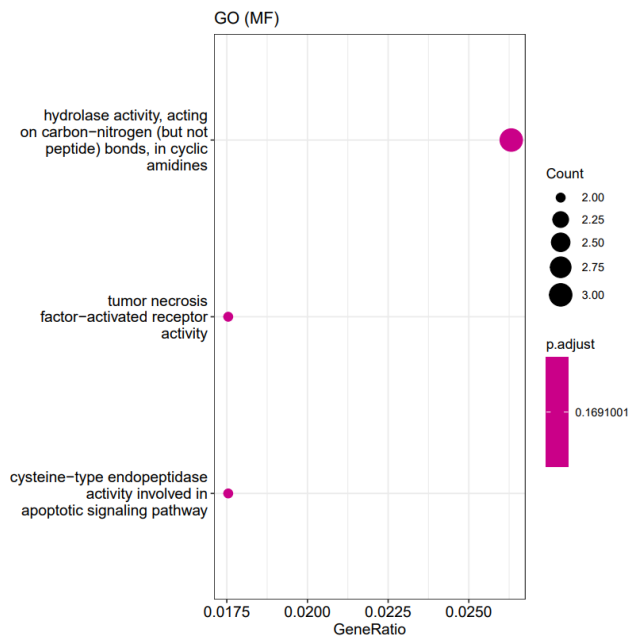
**Figure 11.** GO enrichment dotplot of BP for METABRIC dataset, for the Biphasic Gens that act as a TSG in early stages of cancer and as oncogenes in late stages of cancer. The plot shows an enrichment link to the immune system.

**Figure 12. GO enrichment dotplot of CC**



**Figure 12.** GO enrichment dotplot of CC for METABRIC dataset, for the Biphasic Gens that act as a TSG in early stages of cancer and as oncogenes in late stages of cancer. The plot shows an enrichment link to RNase complexes and specific granule.

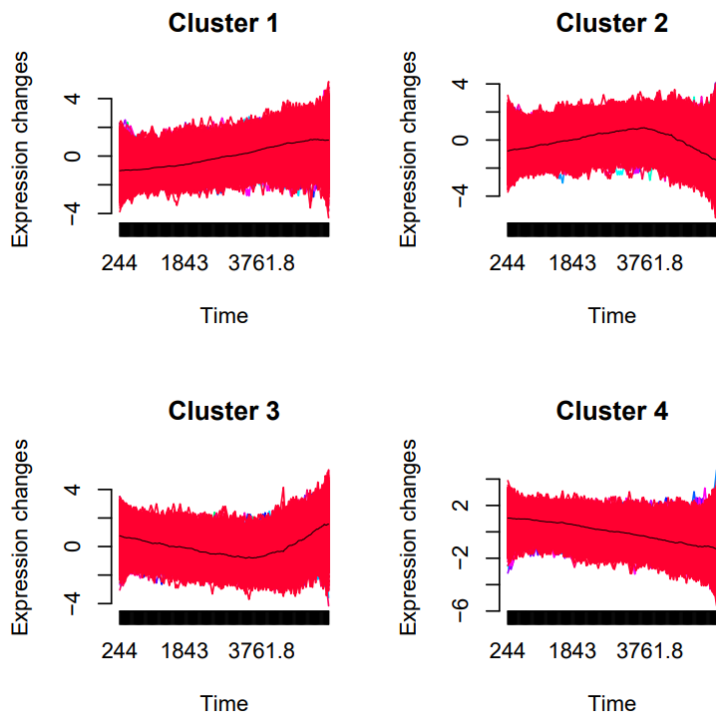
**Figure 13. GO enrichment dotplot of MF**



**Figure 13.** GO enrichment dotplot of CC for METABRIC dataset, for the Biphasic Gens that act as a TSG in early stages of cancer and as oncogenes in late stages of cancer. The plot shows an enrichment link to hydrolase activity, tumour necrosis and apoptotic signaling pathway.

### 6.1 Future directions

**Figure 14. Plot of 4 different cluster and its distributions**



**Figure 14.** Plot of 4 different cluster of the distributions of each gene that it was found in the METABRIC dataset.