

Identifying genetic markers in Skin cancer cells

TEAM 7

TRIPTI SINGH

EMRE KARAKUS

Motivation

Drug treatment researches have indicated a relationship between certain genetic markers and melanoma cancerous & non-cancerous cells. Such genetic markers can be used in improving drug therapies to treat skin cancer (melanoma).

We intend to explore such associations using data mining methods.

Hypotheses

There are some genetic markers that can assist in classifying Benign and Malignant cells

Data Source

Data is provided by Benjamin Soibam, PhD

Data

5

4645 Observation

**23689
attributes**

	A	B	C	D	E	F	G	H	I	J
1	Cell	Cy72_CD4	CY58_1_Cl	Cy71_CD4	Cy81_FNA	Cy80_II_Cl	Cy81_Bulk	Cy72_CD4	Cy74_CD4	Cy71_CD4
2	tumor	72	58	71	81	80	81	72	74	71
3	malignant(1=no,2=yes,0=unresolved)	1	1	2	2	2	2	1	1	2
4	non-malignant cell type (1=T,2=B,3=Macro.4=Endo.,5=CAF;6=NK)	2	1	0	0	0	0	1	1	0
5	C9orf152	0	0	0	0	0	0	0	0	0
6	RPS11	9.2172	8.3745	9.313	7.8876	8.3291	7.8336	8.3737	8.1338	8.4373
7	ELMO2	0	0	2.1263	0	0	0.774	0	0	0
8	CREB3L1	0	0	0	0	0	0	0	0	0
9	PNMA1	0	0	0	0	0	0	0	0	0
10	MMP2	0	0	0.73812	0	0	0	0	0	0
11	TMEM216	0	0	0	0	3.7949	0	0	0	0
12	TRAF3IP2-AS1	2.8514	2.0983	0.6173	0.96495	1.4735	3.158	1.308	1.3802	0
13	LRRC37A5P	0	0	0	0	0	0	0	0	0
14	LOC653712	0	0	0	0	0	0	0	0	0
15	C10orf90	0	0	0	3.4069	1.4468	2.7876	0	0	1.7398
16	ZHX3	0	0.52907	0	0.51197	0	1.4952	1.1097	0	1.2821
17	ERCC5	0	0	0	0	2.2866	2.3741	0	0	0
18	GPR98	0	0	0	0	0	0.0229	0	0	0

Data Preparation

removed columns with '0' values :

7

22848 attributes

**4645
Observation**

	A	B	C	D	E	F	G	H	I	J
1	Cell	Cy72_CD4	CY58_1_Cl	Cy71_CD4	Cy81_FNA	Cy80_II_Cl	Cy81_Bulk	Cy72_CD4	Cy74_CD4	Cy71_CD4
2	tumor	72	58	71	81	80	81	72	74	71
3	malignant(1=no,2=yes,0=unresolved)	1	1	2	2	2	2	1	1	2
4	non-malignant cell type (1=T,2=B,3=Macro.4=Endo.,5=CAF;6=NK)	2	1	0	0	0	0	1	1	0
5	C9orf152	0	0	0	0	0	0	0	0	0
6	RPS11	9.2172	8.3745	9.313	7.8876	8.3291	7.8336	8.3737	8.1338	8.4373
7	ELMO2	0	0	2.1263	0	0	0.774	0	0	0
8	CREB3L1	0	0	0	0	0	0	0	0	0
9	PNMA1	0	0	0	0	0	0	0	0	0
10	MMP2	0	0	0.73812	0	0	0	0	0	0
11	TMEM216	0	0	0	0	3.7949	0	0	0	0
12	TRAF3IP2-AS1	2.8514	2.0983	0.6173	0.96495	1.4735	3.158	1.308	1.3802	0
13	LRRC37A5P	0	0	0	0	0	0	0	0	0
14	LOC653712	0	0	0	0	0	0	0	0	0
15	C10orf90	0	0	0	3.4069	1.4468	2.7876	0	0	1.7398
16	ZHX3	0	0.52907	0	0.51197	0	1.4952	1.1097	0	1.2821
17	ERCC5	0	0	0	0	2.2866	2.3741	0	0	0
18	GPR98	0	0	0	0	0	0.0229	0	0	0

Curse of Dimensionality

removed columns with near zero variance

9

5628 attributes

4645
Observation

	A	B	C	D	E	F	G	H	I	J
1	Cell	Cy72_CD4	CY58_1_Cl	Cy71_CD4	Cy81_FNA	Cy80_II_Cl	Cy81_Bulk	Cy72_CD4	Cy74_CD4	Cy71_CD4
2	tumor	72	58	71	81	80	81	72	74	71
3	malignant(1=no,2=yes,0=unresolved)	1	1	2	2	2	2	1	1	2
4	non-malignant cell type (1=T,2=B,3=Macro.4=Endo.,5=CAF;6=NK)	2	1	0	0	0	0	1	1	0
5	C9orf152	0	0	0	0	0	0	0	0	0
6	RPS11	9.2172	8.3745	9.313	7.8876	8.3291	7.8336	8.3737	8.1338	8.4373
7	ELMO2	0	0	2.1263	0	0	0.774	0	0	0
8	CREB3L1	0	0	0	0	0	0	0	0	0
9	PNMA1	0	0	0	0	0	0	0	0	0
10	MMP2	0	0	0.73812	0	0	0	0	0	0
11	TMEM216	0	0	0	0	3.7949	0	0	0	0
12	TRAF3IP2-AS1	2.8514	2.0983	0.6173	0.96495	1.4735	3.158	1.308	1.3802	0
13	LRRC37A5P	0	0	0	0	0	0	0	0	0
14	LOC653712	0	0	0	0	0	0	0	0	0
15	C10orf90	0	0	0	3.4069	1.4468	2.7876	0	0	1.7398
16	ZHX3	0	0.52907	0	0.51197	0	1.4952	1.1097	0	1.2821
17	ERCC5	0	0	0	0	2.2866	2.3741	0	0	0
18	GPR98	0	0	0	0	0	0.0229	0	0	0

removed correlated variables

5564 attributes

4645
Observation

	A	B	C	D	E	F	G	H	I	J
1	Cell	Cy72_CD4	CY58_1_Cl	Cy71_CD4	Cy81_FNA	Cy80_II_Cl	Cy81_Bulk	Cy72_CD4	Cy74_CD4	Cy71_CD4
2	tumor	72	58	71	81	80	81	72	74	71
3	malignant(1=no,2=yes,0=unresolved)	1	1	2	2	2	2	1	1	2
4	non-malignant cell type (1=T,2=B,3=Macro.4=Endo.,5=CAF;6=NK)	2	1	0	0	0	0	1	1	0
5	C9orf152	0	0	0	0	0	0	0	0	0
6	RPS11	9.2172	8.3745	9.313	7.8876	8.3291	7.8336	8.3737	8.1338	8.4373
7	ELMO2	0	0	2.1263	0	0	0.774	0	0	0
8	CREB3L1	0	0	0	0	0	0	0	0	0
9	PNMA1	0	0	0	0	0	0	0	0	0
10	MMP2	0	0	0.73812	0	0	0	0	0	0
11	TMEM216	0	0	0	0	3.7949	0	0	0	0
12	TRAF3IP2-AS1	2.8514	2.0983	0.6173	0.96495	1.4735	3.158	1.308	1.3802	0
13	LRRC37A5P	0	0	0	0	0	0	0	0	0
14	LOC653712	0	0	0	0	0	0	0	0	0
15	C10orf90	0	0	0	3.4069	1.4468	2.7876	0	0	1.7398
16	ZHX3	0	0.52907	0	0.51197	0	1.4952	1.1097	0	1.2821
17	ERCC5	0	0	0	0	2.2866	2.3741	0	0	0
18	GPR98	0	0	0	0	0	0.0229	0	0	0

Keeping only Benign & Malignant observations

11

5564 attributes

4613
Observation

	A	B	C	D	E	F	G	H	I	J
1	Cell	Cy72_CD4	CY58_1_Cl	Cy71_CD4	Cy81_FNA	Cy80_II_Cl	Cy81_Bulk	Cy72_CD4	Cy74_CD4	Cy71_CD4
2	tumor	72	58	71	81	80	81	72	74	71
3	malignant(1=no,2=yes,0=unresolved)	1	1	2	2	2	2	1	1	2
4	non-malignant cell type (1=T,2=B,3=Macro,4=Endo.,5=CAF;6=NK)	2	1	0	0	0	0	1	1	0
5	C9orf152	0	0	0	0	0	0	0	0	0
6	RPS11	9.2172	8.3745	9.313	7.8876	8.3291	7.8336	8.3737	8.1338	8.4373
7	ELMO2	0	0	2.1263	0	0	0.774	0	0	0
8	CREB3L1	0	0	0	0	0	0	0	0	0
9	PNMA1	0	0	0	0	0	0	0	0	0
10	MMP2	0	0	0.73812	0	0	0	0	0	0
11	TMEM216	0	0	0	0	3.7949	0	0	0	0
12	TRAF3IP2-AS1	2.8514	2.0983	0.6173	0.96495	1.4735	3.158	1.308	1.3802	0
13	LRRC37A5P	0	0	0	0	0	0	0	0	0
14	LOC653712	0	0	0	0	0	0	0	0	0
15	C10orf90	0	0	0	3.4069	1.4468	2.7876	0	0	1.7398
16	ZHX3	0	0.52907	0	0.51197	0	1.4952	1.1097	0	1.2821
17	ERCC5	0	0	0	0	2.2866	2.3741	0	0	0
18	GPR98	0	0	0	0	0	0.0229	0	0	0

Reduced Data used for all further data mining

5564 attributes

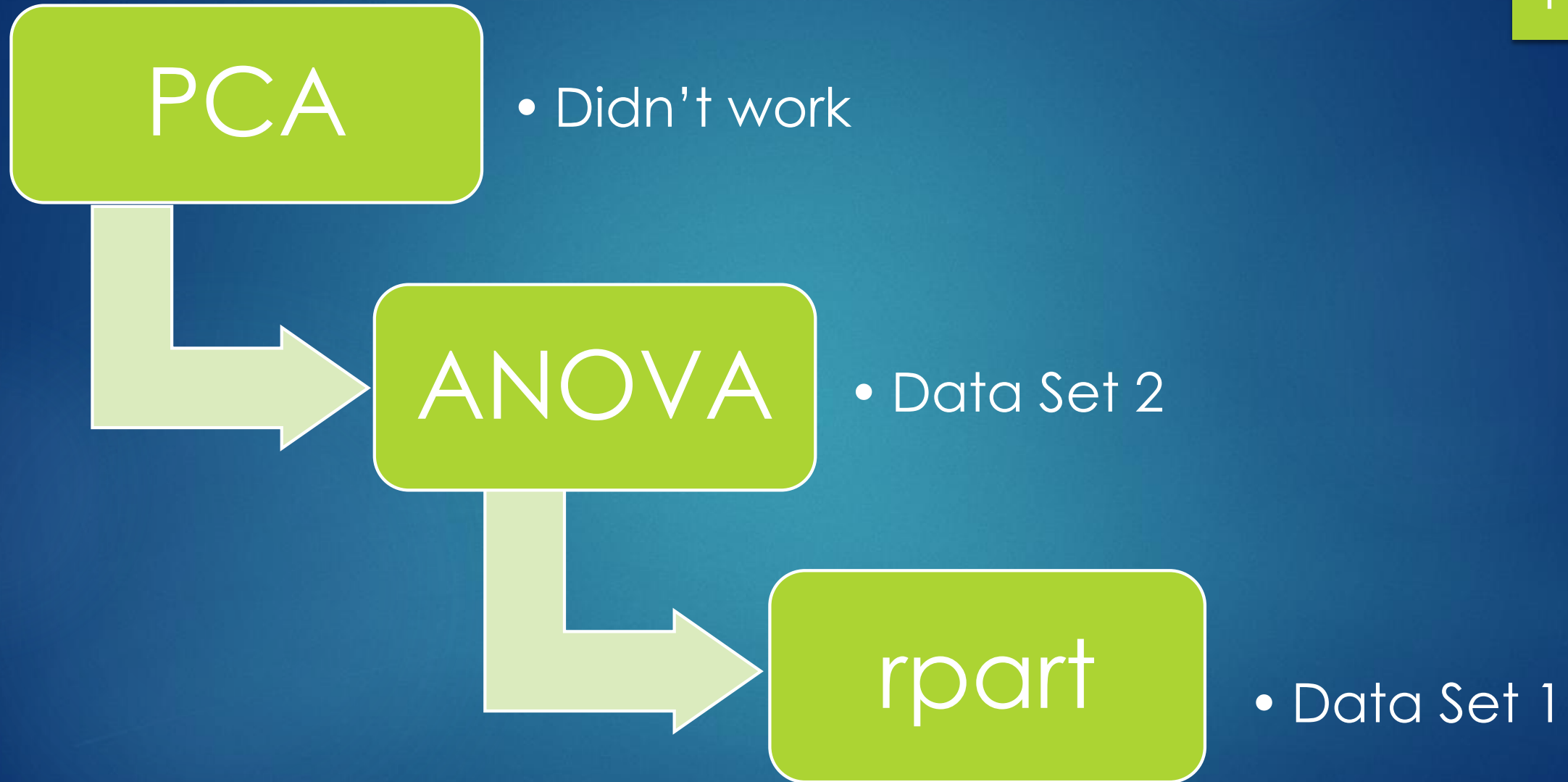
**4613
Observation**

	A	B	C	D	E	F	G	H	I	J
1 Cell	Cy72_CD4	CY58_1_Cl	Cy71_CD4	Cy81_FNA	Cy80_II_Cl	Cy81_Bulk	Cy72_CD4	Cy74_CD4	Cy71_CD4	
2 tumor	72	58	71	81	80	81	72	74	71	
3 malignant(1=no,2=yes,0=unresolved)	1	1	2	2	2	2	1	1	2	
4 non-malignant cell type (1=T,2=B,3=Macro.4=Endo.,5=CAF;6=NK)	2	1	0	0	0	0	1	1	0	
5 C9orf152	0	0	0	0	0	0	0	0	0	
6 RPS11	9.2172	8.3745	9.313	7.8876	8.3291	7.8336	8.3737	8.1338	8.4373	
7 ELMO2	0	0	2.1263	0	0	0.774	0	0	0	
8 CREB3L1	0	0	0	0	0	0	0	0	0	
9 PNMA1	0	0	0	0	0	0	0	0	0	
10 MMP2	0	0	0.73812	0	0	0	0	0	0	
11 TMEM216	0	0	0	0	3.7949	0	0	0	0	
12 TRAF3IP2-AS1	2.8514	2.0983	0.6173	0.96495	1.4735	3.158	1.308	1.3802	0	
13 LRRC37A5P	0	0	0	0	0	0	0	0	0	
14 LOC653712	0	0	0	0	0	0	0	0	0	
15 C10orf90	0	0	0	3.4069	1.4468	2.7876	0	0	1.7398	
16 ZHX3	0	0.52907	0	0.51197	0	1.4952	1.1097	0	1.2821	
17 ERCC5	0	0	0	0	2.2866	2.3741	0	0	0	
18 GPR98	0	0	0	0	0	0.0229	0	0	0	

Curse of Dimensionality

13

Still not able to use most of the algorithms to develop classification models



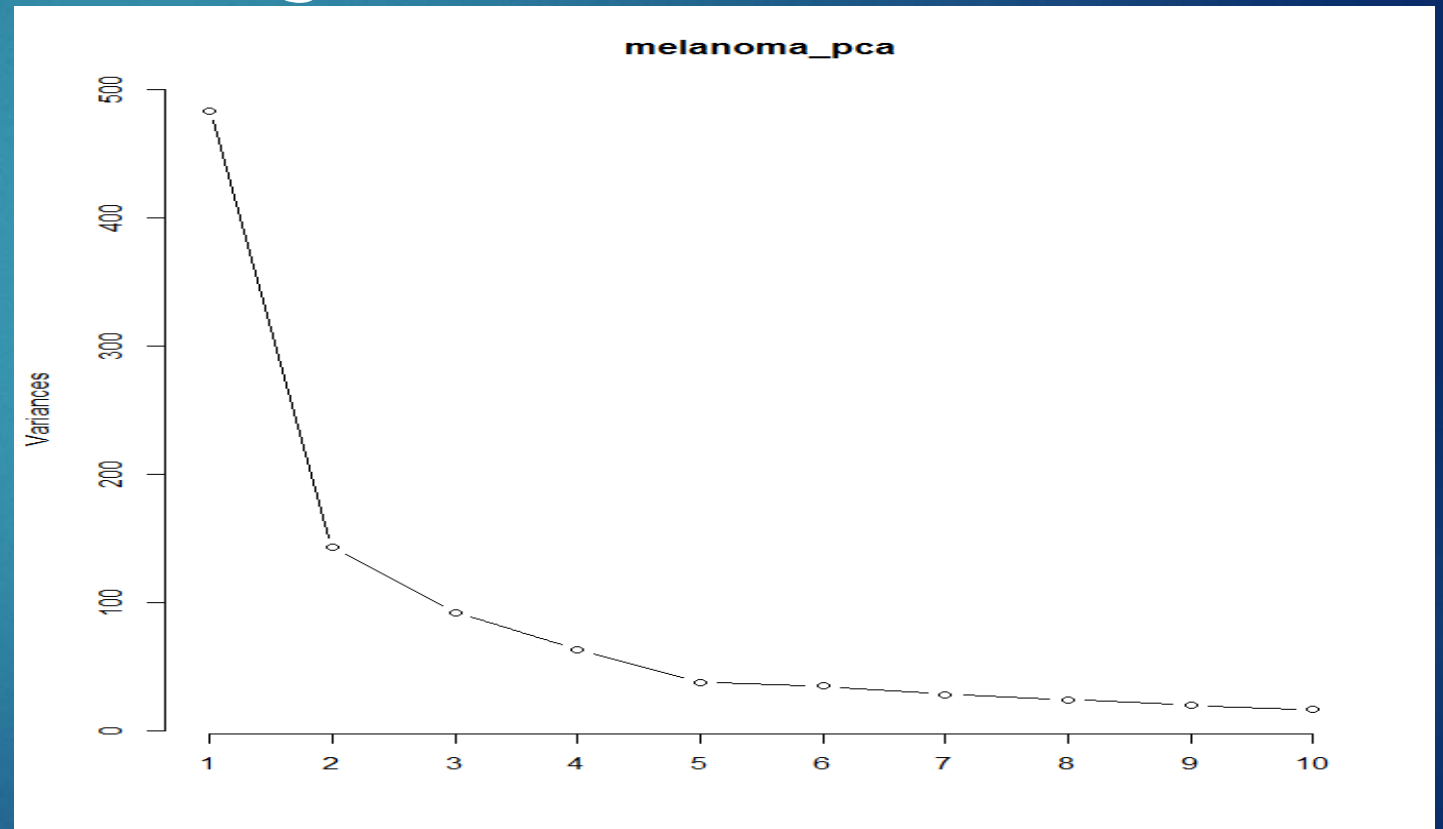
PCA

15

first 5 PCAs explaining
only 15% variance

```
cumsum_pca[1:5]
```

```
[1] 8.687328  
[2] 11.255989  
[3] 12.909616  
[4] 14.039679  
[5] 14.716901
```



ANOVA

16

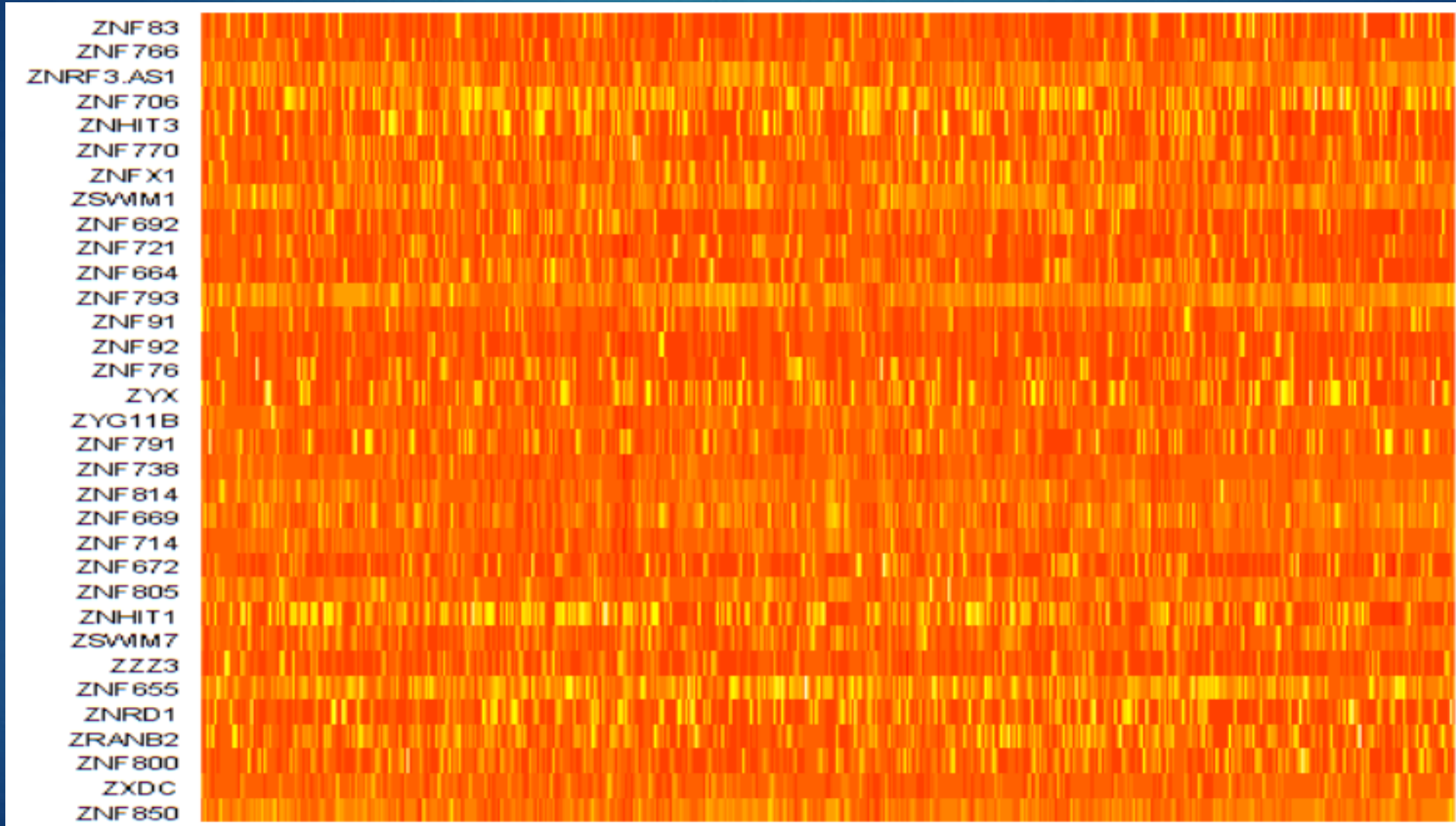
```
sep <- vb/ vw
```

```
> var_MaxSep
[1] "ZZZ3"      "ZYX"      "ZYG11B"   "ZXDC"     "ZSWIM7"   "ZSWIM1"
[7] "ZRANB2"    "ZNRFB3.AS1" "ZNRD1"    "ZNHIT3"   "ZNHIT1"   "ZNFx1"
[13] "ZNF92"     "ZNF91"     "ZNF850"   "ZNF83"    "ZNF814"   "ZNF805"
[19] "ZNF800"    "ZNF793"    "ZNF791"   "ZNF770"   "ZNF766"   "ZNF76"
[25] "ZNF738"    "ZNF721"    "ZNF714"   "ZNF706"   "ZNF692"   "ZNF672"
[31] "ZNF669"    "ZNF664"    "ZNF655"
```

a dataframe of 4513 x 33 -> data_set2 created to develop classification models

Gene Heatmap on Reduced Data

17



ANOVA
rpart

```
> rpart2_train
```

```
CART
```

```
4513 samples
```

```
 33 predictor
```

```
 2 classes: 'Benign', 'Malignant'
```

```
No pre-processing
```

```
Resampling: Cross-validated (10 fold)
```

```
Summary of sample sizes: 4062, 4062, 4062, 4062, 4062, 4061, ...
```

```
Resampling results across tuning parameters:
```

cp	Accuracy	Kappa
0.010	0.8701552	0.6612847
0.014	0.8630657	0.6395051
0.018	0.8601808	0.6332123
0.021	0.8601808	0.6332123
0.039	0.8528657	0.6223091
0.102	0.8335943	0.5467710
0.718	0.7214717	0.0000000

```
Accuracy was used to select the optimal model using the one SE rule.  
The final value used for the model was cp = 0.01.
```

87% Accuracy with 66 % kappa

ANOVA knn

```
> knn_train  
k-Nearest Neighbors  
  
4513 samples  
  33 predictor  
   2 classes: 'Benign', 'Malignant'
```

No pre-processing

Resampling: Cross-validated (10 fold, repeated 10 times)

Summary of sample sizes: 4061, 4062, 4062, 4061, 4062, 4061, ...

Resampling results across tuning parameters:

k	Accuracy	Kappa
3	0.8704442	0.6719509
5	0.8768047	0.6904313
7	0.8795751	0.6966258
9	0.8798626	0.6973234
11	0.8795322	0.6966849

Accuracy was used to select the optimal model using the one SE rule.

The final value used for the model was k = 11.

87% Accuracy with 67 % kappa

```
> naive_train
```

```
Naive Bayes
```

```
4513 samples
```

```
 33 predictor
```

```
 2 classes: 'Benign', 'Malignant'
```

```
No pre-processing
```

```
Resampling: Cross-validated (10 fold, repeated 10 times)
```

```
Summary of sample sizes: 4061, 4061, 4062, 4061, 4062, 4062, ...
```

```
Resampling results:
```

Accuracy	Kappa
----------	-------

0.8990478	0.7448179
-----------	-----------

```
Tuning parameter 'fL' was held constant at a value of 1
```

```
Tuning parameter
```

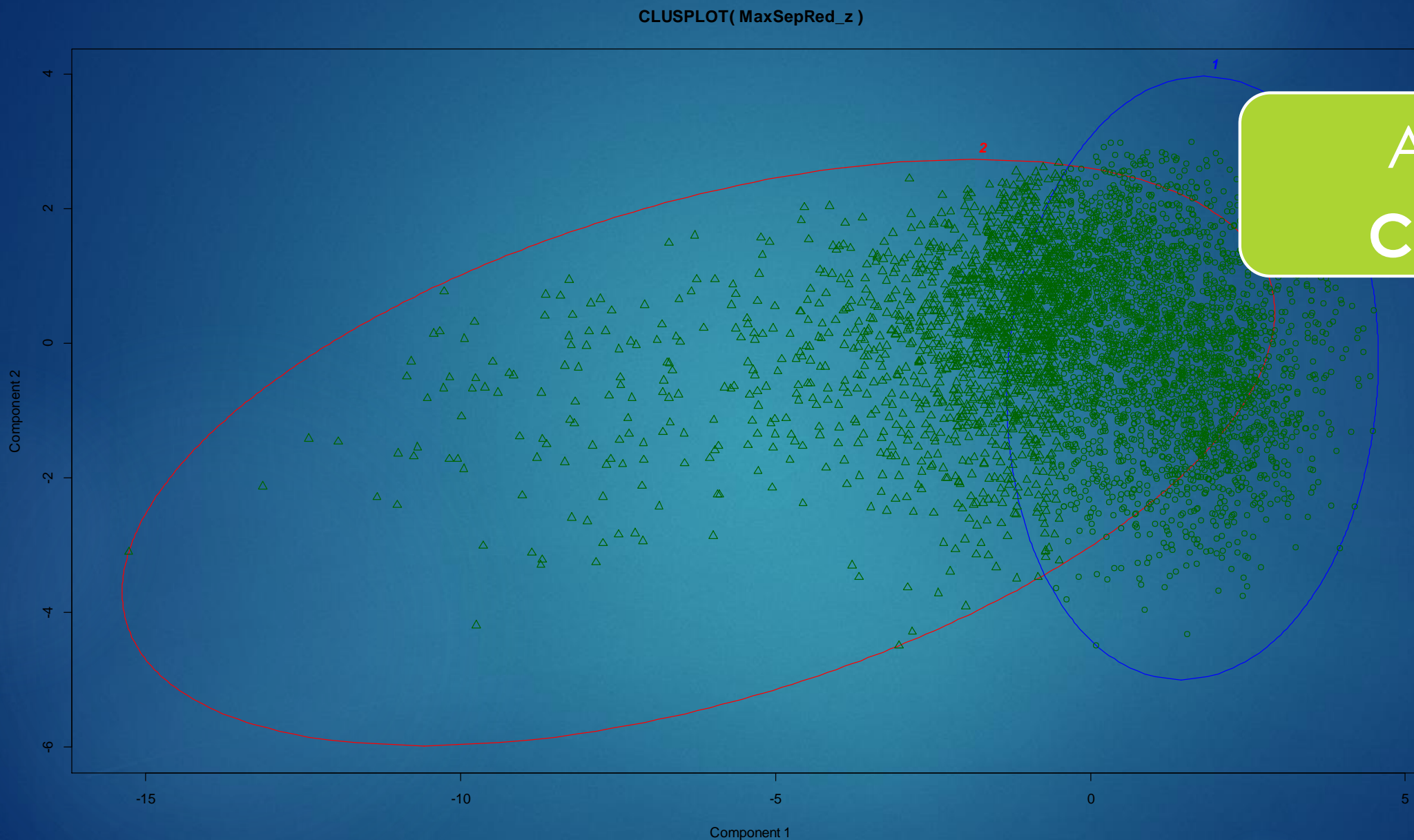
```
'usekernel' was held constant at a value of TRUE
```

```
Tuning parameter 'adjust' was
```

```
held constant at a value of 1
```

90% Accuracy with 75 % kappa

ANOVA CLUSTERS



These two components explain 20.2 % of the point variability.

rpart

22

```
> names(var_rpart)
[1] "PTPRC"      "ARHGDIB"    "CALD1"      "CNN3"       "LAPTM5"     "CTTN"
[7] "ERBB3"      "ASAHL"      "TCF4"       "AP1S2"      "ATP1A1"     "FDFT1"
[13] "DUSP4"      "CD63"       "TMEM98"     "GIMAP7"     "ITM2A"      "ATP1B1"
[19] "CD37"       "LCP1"       "TNFRSF1B"   "CD52"       "CXCR4"      "GAPDH"
[25] "RHOH"       "GDF15"      "PHLDA1"     "NT5DC3"     "SPP1"       "CAPN3"
[31] "CD59"       "FKBP10"     "LINC00673"
```

a dataframe of 4513 x 33 -> data_set1 created to develop classification models

```
> rpart_train
```

```
CART
```

```
3385 samples
```

```
 33 predictor
```

```
 2 classes: 'Benign', 'Malignant'
```

```
No pre-processing
```

```
Resampling: Cross-validated (10 fold)
```

```
Summary of sample sizes: 3047, 3046, 3047, 3046, 3047, 3047, ...
```

```
Resampling results across tuning parameters:
```

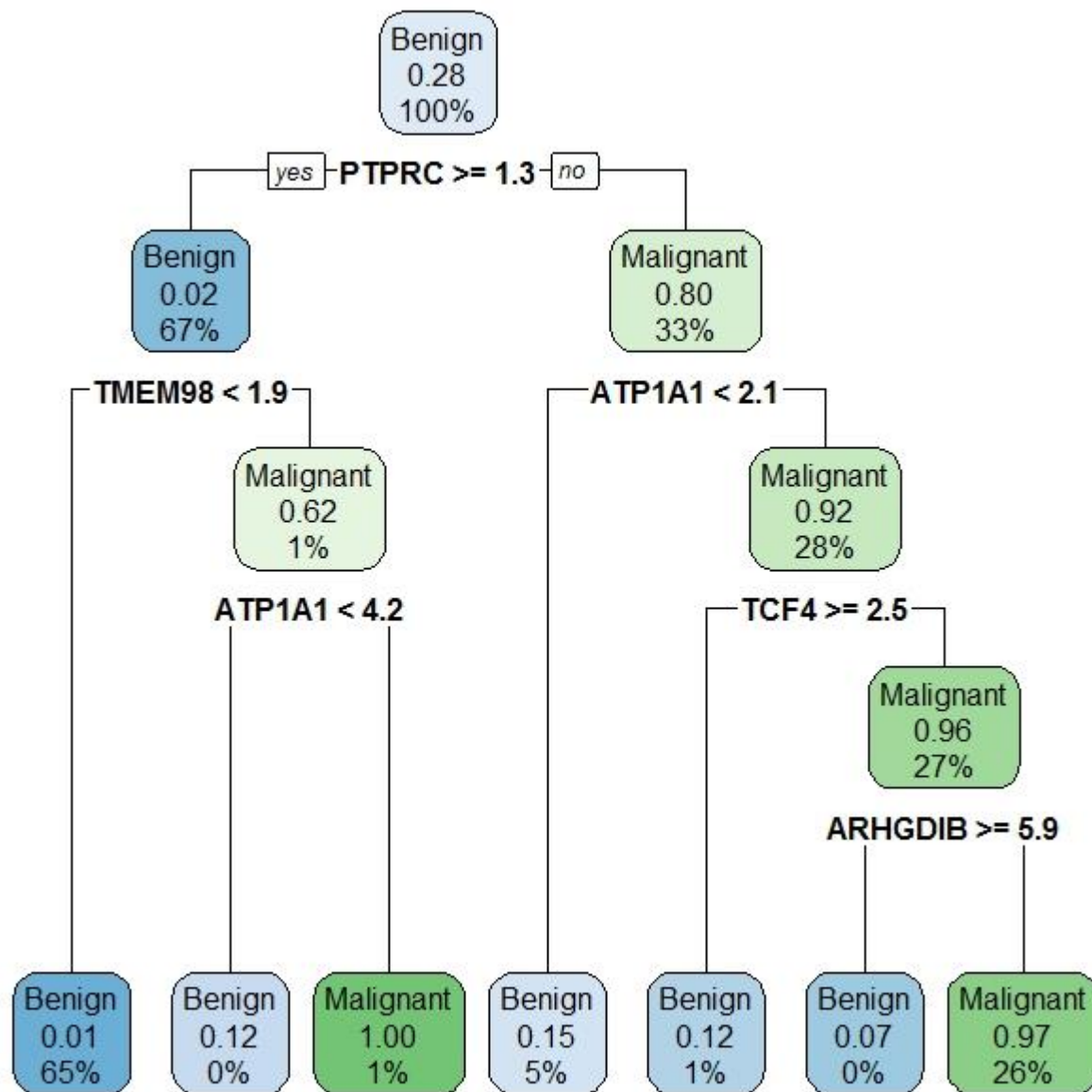
cp	Accuracy	Kappa
0.010	0.9695694	0.9241018
0.014	0.9636601	0.9088142
0.018	0.9648426	0.9115008
0.021	0.9639577	0.9095323
0.039	0.9521425	0.8812046
0.102	0.9518467	0.8804896
0.718	0.7559839	0.1509727

```
Accuracy was used to select the optimal model using the one SE rule.
```

```
The final value used for the model was cp = 0.01.
```

97% Accuracy with 92 % kappa

rpart



n= 3385

node), split, n, loss, yval, (yprob)

* denotes terminal node

- 1) root 3385 943 Benign (0.721418021 0.278581979)
- 2) PTPRC>=1.325 2253 41 Benign (0.981802042 0.018197958)
- 4) TMEM98< 1.92 2216 18 Benign (0.991877256 0.008122744) *
- 5) TMEM98>=1.92 37 14 Malignant (0.378378378 0.621621622)
- 10) ATP1A1< 4.19 16 2 Benign (0.875000000 0.125000000) *
- 11) ATP1A1>=4.19 21 0 Malignant (0.000000000 1.000000000) *
- 3) PTPRC< 1.325 1132 230 Malignant (0.203180212 0.796819788)
- 6) ATP1A1< 2.14 176 26 Benign (0.852272727 0.147727273) *
- 7) ATP1A1>=2.14 956 80 Malignant (0.083682008 0.916317992)
- 14) TCF4>=2.535 48 6 Benign (0.875000000 0.125000000) *
- 15) TCF4< 2.535 908 38 Malignant (0.041850220 0.958149780)
- 30) ARHGDIB>=5.895 14 1 Benign (0.928571429 0.071428571) *
- 31) ARHGDIB< 5.895 894 25 Malignant (0.027964206 0.972035794) *

97% Accuracy with 92 % kappa

```
> CrossTable(Test_set1$factor_redData,predict_rpart)
```

cell contents

	N
Chi-square contribution	
N / Row Total	
N / Col Total	
N / Table Total	

Total observations in Table: 1128

Test_set1\$factor_redData	predict_rpart		Row Total
	Benign	Malignant	
Benign	803	11	814
	71.839	196.488	
	0.986	0.014	0.722
	0.972	0.036	
	0.712	0.010	
Malignant	23	291	314
	186.233	509.367	
	0.073	0.927	0.278
	0.028	0.964	
	0.020	0.258	
column Total	826	302	1128
	0.732	0.268	

25

rpart

rpart

Confusion matrix on predicting
Class of Blind Holdout

97% Accuracy

```
> knn_train  
k-Nearest Neighbors
```

```
3385 samples  
 33 predictor  
 2 classes: 'Benign', 'Malignant'
```

No pre-processing

Resampling: Cross-validated (10 fold, repeated 10 times)

Summary of sample sizes: 3046, 3047, 3045, 3047, 3047, 3047, ...

Resampling results across tuning parameters:

k	Accuracy	Kappa
3	0.9915521	0.9790008
5	0.9906956	0.9768775
7	0.9899864	0.9751339
9	0.9901637	0.9755640
11	0.9896023	0.9741869

Accuracy was used to select the optimal model using the one SE rule.
The final value used for the model was k = 3.

99% Accuracy with 98 % kappa


```
> CrossTable(Test_set1$factor_redData, predict_knn)
```

```
cell contents
```

	N
Chi-square contribution	
N / Row Total	
N / Col Total	
N / Table Total	

```
Total observations in Table: 1128
```

Test_set1\$factor_redData	predict_knn		Row Total
	Benign	Malignant	
Benign	810	4	814
	81.138	215.056	
	0.995	0.005	0.722
	0.989	0.013	
	0.718	0.004	
Malignant	9	305	314
	210.339	557.501	
	0.029	0.971	0.278
	0.011	0.987	
	0.008	0.270	
column Total	819	309	1128
	0.726	0.274	

Confusion matrix on
predicting Class of
Blind Holdout

99% Accuracy

```
> naive_train
```

```
Naive Bayes
```

```
3385 samples
```

```
 33 predictor
```

```
 2 classes: 'Benign', 'Malignant'
```

```
No pre-processing
```

```
Resampling: Cross-Validated (10 fold, repeated 10 times)
```

```
Summary of sample sizes: 3046, 3047, 3047, 3046, 3045, 3047, ...
```

```
Resampling results:
```

Accuracy	Kappa
0.9835167	0.9594681

```
Tuning parameter 'fL' was held constant at a value of 1
```

```
Tuning parameter
```

```
'usekernel' was held constant at a value of TRUE
```

```
Tuning parameter 'adjust' was  
held constant at a value of 1
```

98% Accuracy with 96 % kappa

rpart naive

```
> CrossTable(Test_set1$factor_redData, predict_naive)
```

Cell Contents

	N
Chi-square contribution	
N / Row Total	
N / Col Total	
N / Table Total	

Total observations in Table: 1128

Test_set1\$factor_redData	predict_naive		Row Total
	Benign	Malignant	
Benign	798	16	814
	82.412	203.622	
	0.980	0.020	0.722
	0.994	0.049	
	0.707	0.014	
Malignant	5	309	314
	213.642	527.860	
	0.016	0.984	0.278
	0.006	0.951	
	0.004	0.274	
Column Total	803	325	1128
	0.712	0.288	

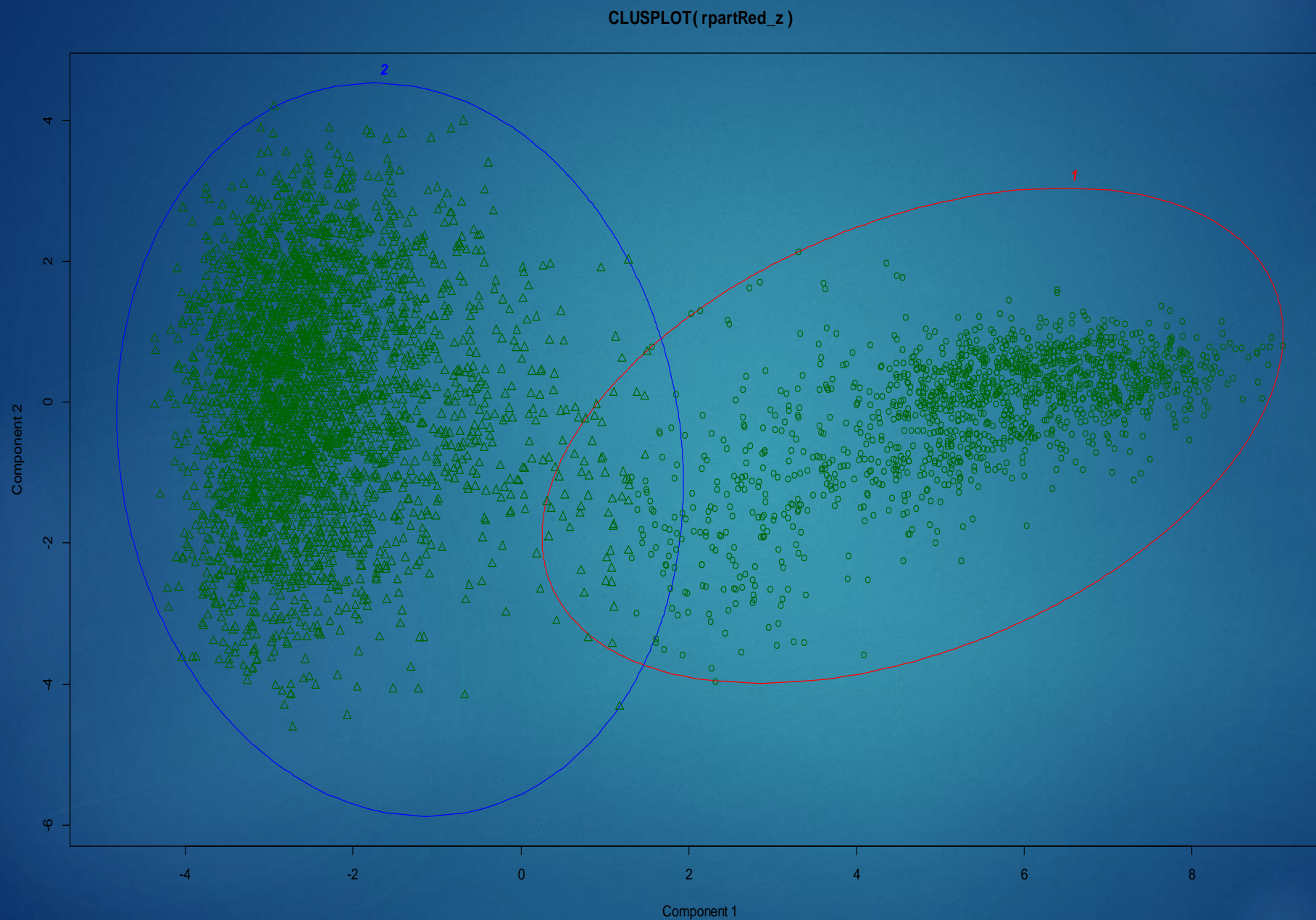
Confusion matrix on predicting
Class of Blind Holdout

only 5 Malignant are misclassified as
Benign*

98% Accuracy

30

rpart CLUSTERS



rpart bagged

```
> bagged_rpart$results
parameter Accuracy Kappa AccuracySD KappaSD
1 none 0.9853702 0.9635066 0.005051538 0.01269591
> predict_bagged <- predict(bagged_rpart, Test_set1)
> CrossTable(Test_set1$factor_redData, predict_bagged)
```

Cell Contents

	N
Chi-square contribution	
N / Row Total	
N / Col Total	
N / Table Total	

Total observations in Table: 1128

Test_set1\$factor_redData	predict_bagged		Row Total
	Benign	Malignant	
Benign	814	0	814
	86.745	225.871	
	1.000	0.000	0.722
	0.999	0.000	
	0.722	0.000	
Malignant	1	313	314
	224.875	585.537	
	0.003	0.997	0.278
	0.001	1.000	
	0.001	0.277	
Column Total	815	313	1128
	0.723	0.277	

Confusion matrix on predicting Class of Blind Holdout :

- Only 1 Malignant is misclassified as Benign*
- '0' Benign is misclassified as Malignant

rpart rf

```
> rf_rpart$results
  mtry Accuracy      Kappa AccuracySD      KappaSD
1     3 0.9918009 0.9795504 0.005228953 0.01308589
2     6 0.9915797 0.9789938 0.004650786 0.01166644
> predict_rf <- predict(rf_rpart, Test_set1)
> CrossTable(Test_set1$factor_redData, predict_rf)
```

cell contents

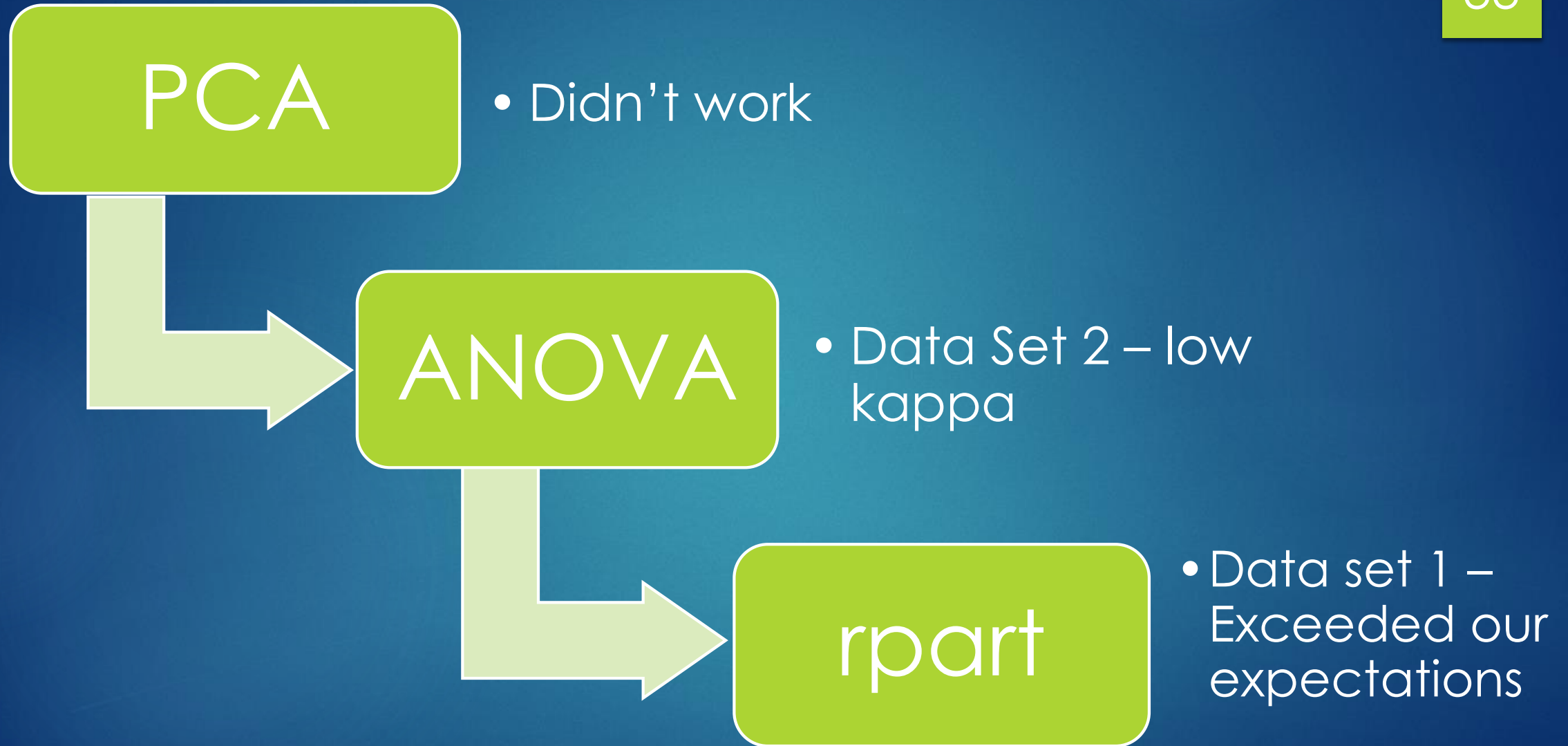
	N
Chi-square contribution	
N / Row Total	
N / Col Total	
N / Table Total	

Total Observations in Table: 1128

Test_set1\$factor_redData	predict_rf		Row Total
	Benign	Malignant	
Benign	814	0	814
	87.408	226.592	
	1.000	0.000	0.722
	1.000	0.000	
	0.722	0.000	
Malignant	0	314	314
	226.592	587.408	
	0.000	1.000	0.278
	0.000	1.000	
	0.000	0.278	
Column Total	814	314	1128
	0.722	0.278	

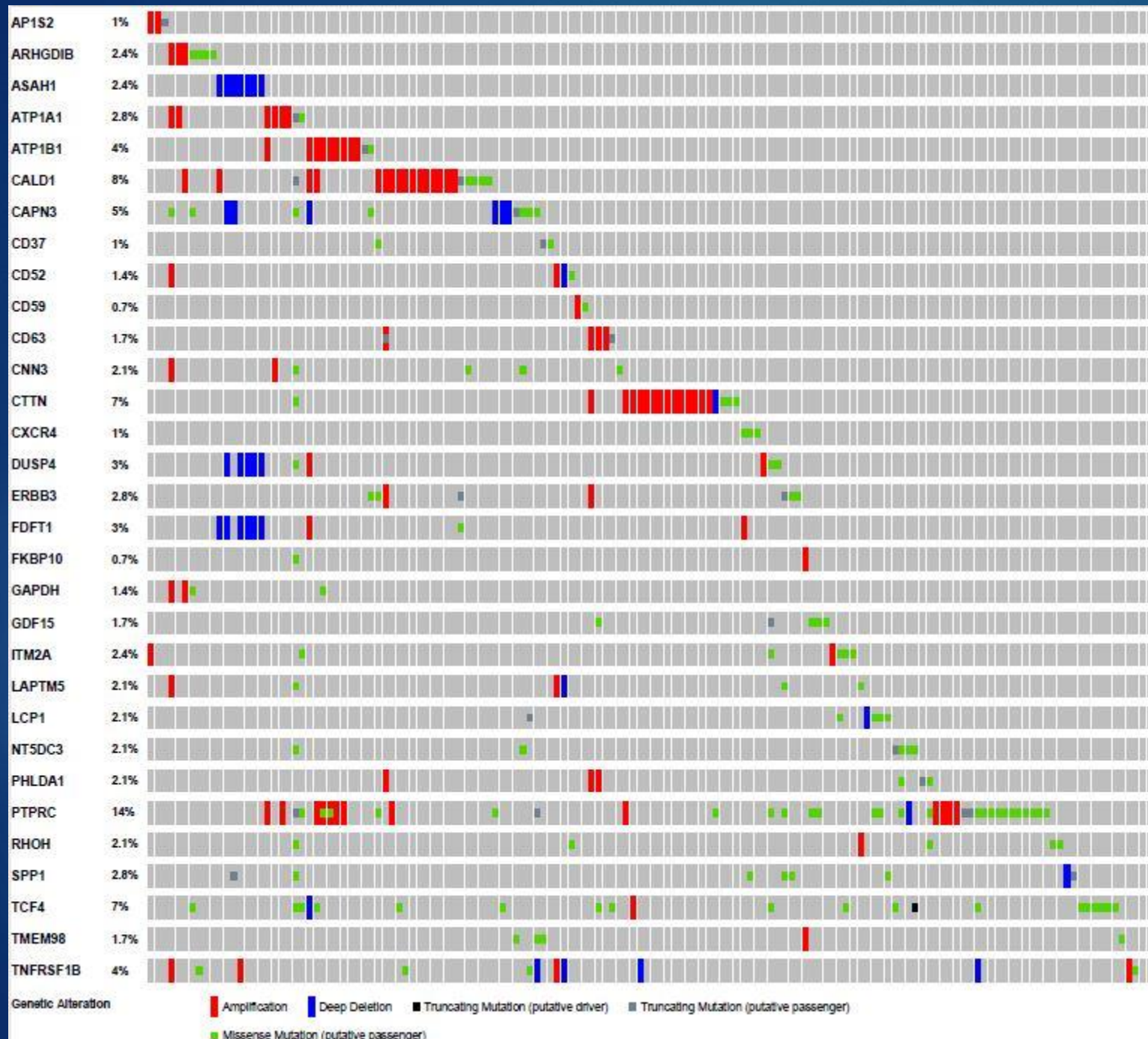
Confusion matrix on predicting Class of Blind Holdout :

- '0' Malignant is misclassified as Benign*
- '0' Benign is misclassified as Malignant



Eyeballing, some of the genes are related to immune response triggered in early response to a pathogen.

1	Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
2	SP_PIR_KEYWORDS	sodium/potassium transport	2	6.25	0.01282	ATP1B1, ATP1A1	32	8	19235	150.2734375	0.810783421	0.211669272	13.93686213
3	GOTERM_MF_FAT	GO:0005391~sodium:potassium-exchanging ATPase activity	2	6.25	0.01682	ATP1B1, ATP1A1	23	10	12983	112.8956522	0.895274319	0.895274319	17.99367858
4	GOTERM_CC_FAT	GO:0005890~sodium:potassium-exchanging ATPase complex	2	6.25	0.01801	ATP1B1, ATP1A1	30	8	12782	106.5166667	0.89111769	0.308976388	18.88269812
5	GOTERM_MF_FAT	GO:0005523~tropomyosin binding	2	6.25	0.02182	CNN3, CALD1	23	13	12983	86.84280936	0.946798763	0.623896614	22.73397508
6	GOTERM_BP_FAT	GO:0002920~regulation of humoral immune response	2	6.25	0.02196	PTPRC, CD37	26	12	13528	86.71794872	0.999998328	0.669926058	27.95561016
7	SP_PIR_KEYWORDS	blocked carboxyl end	2	6.25	0.02863	CD59, CD52	32	18	19235	66.78819444	0.976409635	0.340533498	28.66536109
8	COG_ONTOLOGY	Cytoskeleton	2	6.25	0.03155	CNN3, LCP1	3	31	1950	41.93548387	0.062104927	0.062104927	6.981634769
9	GOTERM_MF_FAT	GO:0017022~myosin binding	2	6.25	0.03665	CXCR4, CALD1	23	22	12983	51.31620553	0.993031832	0.711078634	35.37860124
10	GOTERM_BP_FAT	GO:0051017~actin filament bundle formation	2	6.25	0.03813	CALD1, LCP1	26	21	13528	49.55311355	1	0.766672115	43.67292147
11	INTERPRO	IPR018503:Tetraspanin, conserved site	2	6.25	0.04432	CD37, CD63	30	26	16659	42.71538462	0.986522855	0.883908895	39.24476767
12	GOTERM_BP_FAT	GO:0002637~regulation of immunoglobulin production	2	6.25	0.047	PTPRC, CD37	26	26	13528	40.02366864	1	0.763478818	50.87449701
13	GOTERM_BP_FAT	GO:0042098~T cell proliferation	2	6.25	0.04876	PTPRC, CXCR4	26	27	13528	38.54131054	1	0.759714456	52.20061329
14	SP_PIR_KEYWORDS	phosphatidylinositol linkage	2	6.25	0.04881	CD59, CD52	32	31	19235	38.78024194	0.998427689	0.475610663	44.11886017
15	GOTERM_BP_FAT	GO:0031032~actomyosin structure organization	2	6.25	0.05052	CNN3, CAPN3	26	28	13528	37.16483516	1	0.740812467	53.49102615
16	INTERPRO	IPR000301:Tetraspanin, subgroup	2	6.25	0.05097	CD37, CD63	30	30	16659	37.02	0.993056411	0.809222125	43.73234446
17	PIR_SUPERFAMILY	PIRSF002419:CD9 antigen	2	6.25	0.05227	CD37, CD63	23	18	7396	35.7294686	0.800216677	0.800216677	36.70126362
18	INTERPRO	IPR018499:Tetraspanin	2	6.25	0.05593	CD37, CD63	30	33	16659	33.65454545	0.995777904	0.597973496	46.87977507
19	GOTERM_CC_FAT	GO:0005765~lysosomal membrane	2	6.25	0.05954	LAPTM5, CD63	30	27	12782	31.56049383	0.999441132	0.437926648	50.67927244
20	GOTERM_BP_FAT	GO:0002700~regulation of production of molecular mediator of immune response	2	6.25	0.07314	PTPRC, CD37	26	41	13528	25.38086304	1	0.791702446	67.42051988
21	GOTERM_BP_FAT	GO:0046651~lymphocyte proliferation	2	6.25	0.07486	PTPRC, CXCR4	26	42	13528	24.77655678	1	0.788502065	68.30095094
22	GOTERM_BP_FAT	GO:0007162~negative regulation of cell adhesion	2	6.25	0.07657	ERBB3, ARHGD1B	26	43	13528	24.20035778	1	0.764487001	69.15765186
23	GOTERM_BP_FAT	GO:0032943~mononuclear cell proliferation	2	6.25	0.07828	PTPRC, CXCR4	26	44	13528	23.65034965	1	0.752196464	69.99126054
24	GOTERM_BP_FAT	GO:0002699~positive regulation of immune effector process	2	6.25	0.07828	PTPRC, CD37	26	44	13528	23.65034965	1	0.752196464	69.99126054
25	GOTERM_BP_FAT	GO:0070661~leukocyte proliferation	2	6.25	0.07828	PTPRC, CXCR4	26	44	13528	23.65034965	1	0.752196464	69.99126054
26	GOTERM_CC_FAT	GO:0043197~dendritic spine	2	6.25	0.08281	CNN3, CALD1	30	38	12782	22.4245614	0.999973688	0.462228528	63.03583772
27	GOTERM_MF_FAT	GO:0015662~ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	2	6.25	0.08613	ATP1B1, ATP1A1	23	53	12983	21.30106645	0.999993724	0.819362852	65.11554623
28	GOTERM_BP_FAT	GO:0032102~negative regulation of response to external stimulus	2	6.25	0.08849	TNFRSF1B, SPP1	26	50	13528	20.81230769	1	0.776870095	74.54160001
29	GOTERM_CC_FAT	GO:0005884~actin filament	2	6.25	0.08906	CALD1, LCP1	30	41	12782	20.78373984	0.999988571	0.468571055	65.83327329
30	SP_PIR_KEYWORDS	EF hand	2	6.25	0.09238	CAPN3, LCP1	32	60	19235	20.03645833	0.999996286	0.617795704	67.60588844
31	SP_PIR_KEYWORDS	actin binding	3	9.375	0.00189	CNN3, CALD1, LCP1	32	40	19235	45.08203125	0.216272572	0.114716188	2.172968065
32	GOTERM_BP_FAT	GO:0048585~negative regulation of response to stimulus	3	9.375	0.01453	PTPRC, TNFRSF1B, SPP1	26	100	13528	15.60923077	0.999844221	0.714193125	19.43581476
33	UP_SEQ_FEATURE	calcium-binding region:2	3	9.375	0.01489	FKBP10, CAPN3, LCP1	32	115	19113	15.58125	0.889839159	0.668095133	16.35299631
34	GOTERM_BP_FAT	GO:0007204~elevation of cytosolic calcium ion concentration	3	9.375	0.0174	PTPRC, CXCR4, CD52	26	110	13528	14.19020979	0.999972861	0.650570152	22.83253068



Oncoprint

144 (50%) of 287 patients had an alteration in at least one of this 33 gene set.

Publishing a paper

It would be great if subject matter experts can look at our findings so we are planning to publish a conference paper.

Future Work

37

Expanding the technique to other cancer types

Q & A