Contents

[Figures for abstract: 1](#_Toc506068926)

[Clustering agreement (external validation) 2](#_Toc506068927)

[Gene Expression heat map For 4 classes: 3](#_Toc506068928)

[Important genes (2 classes (TAC3a, TAC3b)): 3](#_Toc506068929)

[Gene Ranking (TAC1,TAC2,TAC3a.TAC3b) 6](#_Toc506068930)

[Association to Clinical variables: 8](#_Toc506068931)

[Classes (TAC1and2, against TAC3aandb) 8](#_Toc506068932)

[Classes (TAC3a, TAC3b) 9](#_Toc506068933)

[Box plots - Bivariate Analysis: 10](#_Toc506068934)

[TAC3a and TAC3b 10](#_Toc506068935)

[TAC12 and TAC3a3b: 11](#_Toc506068936)

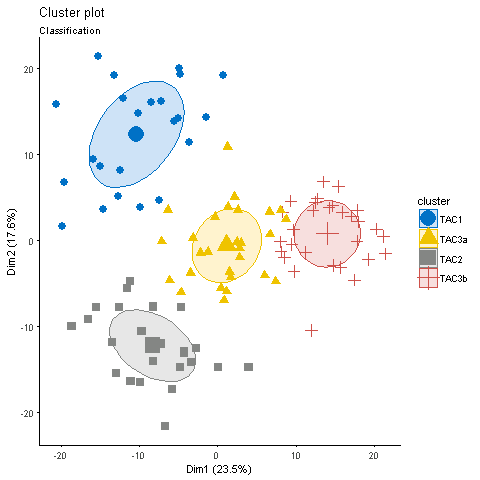
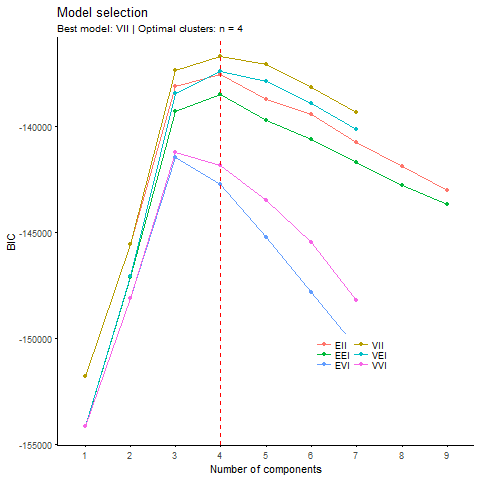
[Between all TACs 14](#_Toc506068937)

[Pathway Analysis: 15](#_Toc506068938)

[TAC3a and TAC3b 15](#_Toc506068939)

[TAC1:3b 16](#_Toc506068940)

# Figures for abstract:



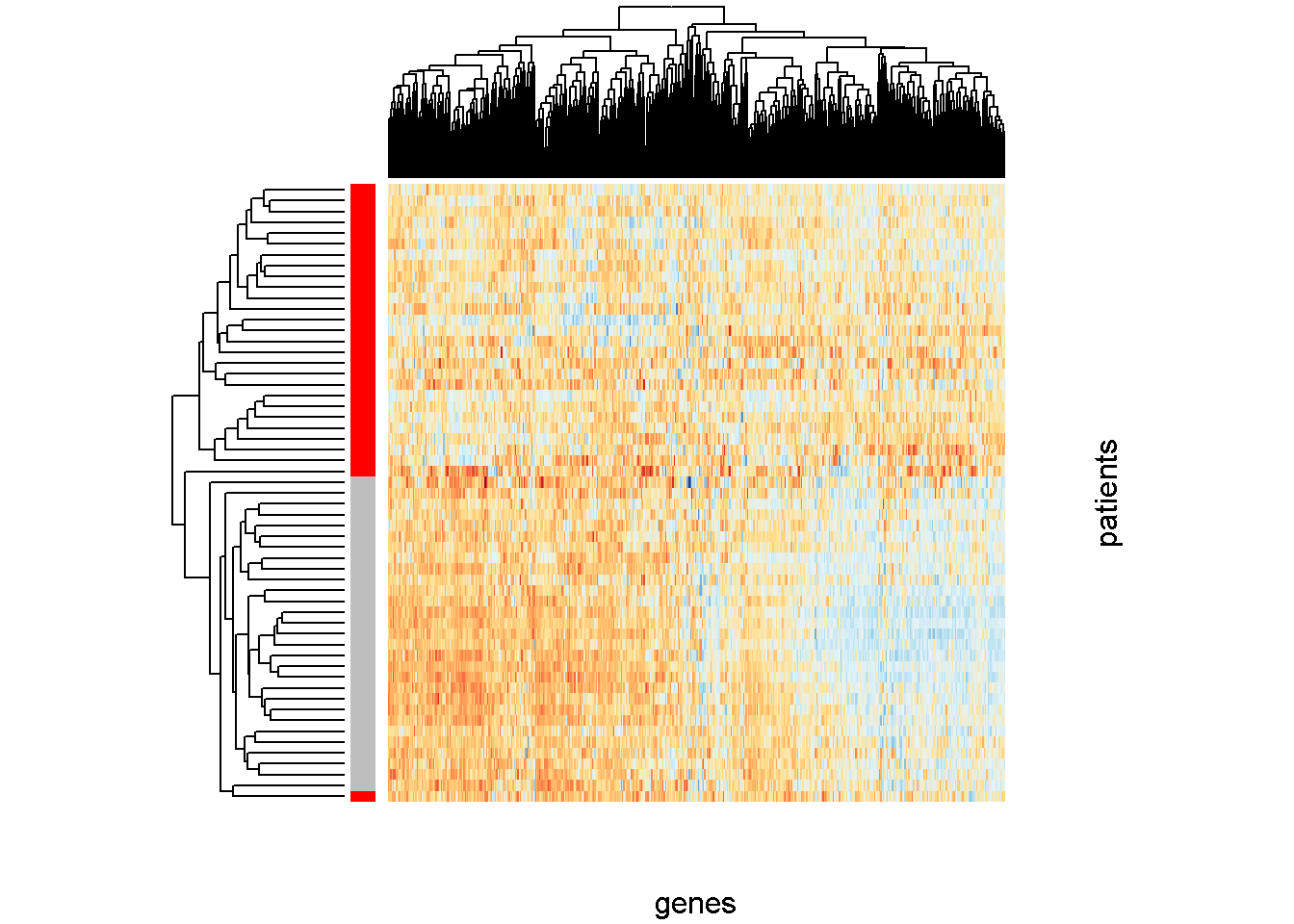
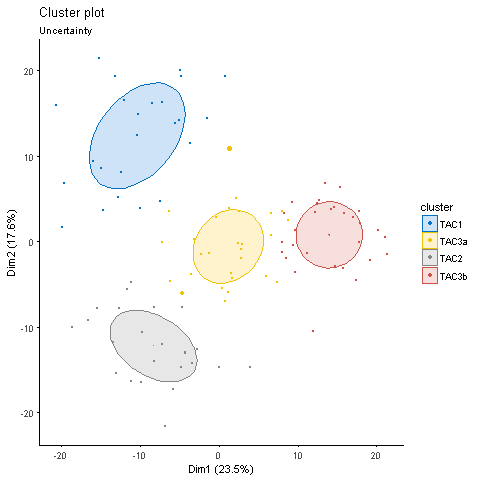


Figure 1- Top left figure Shows traces of Bayesian Information Criterion (BIC) as the function of number of clusters for all the considered models. There is a clear indication of the existence of4 clusters with covariance having variable volume but the same shape and orientation (VII). Top right PCA plot of the data that visualizes the contour of the estimated clusters. The horizontal axis is the first principal component and the vertical axis is the second. Bottom left class membership uncertainty plot that represents how the fitted model recovers the true labels. Each dot represent one patient in the data set and the size of the dot reflects the uncertainty levels of the patient. There are only two patients with relatively high uncertainty. Bottom right visualises the gene expression levels of the patients in the TAC3a and TAC3b clusters. In the colour coded band attached to the heat map, red colour correspond to the patients in the TAC3a cluster and grey colour represent the patients in the TAC3b.

# Clustering agreement (external validation)

Confusion matrix:

## myLabels TAC1 TAC2 TAC3

## tacLabels

## TAC1 23 1 6

## TAC2 0 22 0

## TAC3 0 1 51

The model based clustering moves 7 patients from TAC1 (based on Kuo et al. ) to other TACS. To be precise, 1 to TAC 2 and 6 to TAC3. TAC2 remains intact. One patient from TAC3 is replaced to TAC2 by model based clustering.

Finally, TAC1 cluster in our study has 7 less patients (1 is assigned to TAC2 and 6 to TAC3). TAC2 has two more patients, one patient was allocated to TAC1 by Kuo et al. and the other to TAC3. TAC3 in our study has 6 more patients, all these patients were considered to be TAC1 by Kuo et al.

Paul’s code:

mutual\_information(tacLabels,myLabels)

##

## 0.7698318

homogeneity\_completeness\_vmeasure(tacLabels,myLabels)

## $homogeneity

##

## 0.7446766

##

## $completeness

##

## 0.7685528

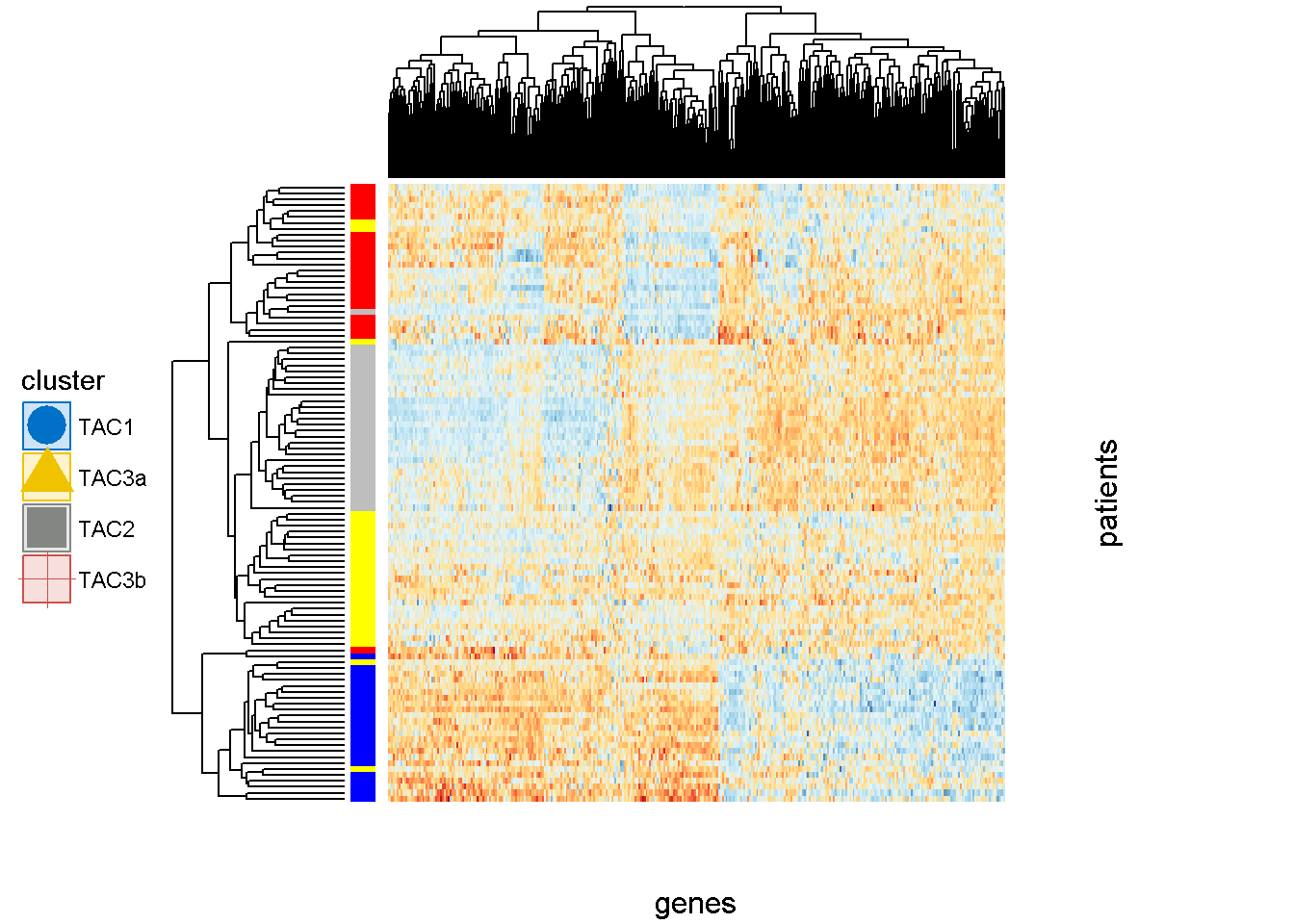
##

## $vmeasure

##

## 0.7564263

# Gene Expression heat map For 4 classes:



# Important genes (2 classes (TAC3a, TAC3b)):

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | t.test | welch.test | wilcox.test | f.test | kruskal.test | limma | rfe | rf | lasso | elasticnet | boosting | golub | shrinkcat | foldChange | majority voting |
| CPA3 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -2.48654 | 6 |
| ZKSCAN8 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.297125 | 5 |
| FCGR1B | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.49851 | 5 |
| LCP2 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0.070676 | 5 |
| TIPRL | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1.42477 | 4 |
| GGT5 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | -0.1919 | 4 |
| UMPS | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.679818 | 4 |
| LOC284454 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0.968724 | 4 |
| ATG2A | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.01543 | 4 |
| SH2D2A | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.705876 | 4 |
| MAP3K1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0.697207 | 4 |
| MRPS33 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.127616 | 4 |
| FTSJ2 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7.360166 | 4 |
| GTF2H2B | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -4.10763 | 3 |
| ETS2 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | -0.16952 | 3 |
| PRDX3 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9.254599 | 3 |
| BTG1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0.212585 | 3 |
| TGOLN2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | -4.17153 | 3 |
| GINM1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1.62877 | 3 |
| P2RY10 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.263826 | 3 |
| RNF135 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 3.710306 | 3 |
| NR4A1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.162824 | 3 |
| ZNF395 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.40605 | 3 |
| IGLV@ | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.29408 | 3 |
| GZMB | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -5.92302 | 3 |
| SNX30 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.236887 | 3 |
| DESI2 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1.20412 | 3 |
| SLC25A36 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.63468 | 3 |
| ADORA2A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0.155223 | 3 |
| IFITM2 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.52304 | 3 |
| IL1RL1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | -1.11874 | 2 |
| LGALS12 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1.89022 | 2 |
| TLR7 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.616064 | 2 |
| UTRN | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -0.09813 | 2 |
| ARHGAP1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1.12318 | 2 |
| SCOC | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.982288 | 2 |
| REL | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.469703 | 2 |
| TBC1D2B | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | -9.83413 | 2 |
| BANK1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.172962 | 2 |
| ANKRD10 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.125352 | 2 |
| LINC01094 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.06349 | 2 |
| SCARB2 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -8.96869 | 2 |
| TPSB2 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.90462 | 2 |
| C2CD5 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -2.36878 | 2 |
| SLC7A11 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19.12891 | 2 |
| ZYG11B | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.148143 | 2 |
| ACAA1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.113046 | 2 |
| SFMBT2 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.24747 | 2 |
| ADAM10 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2.44524 | 2 |
| DNASE1L3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | -5.50376 | 2 |
| NFKBID | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.33843 | 2 |
| TANC2 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.322528 | 2 |
| ASAP1-IT1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.04563 | 2 |
| OAS3 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1.025127 | 2 |
| SLC7A11-AS1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.97076 | 2 |
| FGFR1OP2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | -7.0176 | 2 |
| FLOT1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.43086 | 2 |
| GTPBP10 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8.368588 | 2 |
| MGAM | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.96917 | 2 |
| VPS36 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1.23616 | 2 |
| BBS10 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.006928 | 2 |
| KRT23 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.64371 | 2 |
| PPFIBP2 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.06133 | 2 |
| NOLC1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.270837 | 2 |
| ATP1B1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 4.324306 | 2 |
| MIOS | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -3.23651 | 2 |
| MYH10 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.929009 | 2 |
| PHF19 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5.472933 | 2 |
| BIRC3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -8.86015 | 1 |
| RUNX3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | -0.48033 | 1 |
| RNF146 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -3.51611 | 1 |
| CTSS | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -3.05895 | 1 |
| LOC101060424 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -3.88961 | 1 |
| COA7 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.755941 | 1 |
| CYLD | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.549025 | 1 |
| XRCC5 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4.178892 | 1 |
| DUSP4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | -0.41386 | 1 |
| DAD1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3040.922 | 1 |
| PLBD1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3.70161 | 1 |
| STX7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | -1.00859 | 1 |
| CAMP | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 34.02843 | 1 |
| ASXL1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3.489065 | 1 |
| ADAM19 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1.04883 | 1 |
| IL2RA | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.251843 | 1 |
| RNPEP | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -6.94433 | 1 |
| BTBD19 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.12912 | 1 |
| COPG1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -1.29802 | 1 |
| C8orf60 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.236308 | 1 |
| HN1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.311273 | 1 |
| C11orf49 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1.27669 | 1 |
| BROX | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11.1949 | 1 |
| COQ2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.01679 | 1 |
| CSTA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 21.68059 | 1 |
| IRF4 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.71919 | 1 |
| LOC100130264 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.68727 | 1 |
| ZCRB1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -4.60628 | 1 |
| PDIA6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 4.577526 | 1 |
| ZFYVE26 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.963537 | 1 |
| FCGR3B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -0.10885 | 1 |
| BRCC3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | -3.75219 | 1 |
| PCCA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | -2.52766 | 1 |
| ZNF587B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 9.457273 | 1 |
| IL3RA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | -1.25287 | 1 |
| HSD17B11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -0.53017 | 1 |
| DACH1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1.48812 | 1 |
| NT5DC1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | -0.04468 | 1 |
| KRCC1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.3827 | 1 |
| YES1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.267401 | 1 |
| AP3M1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.527138 | 1 |
| CR1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | -0.45791 | 1 |
| MARCKSL1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.64897 | 1 |
| MRVI1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.49563 | 1 |
| LOC158402 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | -12.3334 | 1 |
| PRF1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.300393 | 1 |
| CLEC12A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0.501283 | 1 |
| ING2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0.229712 | 1 |
| SLC25A37 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.285616 | 1 |
| MYO6 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1.11351 | 1 |
| TTC7A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0.973929 | 1 |
| SAMSN1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0.041328 | 1 |
| NUP98 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.84177 | 1 |
| GNG2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -0.25842 | 1 |
| PREPL | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0.310278 | 1 |
| FAM65B | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.612149 | 1 |
| IFITM1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0.695653 | 1 |
| POLR2M | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.352445 | 1 |
| FAM162A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 11.2957 | 1 |
| SMCHD1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | -0.95891 | 1 |
| NINJ1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | -0.25245 | 1 |
| HLA-DMB | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 3.850917 | 1 |
| LIPA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2.453322 | 1 |
| MRPL57 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1.410857 | 1 |
| CLC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -2.24538 | 0 |

# Gene Ranking (TAC1,TAC2,TAC3a.TAC3b)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Gene identifier | t.test | kruskalWalis | Random forest | Component boosting | Fold change | Sum of votes |
| CCNYL1 | 1 | 1 | 1 | 0 | 0.824561 | 3 |
| TRAM1 | 1 | 0 | 1 | 0 | 0.824561 | 2 |
| FCGR3B | 1 | 1 | 0 | 0 | 0.824561 | 2 |
| FCGR1B | 1 | 1 | 0 | 0 | 0.824561 | 2 |
| SBF2 | 1 | 1 | 0 | 0 | 0.824561 | 2 |
| CASP5 | 1 | 0 | 1 | 0 | -1.21277 | 2 |
| SYNE1 | 1 | 1 | 0 | 0 | -1.21277 | 2 |
| NSL1 | 1 | 1 | 0 | 0 | 0.824561 | 2 |
| OR2A7 | 1 | 1 | 0 | 0 | 0.824561 | 2 |
| CPA3 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| GATA2 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| SMOX | 0 | 0 | 1 | 0 | -1.21277 | 1 |
| FCER1A | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| TRGV9 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| CD207 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| ADORA3 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| IL18R1 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| DNAJB1 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| STX7 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| ARPC3 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| PTBP3 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| SCARB2 | 0 | 0 | 1 | 0 | 0.824561 | 1 |
| ETS2 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| FCF1 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| CUL2 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| TMEM120A | 0 | 0 | 1 | 0 | -1.21277 | 1 |
| LOC284454 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| C2CD5 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| BTG1 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| ERN1 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| MAP3K1 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| RNF135 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| PDIA6 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| SFMBT2 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| FBXO3 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| ZFYVE26 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| SRFBP1 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| FCER2 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| MMP12 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| TLR1 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| IFI16 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| NR4A1 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| CASP4 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| GIMAP4 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| TLR8 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| F13A1 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| RRN3P2 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| NFKBID | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| HSD17B11 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| UBE2D1 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| MSL2 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| DPH1 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| ARHGEF3 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| OLIG2 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| ACSL1 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| DACH1 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| CPEB4 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| VAV1 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| HMG20B | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| NT5DC1 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| RNF149 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| CLEC4D | 0 | 0 | 1 | 0 | -1.21277 | 1 |
| STK3 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| CDC73 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| TNFSF10 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| NDFIP2 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| LOC101927066 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| IFIH1 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| UTP11L | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| RAB11FIP3 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| OAS3 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| NSUN7 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| BAZ1A | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| GZMB | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| OR4F3 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| PRF1 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| LRP5L | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| LOC100129406 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| TSPAN2 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| BCL9 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| SNX30 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| FBXL13 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| NCR3LG1 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| RRAS2 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| AMFR | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| MRPS25 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| KRT23 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| ADM | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| NOLC1 | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| TNFAIP3 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| PLK3 | 1 | 0 | 0 | 0 | -1.21277 | 1 |
| PREPL | 0 | 1 | 0 | 0 | 0.824561 | 1 |
| ATP1B1 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| SULT1B1 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| MRPS33 | 1 | 0 | 0 | 0 | 0.824561 | 1 |
| KCNJ15 | 0 | 1 | 0 | 0 | -1.21277 | 1 |
| NLRP2 | 1 | 0 | 0 | 0 | 0.824561 | 1 |

# Association to Clinical variables:

## Classes (TAC1and2, against TAC3aandb)

##

## Call:

## glm(formula = Label ~ ., data = significant\_cov)

##

## Deviance Residuals:

## Min 1Q Median 3Q Max

## -0.65666 -0.20271 -0.00471 0.19547 0.68554

##

## Coefficients:

## Estimate Std. Error t value Pr(>|t|)

## (Intercept) -7.257e-01 2.269e-01 -3.198 0.00188 \*\*

## IgE\_Total\_.IU.ml. -1.004e-04 6.228e-05 -1.613 0.11013

## Pct\_Neutrophils 7.796e-03 1.788e-03 4.362 3.28e-05 \*\*\*

## Pct\_Macrophages 1.973e-02 2.029e-03 9.723 6.47e-16 \*\*\*

## Eczema\_Diag0sed1 -1.236e-01 6.831e-02 -1.809 0.07355 .

## Allergic\_Rhinitis1 -1.096e-01 6.574e-02 -1.667 0.09871 .

## Sinusitis\_Diag0sed1 -1.649e-01 7.107e-02 -2.320 0.02249 \*

## FEV1\_Change 6.718e-03 2.293e-03 2.930 0.00425 \*\*

## FVC\_Pct\_.L. 3.052e-03 1.744e-03 1.750 0.08333 .

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## (Dispersion parameter for gaussian family taken to be 0.09894232)

##

## Null deviance: 25.7596 on 103 degrees of freedom

## Residual deviance: 9.3995 on 95 degrees of freedom

## AIC: 65.151

##

## Number of Fisher Scoring iterations: 2

confIntervals = exp(confint(fitted\_model\_best, level = 0.95))

## Waiting for profiling to be done...

print(confIntervals)

## 2.5 % 97.5 %

## (Intercept) 0.3102085 0.7550381

## IgE\_Total\_.IU.ml. 0.9997775 1.0000216

## Pct\_Neutrophils 1.0043021 1.0113640

## Pct\_Macrophages 1.0158771 1.0239900

## Eczema\_Diag0sed1 0.7730032 1.0103352

## Allergic\_Rhinitis1 0.7878445 1.0194128

## Sinusitis\_Diag0sed1 0.7377375 0.9747513

## FEV1\_Change 1.0022259 1.0112751

## FVC\_Pct\_.L. 0.9996341 1.0064906

## Classes (TAC3a, TAC3b)

In the context of available clinical measurement backward logistic regression fitted to the clinical data on TAC3a and TAC3b samples.

##

## Call:

## glm(formula = Label ~ ., data = significant\_cov)

##

## Deviance Residuals:

## Min 1Q Median 3Q Max

## -0.57550 -0.24464 -0.00918 0.18566 1.00282

##

## Coefficients:

## Estimate Std. Error t value Pr(>|t|)

## (Intercept) 4.069828 0.494632 8.228 8.58e-11 \*\*\*

## Periostin\_.ng.mL. -0.009029 0.003617 -2.496 0.0160 \*

## Pct\_Neutrophils -0.012045 0.002471 -4.875 1.18e-05 \*\*\*

## Pct\_Lymphocytes 0.100730 0.038424 2.622 0.0116 \*

## Allergic\_Rhinitis1 -0.170307 0.104257 -1.634 0.1088

## wbc -0.034981 0.021319 -1.641 0.1072

## FEV1\_Change -0.006387 0.003058 -2.089 0.0419 \*

## FEV1.FVC\_PreSalbutamol 0.011818 0.004653 2.540 0.0143 \*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## (Dispersion parameter for gaussian family taken to be 0.1219133)

##

## Null deviance: 14.2456 on 56 degrees of freedom

## Residual deviance: 5.9738 on 49 degrees of freedom

## AIC: 51.185

##

## Number of Fisher Scoring iterations: 2

confIntervals = exp(confint(fitted\_model\_best))

print(confIntervals)

## 2.5 % 97.5 %

## (Intercept) 22.2061009 154.3600624

## Periostin\_.ng.mL. 0.9840124 0.9980616

## Pct\_Neutrophils 0.9832545 0.9928230

## Pct\_Lymphocytes 1.0257466 1.1924845

## Allergic\_Rhinitis1 0.6875316 1.0346190

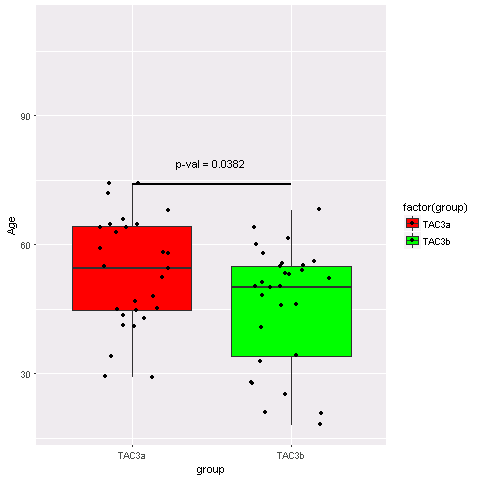
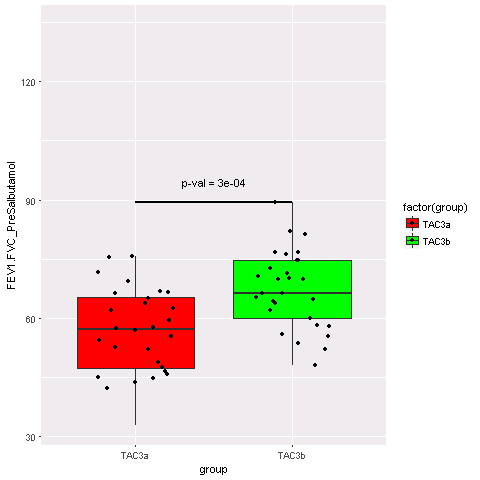
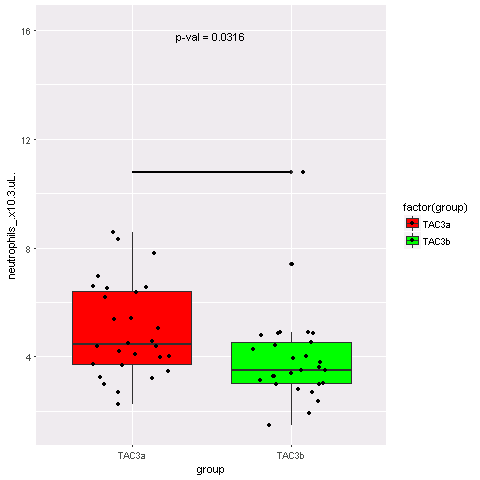
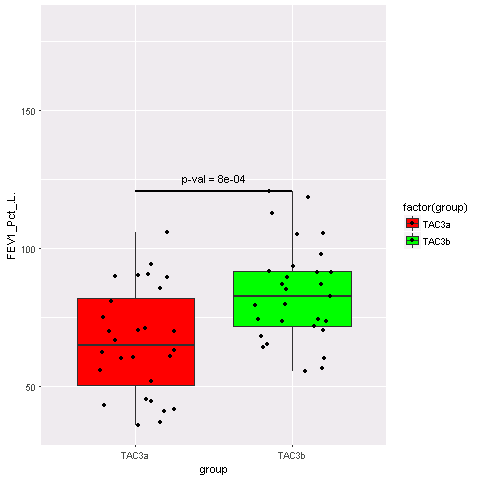
## wbc 0.9261065 1.0068265

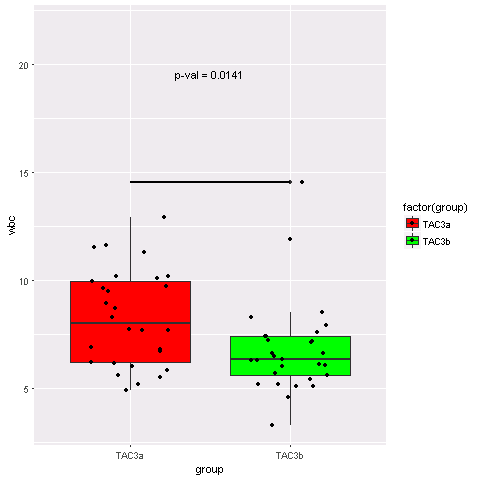
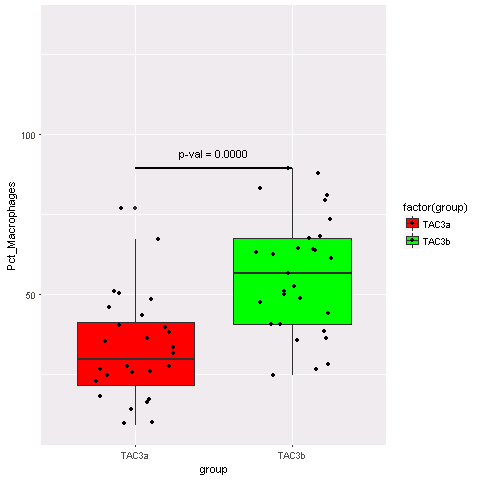
## FEV1\_Change 0.9876953 0.9996063

## FEV1.FVC\_PreSalbutamol 1.0027024 1.0211574

# Box plots - Bivariate Analysis:

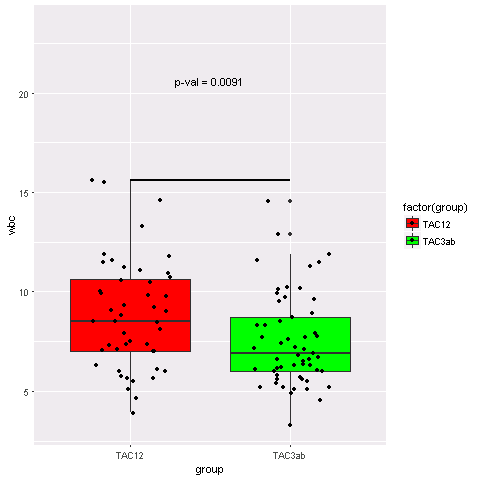
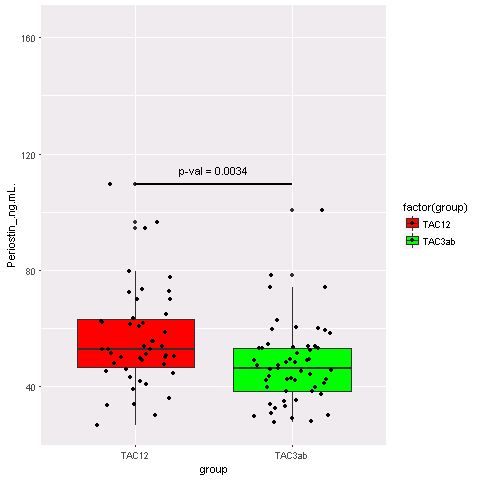
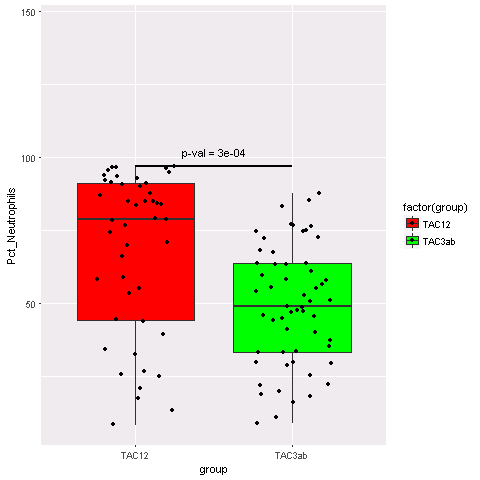
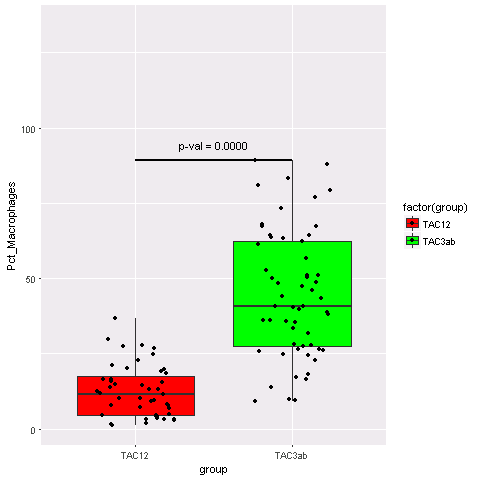
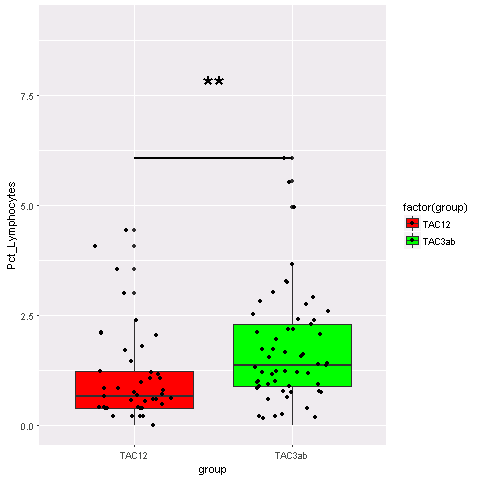
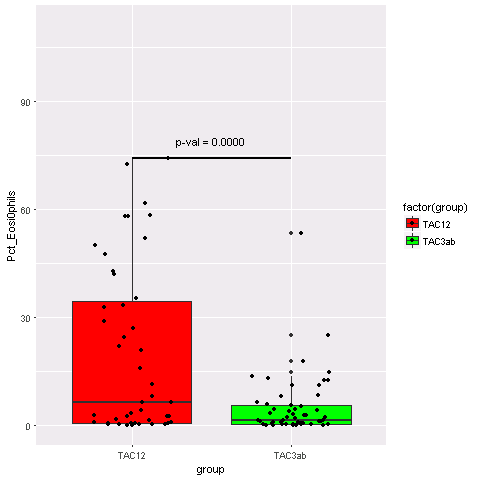
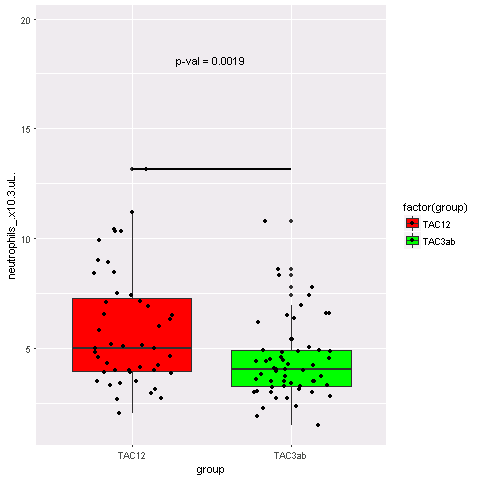
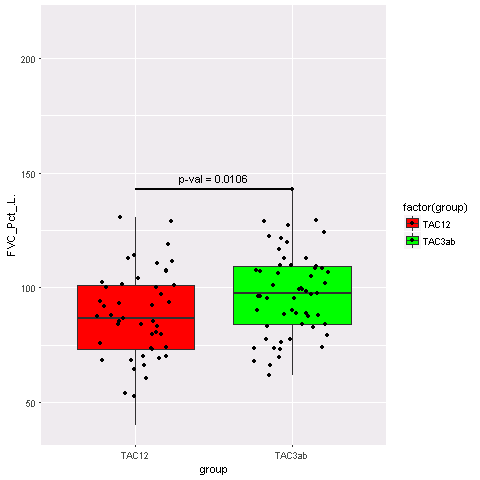
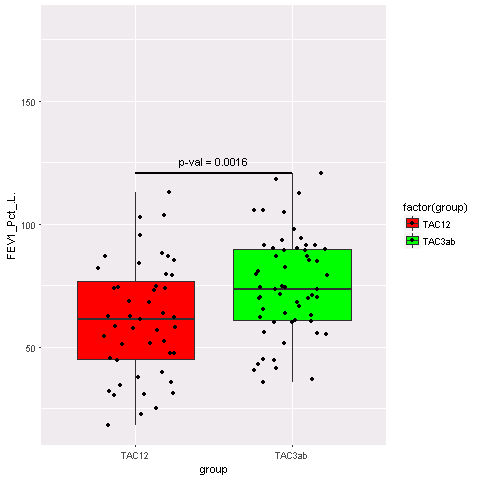
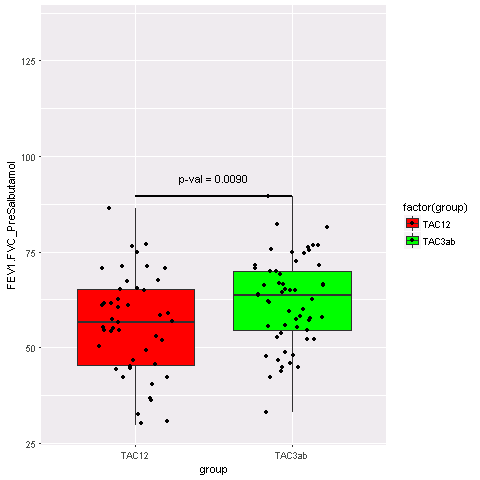
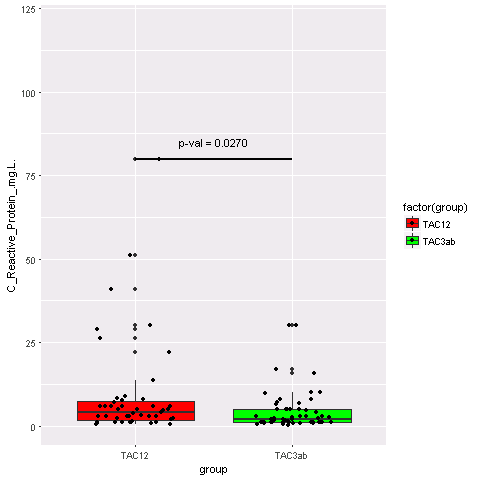
## TAC3a and TAC3b

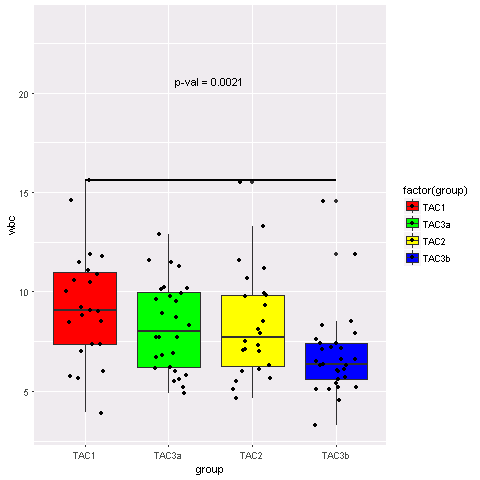
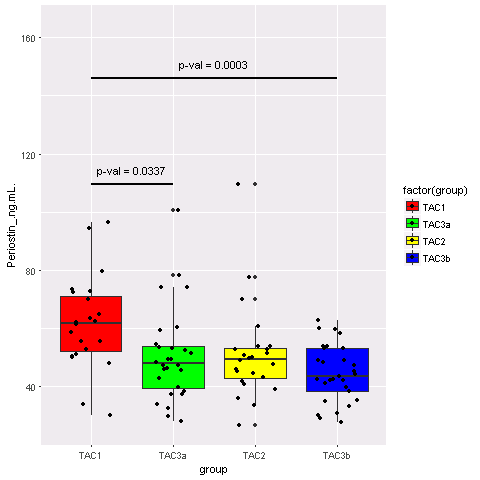
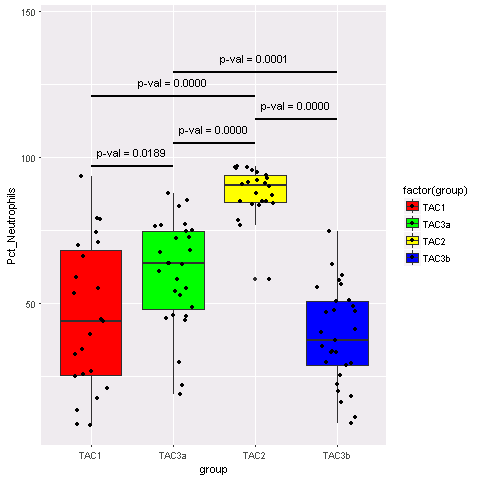
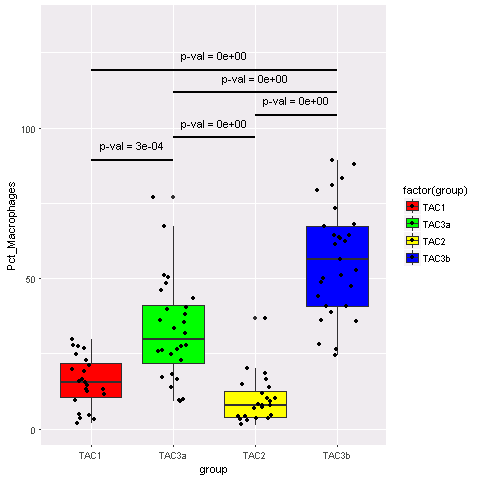
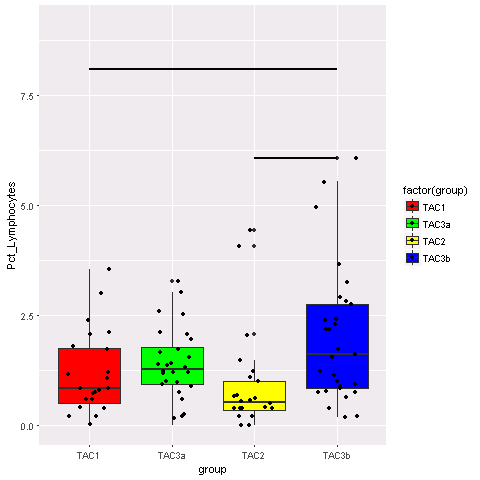
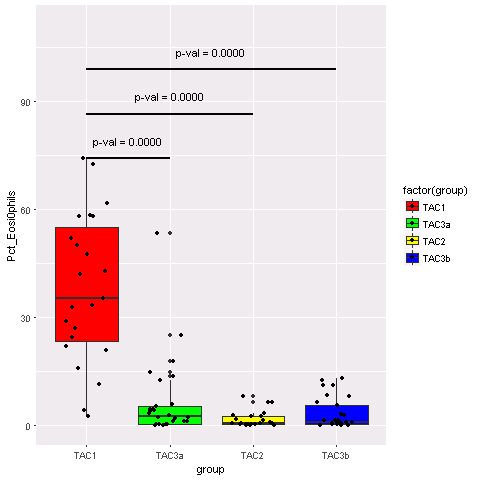
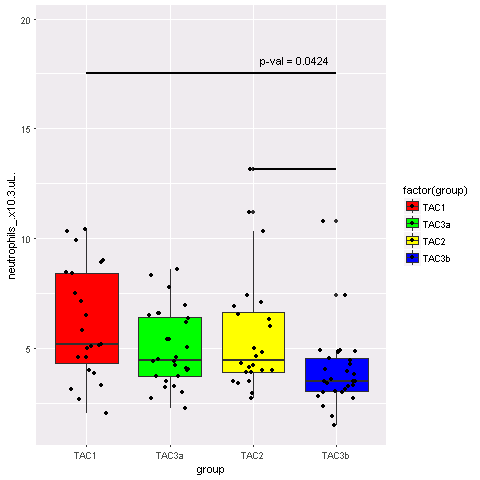
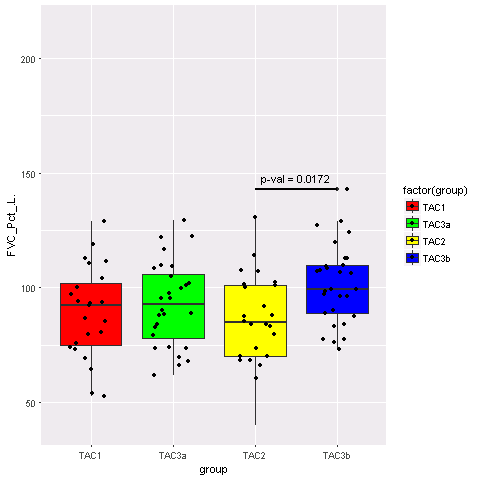
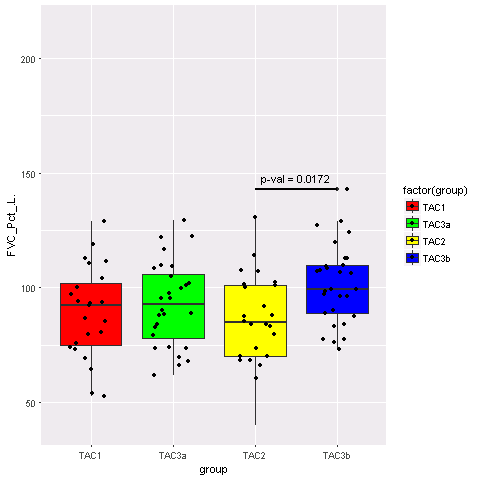
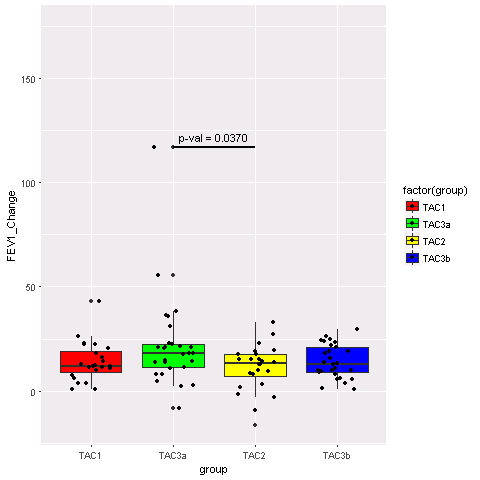
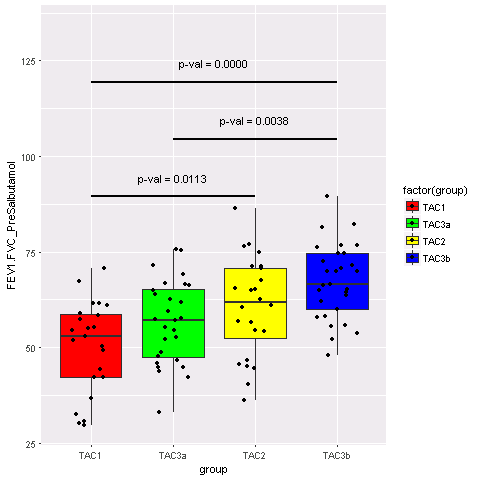
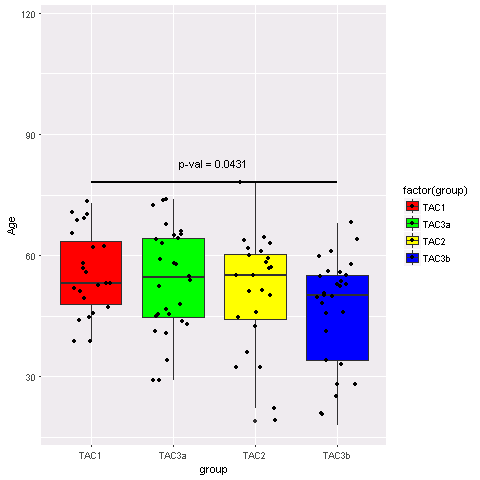


Bivariate Analysis:

## TAC12 and TAC3a3b:



## Between all TACs



# Pathway Analysis:

## TAC3a and TAC3b

13 feature selection techniques were used to find genes that are differently expressed between TAC3a and TAC3b. only 67 were predicted to be important with pValue smaller that 0.05, by more than one technique. Using gprofiler it is shown that 4 genes (of 67) are contributing to ‘ regulation of interferon-beta production’ with pvalue 4.57e-02.

Gprofiler

# signf corr. p-value T Q Q&T Q&T/Q Q&T/T term ID t type t group t name and depth in group Q&T list

#

1 ! 4.57e-02 47 63 [4](https://biit.cs.ut.ee/gprofiler/index.cgi?organism=hsapiens&sort_by_structure=1&significant=1&output=txt&ordered_query=0&user_thr=1.00&query=FLOT1+REL+RNF135+TLR7) 0.063 0.085 [GO:0032648](http://amigo.geneontology.org/amigo/term/GO:0032648) BP 1 regulation of interferon-beta production 1 FLOT1,REL,RNF135,TLR7

From literature

Simpson, Jodie L., et al. "Reduced antiviral interferon production in poorly controlled asthma is associated with neutrophilic inflammation and high-dose inhaled corticosteroids." *Chest* 149.3 (2016): 704-713.

“Asthma is a heterogeneous chronic inflammatory disease in which host defense against respiratory viruses such as human rhinovirus (HRV) may be abnormal. This is a matter of some controversy, with some investigators reporting reduced type I interferon (IFN) synthesis and others suggesting that type I IFN synthesis is relatively normal in asthma.

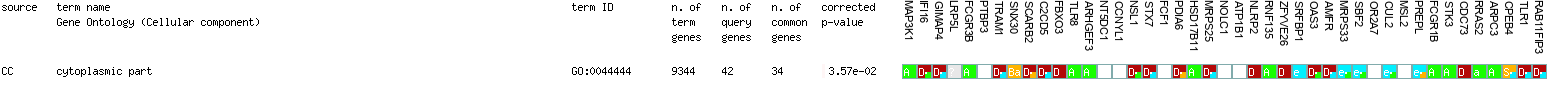
Conclusion Antiviral type I IFN production is impaired in those with neutrophilic airway inflammation and in those prescribed high doses of inhaled corticosteroids. Our study is an important step toward identifying those with poorly controlled asthma who might respond best to inhaled IFN therapy during exacerbations.”

## TAC1:3b

Only genes that are ranked as important by more than one method are used for pathway analysis.

Among them the ones with positive fold change (expressed with higher level in TAC3a and TAC3b compared to TAC1 and TAC2) were selected and were queried from gprofiler.

This resulted in 42 genes. There is a huge overlap between this gene set and the cytoplasmic part (*Cellular component*) path way, 32.



From literature :

Kicic, Anthony, et al. "Decreased fibronectin production significantly contributes to dysregulated repair of asthmatic epithelium." *American journal of respiratory and critical care medicine* 181.9 (2010): 889-898.

“Fifteen cellular component ontologies with the most significantly decreased expression in the asthmatic epithelium relative to healthy non-atopic control

There is a table with cytoplasmic part ”

Under-expressed genes

Only genes that are ranked as important by more than one method are used for pathway analysis.

Among them the ones with negative fold change (expressed with lower level in TAC3a and TAC3b compared to TAC1 and TAC2) were selected and were queried from gprofiler.

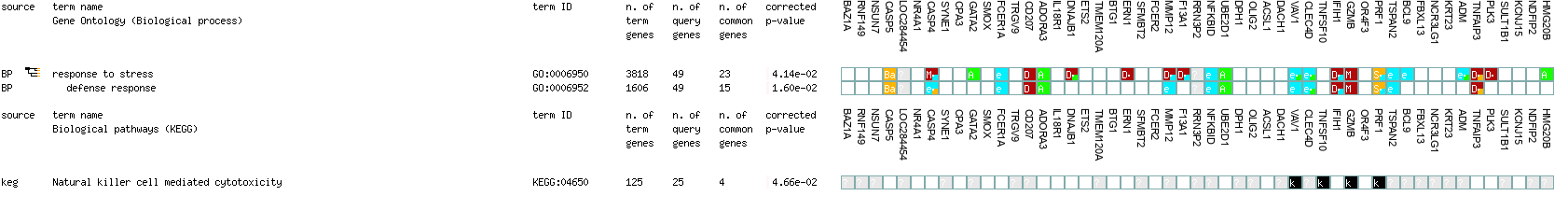
This resulted in 49 genes. There is a relatively high overlap between this gene set and the response to stress (*Cellular component*) 23 and defense response path way, 15.

p-value Q Q&T Q&T/Q t name t depth Q&T list

0.0414 49 23 0.469 response to stress 1 HMG20B,UBE2D1,GZMB,IFIH1,CD207,BCL9,TNFAIP3,F13A1,DNAJB1,TSPAN2,CASP5,VAV1,ADM,CLEC4D,NFKBID,PLK3,ERN1,GATA2,FCER1A,PRF1,CASP4,MMP12,ADORA3

0.016 49 15 0.306 defense response 2 UBE2D1,GZMB,IFIH1,CD207,TNFAIP3,TSPAN2,CASP5,VAV1,CLEC4D,NFKBID,FCER1A,PRF1,CASP4,MMP12,ADORA3

0.0466 25 4 0.16 Natural killer cell mediated cytotoxicity 1 GZMB,TNFSF10,VAV1,PRF1



Literature search

Senapati, J., Devasia, A.J., Alex, A.A., George, B., Albadainah, F., Khader, J., Salah, S. and Salem, A., Hematology/Oncology and Stem Cell Therapy.

“For example, in asthma, NK cells contribute towards the progress of T cell mediated allergic airway response during allergen specific sensitization phase.[38,39](https://www.sciencedirect.com/science/article/pii/S1658387614001083" \l "b0190) Existing evidence also suggests that NK cells are involved in resolving acute allergic airway inflammation.”