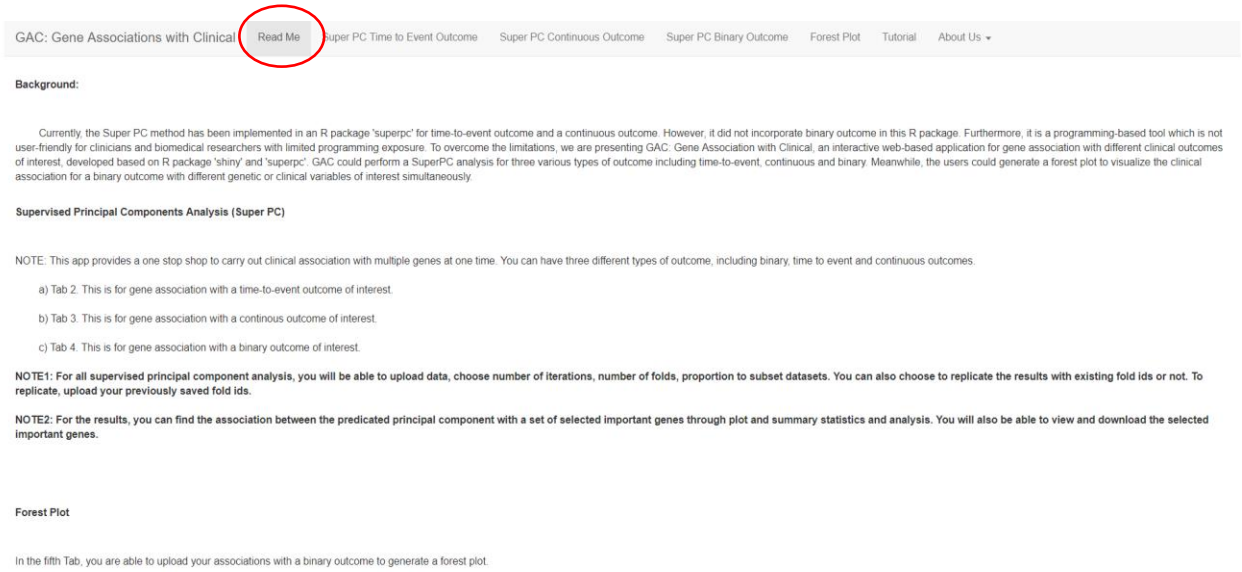
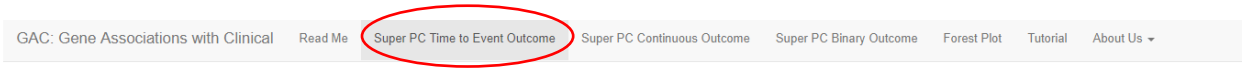
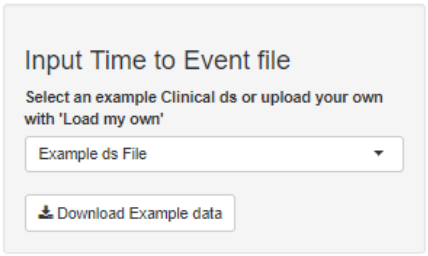
GAC Tutorial

Tab 1: Read Me



Tab 2: Super PC Time to Event Outcome

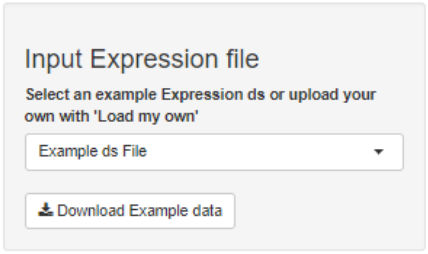




Step 1:

Select example 'Time to event' clinical dataset or upload your own. Uploaded data should be in the same format as example.

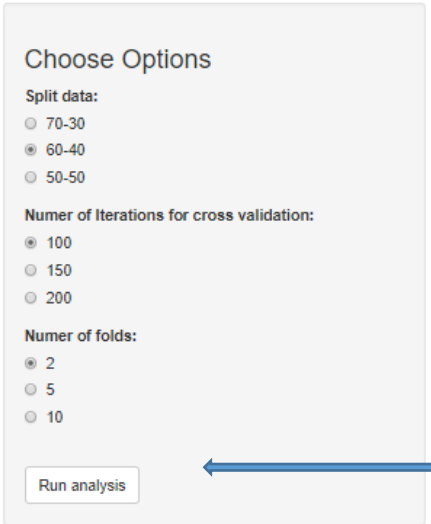
To view example data and format, use download button to view .csv file.



Step 2:

Select example 'expression' dataset or upload your own. The patients should be ordered in same order and should be exact same.

To view example data and format, use download button to view .csv file.



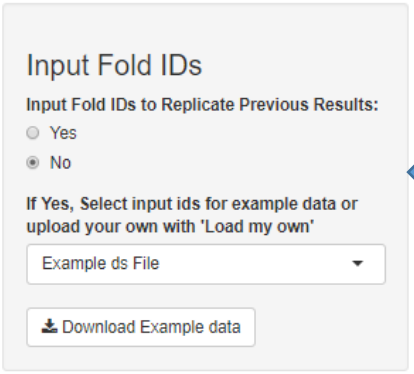
Step 3:

Choose data split into training and validation. Default uses 60-40 i.e. data is split into 40 % for training and remaining 60% for validation. You can also choose the number of interactions. For example purposes a smaller number 100 is used.

Also, select number of folds for cross validation. Default is 2.

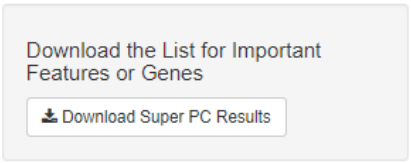
Step 4:

Hit button to run analysis after each change in option.



Step 5 (Optional):

The user can replicate previously generated results by uploading fold-ids (download available below). When running analysis for first time, leave option to 'No'.



Step 6 (Optional):

The user can download the tabular results in .csv file.

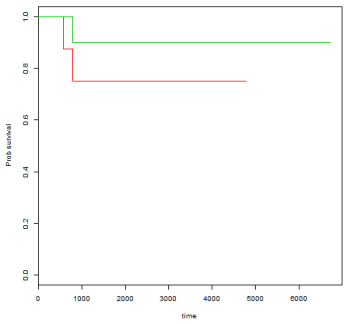
Step 7 (Optional):

The user can download the fold-ids in csv format to replicate results at a later time. These can be uploaded in the input fold id section.

Association between Continuous Predicted Principal Component with Time to Event Outcome

The p-value for the first Principal component is = 0.215364005988728.

KM Plot for Discrete Predicted Principal Component (Cut by Median)



Main panel shows the results from the super PC analysis.

The p-value from 1st supervised PC is reported.

A discrete (categorical) predictor is created by cutting the predictor at its median to form two groups for the Kaplan-Meier analysis.

Important Features Selected

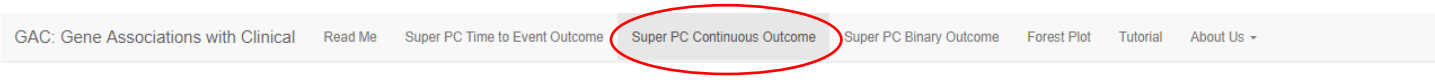
Show 10 entries Search:

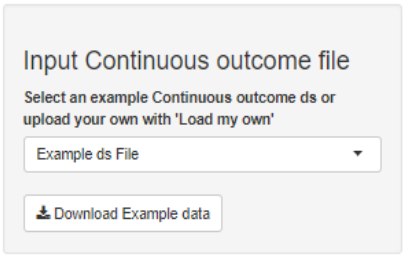
Importance-score	Raw-score	Name
184.147	1.468	MUC6.4588
142.24	1.33	WNK4.65266
127.807	1.722	ZNF385B.151126
126.463	1.638	ELOVL2.54898
116.823	1.487	FLT3.2322
112.227	1.497	SORCS1.114815
105.482	1.406	KCNH1.3756
72.727	1.341	KRT32.3882
71.761	1.26	CYP21A2.1589
66.777	1.343	DOC2B.8447

Showing 1 to 10 of 14 entries Previous 1 2 Next

List of all significant genes, in order of decreasing importance score are reported

Tab 3: Super PC Continuous Outcome

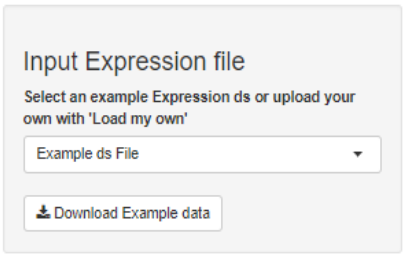




Step 1:

Select example 'continuous' outcome dataset or upload your own. Uploaded data should be in the same format as example.

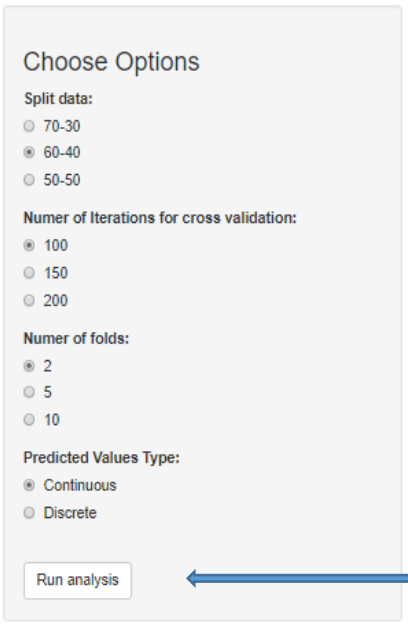
To view example data and format, use download button to view .csv file.



Step 2:

Select example 'expression' dataset or upload your own. The patients should be ordered in same order and should be exact same.

To view example data and format, use download button to view .csv file.



Step 3:

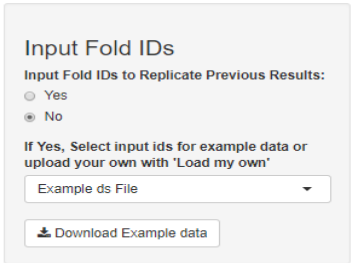
Choose data split into training and validation. Default uses 60-40 i.e. data is split into 40 % for training and remaining 60% for validation. You can also choose the number of interactions. For example purposes a smaller number 100 is used.

Also, select number of folds for cross validation. Default is 2.

Additionally, choose to display predicted values as scatter plots (for continuous predictors with Pearson correlation's) or boxplot (for binary groups cut off at median with t-test).

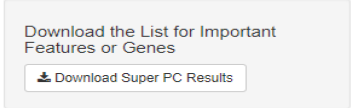
Step 4:

Hit button to run analysis after each change in option.



Step 5 (Optional):

The user can replicate previously generated results by uploading fold-ids (download available below). When running analysis for first time, leave option to 'No'.



Step 6 (Optional):

The user can download the tabular results in .csv file.

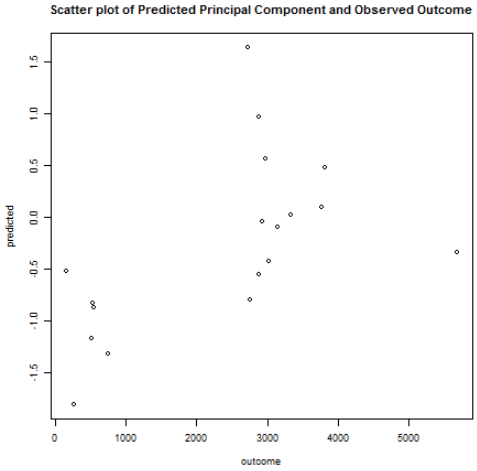
Step 7 (Optional):

The user can download the fold-ids in csv format to replicate results at a later time. These can be uploaded in the input fold id section.

Association between Continuous Predicted Principal Component with Continuous Outcome

```
The p-value for the first Principal component is = 0.009.
```

Plot and Tests for Predicted Principal Component



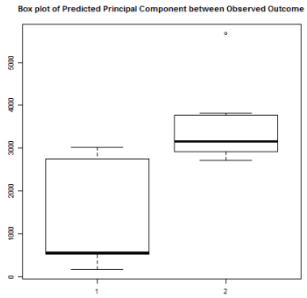
Main panel shows the results from the super PC analysis.

The p-value from 1st supervised PC is reported.

Scatter plot showing correlation of continuous predictor with outcome using Pearson correlation is reported. Discrete predictors are created by cutting the predictor at its median to form two groups and reported as boxplots.

Predicted Values Type:

- ○ Continuous
- ◉ Discrete



Test statistic	df	P value	Alternative hypothesis	cor
2.75	16	0.01422 *	two.sided	0.5666

Table: Pearson's product-moment correlation: `outcome` and `predicted`

Test statistic	df	P value	Alternative hypothesis
-4.331	14.8	0.0006114 * * *	two.sided

Table: Welch Two Sample t-test: `dat$outcome` by `dat$pred.discrete` (continued below)

mean in group 1	mean in group 2
1270	3472

Important Features Selected

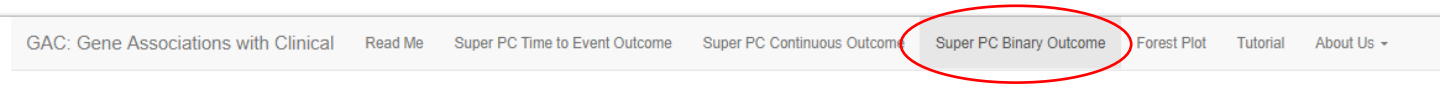
Show 10 entries Search:

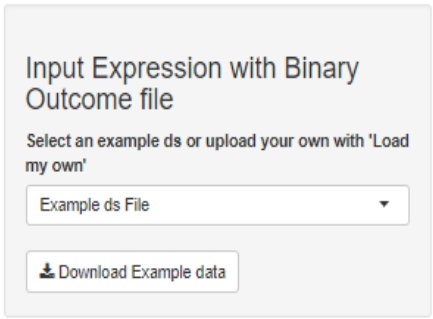
Importance-score	Raw-score	Name
97.196	0.623	CYP4F8.11283
76.61	0.591	NXPH1.30010
69.261	0.773	AFP.174
63.632	0.703	VSIG2.23584
63.19	0.67	TNFRSF13B.23495
61.456	1.131	NNAT.4826
59.825	0.579	LOC389033.389033
55.884	0.726	GPAT2.150763
55.516	0.884	TNN.63923
55.036	1.183	DNASE1L3.1776

Showing 1 to 10 of 75 entries Previous 1 2 3 4 5 … 8 Next

List of all significant genes, in order of decreasing importance score are reported

Tab 4: Super PC Binary Outcome

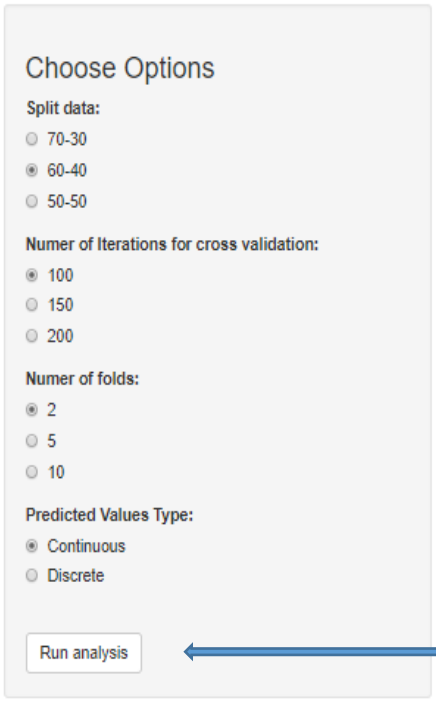




Step 1:

Select example 'expression with binary outcome indicator' dataset or upload your own. Uploaded dataset format should be same as example.

To view example data and format, use download button to view .csv file.



Step 2:

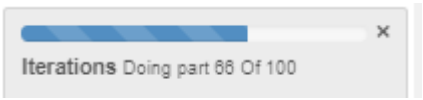
Choose data split into training and validation. Default uses 60-40 i.e. data is split into 40 % for training and remaining 60% for validation. You can also choose the number of interactions. For example purposes a smaller number 100 is used.

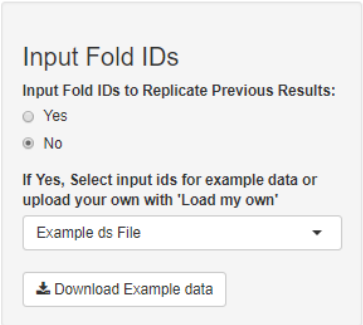
Also, select number of folds for cross validation. Default is 2.

Additionally, choose to display predicted values as boxplots (for continuous predictors with t-test) or bar plot (for discrete groups cut off at median with chisq test).

Step 3:

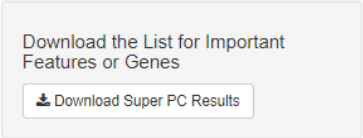
Hit button to run analysis after each change in option. A progress indicator may appear to the right corner of the page.





Step 4 (Optional):

The user can replicate previously generated results by uploading fold-ids (download available below). When running analysis for first time, leave option to 'No'.



Step 5 (Optional):

The user can download the tabular results in .csv file.

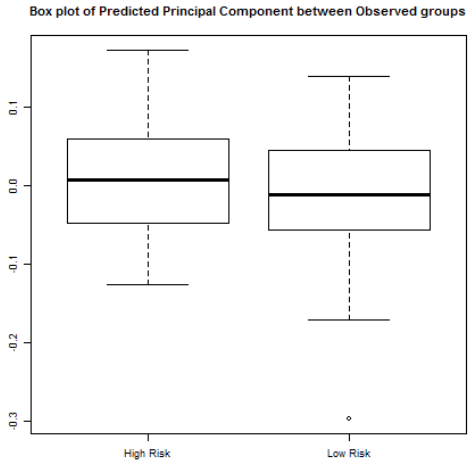
Step 6 (Optional):

The user can download the fold-ids in csv format to replicate results at a later time. These can be uploaded in the input fold id section.

Association between Predicted Principal Component with Observed Outcome

```
The p-value for the most significant Principal component is = 0.013.
```

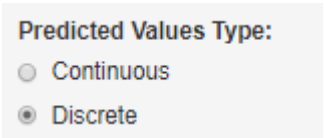
Plots and Statistical Results for Predicted Principal Component between Observed Outcome

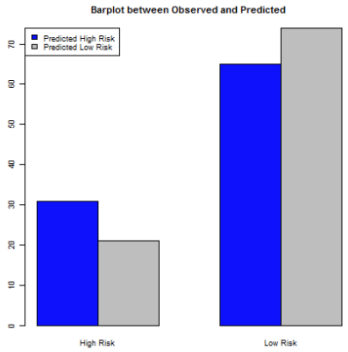


Main panel shows the results from the super PC analysis.

The p-value from 1st supervised PC is reported.

Scatter plot showing correlation of continuous predictor with outcome using Pearson correlation is reported. Discrete predictors are created by cutting the predictor at its median to form two groups and reported as boxplots.





* **High Risk**:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.1252	-0.0479	0.007103	0.009782	0.05932	0.1724

* **Low Risk**:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.2972	-0.05679	-0.01102	-0.008081	0.04541	0.1391

<!-- end of list -->

Test statistic	df	P value	Alternative hypothesis
1.608	97.66	0.1111	two.sided

Table: Welch Two Sample t-test: `dat$pred.contiuous` by `dat$group_label` (continued below)

mean in group High Risk	mean in group Low Risk
0.009782	-0.008081

	High Risk	Low Risk
Predicted High Risk	31	65
Predicted Low Risk	21	74

Test statistic	df	P value
2.013	1	0.156

Table: Pearson's Chi-squared test with Yates' continuity correction: `counts`

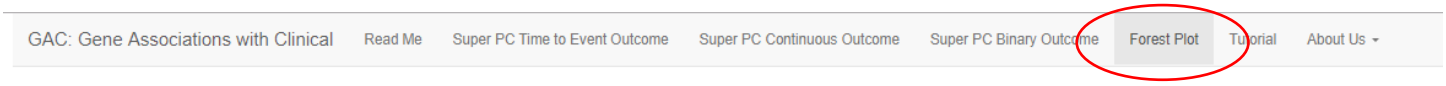
Show 10 entries Search:

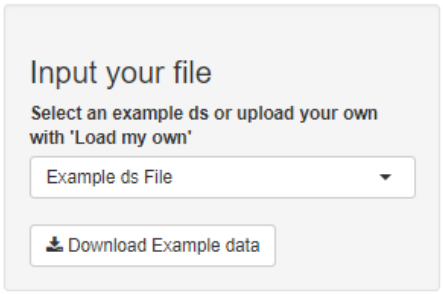
	OR	2.5 %	97.5 %	pvalue
ENSG00000003989.SLC7A2	1.1636	1.0823	1.2519	0
ENSG00000039560.RAI14	1.2379	1.1405	1.3464	0
ENSG00000071205.ARHGAP10	1.1867	1.1046	1.2776	0
ENSG00000096696.DSP	1.2506	1.1588	1.3528	0
ENSG00000099139.PCSK5	1.2188	1.1279	1.3192	0
ENSG00000101134.DOK5	1.1886	1.1021	1.2837	0
ENSG00000101493.ZNF516	1.2064	1.1191	1.3029	0
ENSG00000102271.KLHL4	1.1025	1.0378	1.1711	0.002
ENSG00000104447.TRPS1	1.1855	1.0969	1.284	0
ENSG00000105472.CLEC11A	1.2205	1.1385	1.3106	0

Showing 1 to 10 of 45 entries Previous 1 2 3 4 5 Next

List of all significant genes, in order of decreasing importance score are reported

Tab 4: Forest Plot

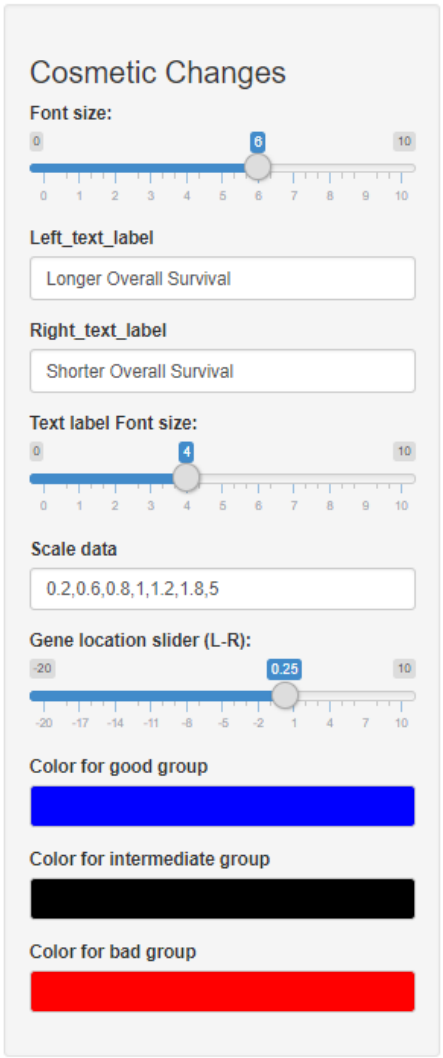




Step 1:

Select example dataset or upload your own.

To view example data and format, use download button to view .csv file.



Step 2:

Change font size for the IDs (genes) displayed to the left of forest plot.

Step 3:

Change label for text to be displayed above the horizontal line to the left and right of the HR= 1 dotted line.

A slider bar to change font size for these labels is also available.

Step 4:

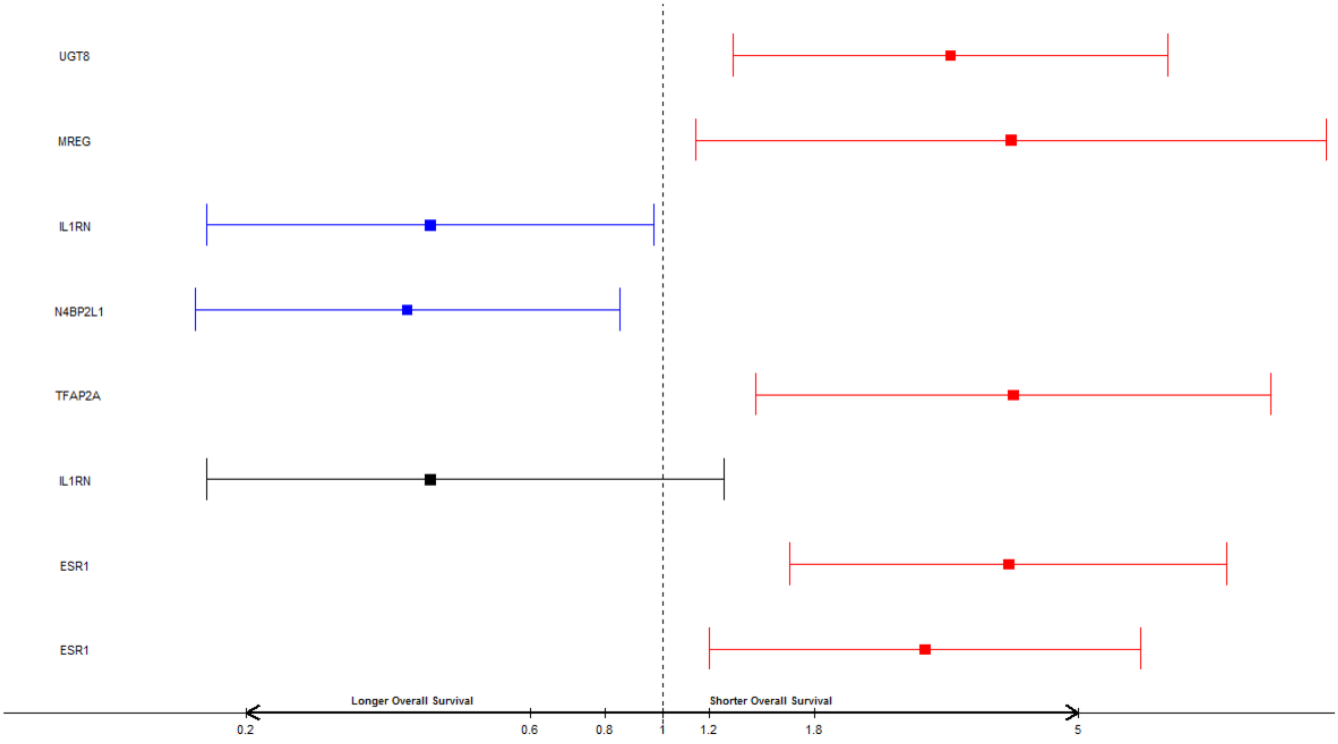
Choose x ticks to display below the horizontal line depending on dataset.

Step 7:

Move IDs (genes) slightly to right or left to avoid overlap with the forest plot.

Step 8:

Change color of outcomes. Genes with HR and their 95% CI below 1 are coded blue (good group), overlapping 1 are coded black (the intermediate group) and those over 1 are coded red (the bad group).



Reporting results from super PC analysis through forest plot. Genes with HR < 1 are coded in blue, those overlapping 1 are coded black and those greater than 1 are coded red.