Chapter IV:

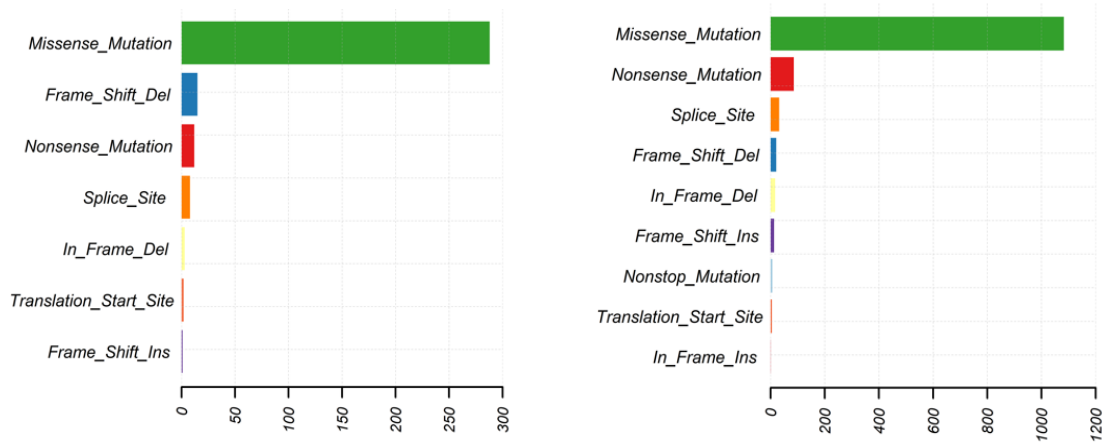
Results

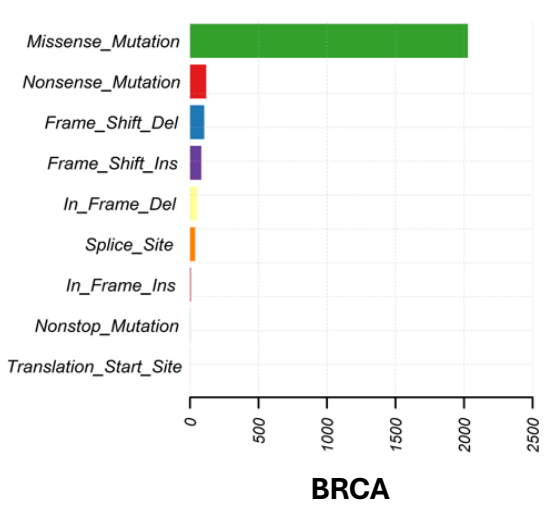
4.1. Characterization of immune related genes with genomic instability in the human genome with respect to cancer

4.1.1. Results:

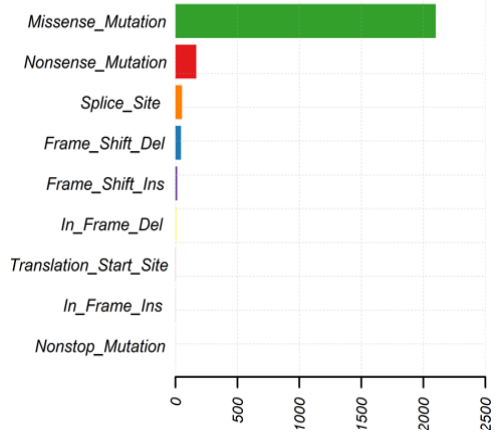
4.1.1.1. Mutational summary of immune related genes in 24 cancers

Mutation profile of immune-related genes in 24 different human cancers (Table 3.1) were analysed. SNPs (Figure 4.1.2) and missense mutation (Figure 4.1.1) were found to be the most prevalent type of mutation. The C>T transition change was more frequent. Cancers like KICH, KIRC, KIRP, LUSC however harboured higher number of C>A transversion (Figure 4.1.3).

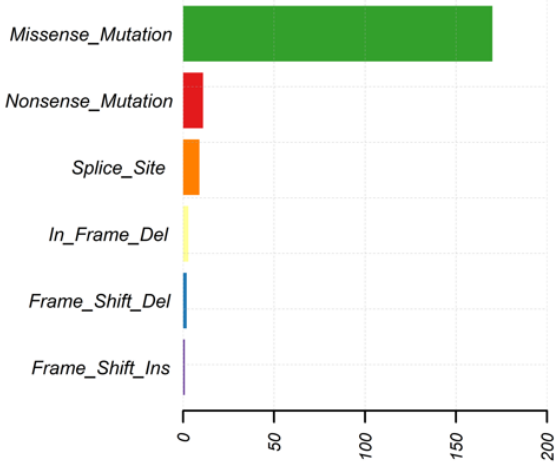




BRCA

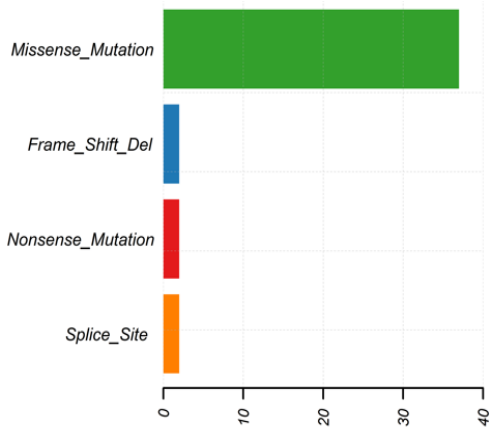


CESC

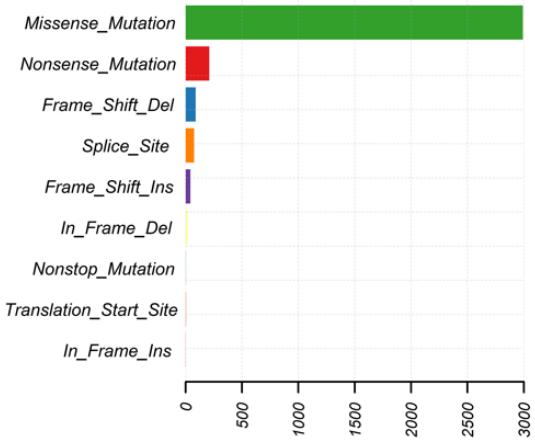


CHOL

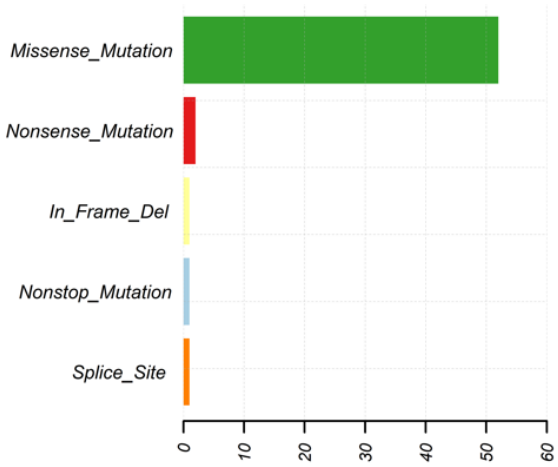
COAD



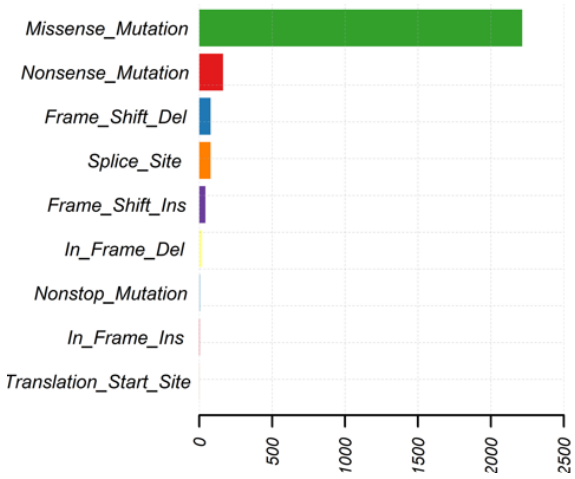
HCC



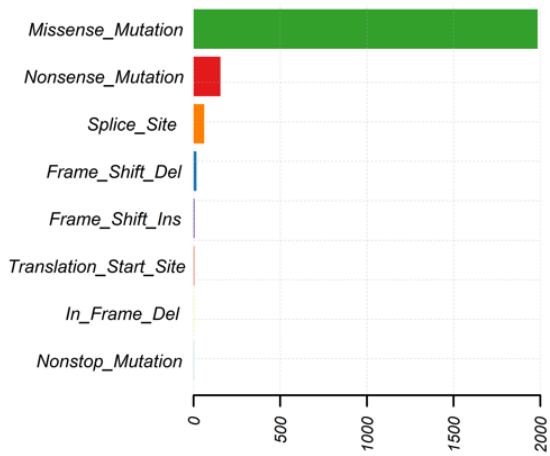
HNSC



KIRP

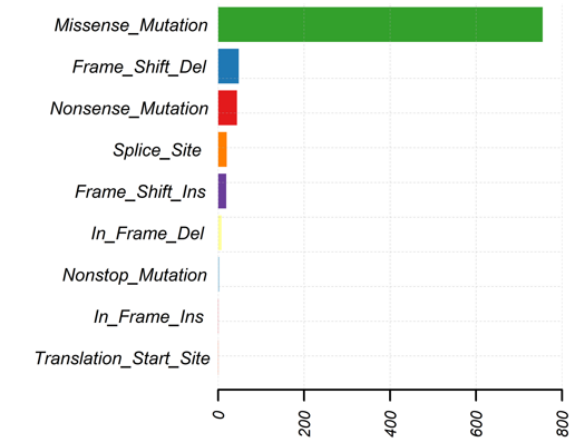


LUAD

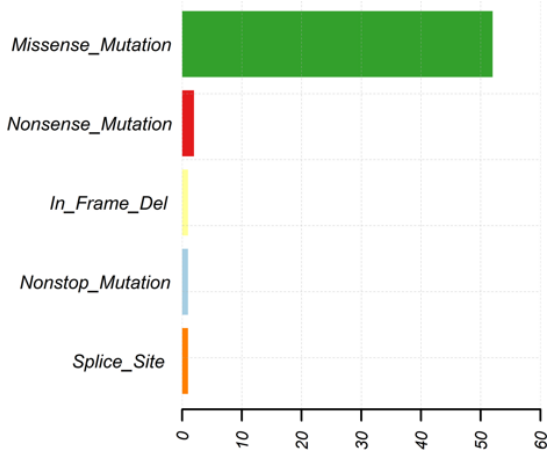


LUSC

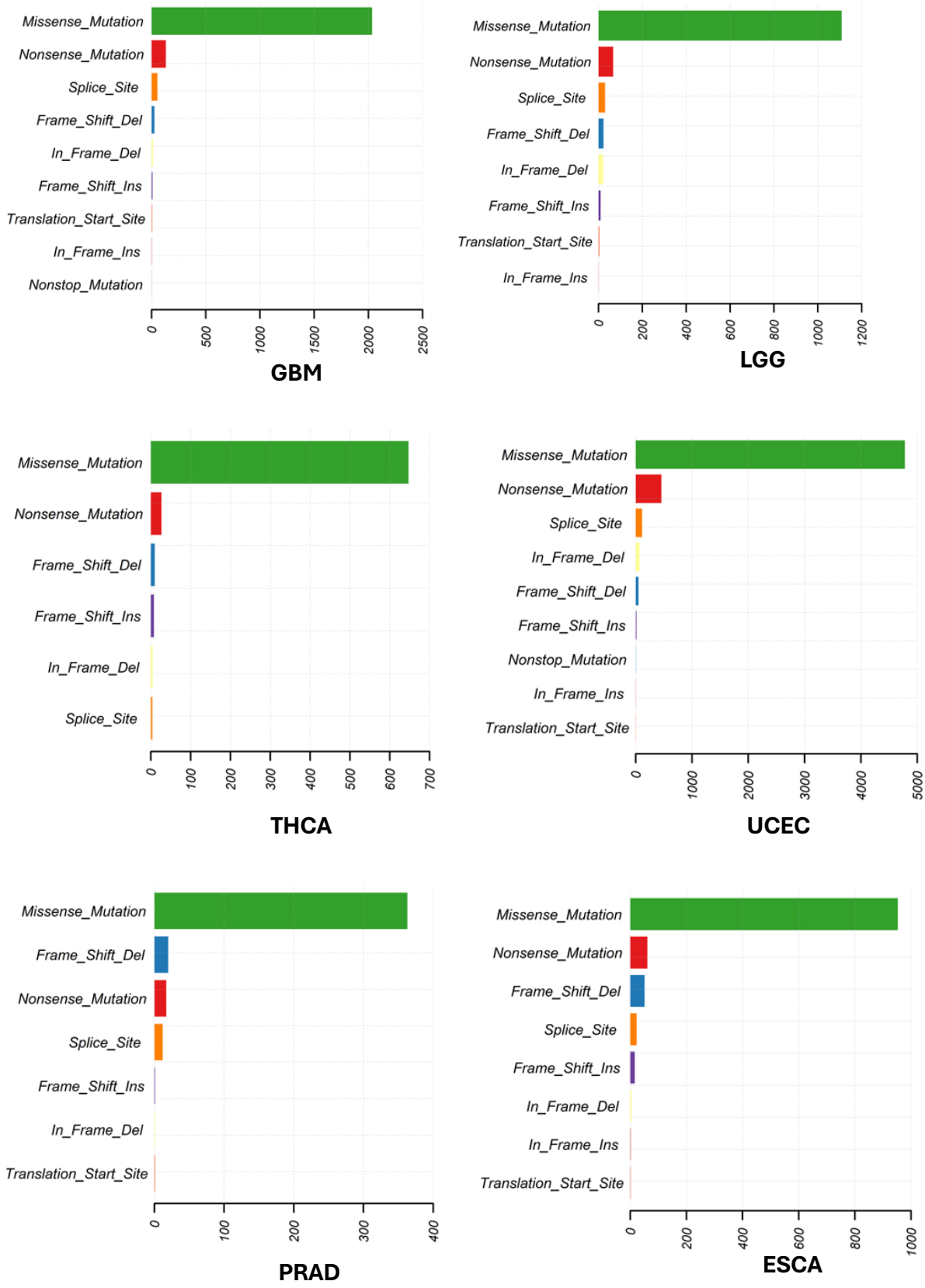
OV



HCC



HNSC


Missense_Mutation
Nonsense_Mutation
Splice_Site
Frame_Shift_Del
In_Frame_Del
Frame_Shift_Ins
Translation_Start_Site
In_Frame_Ins
Nonstop_Mutation
0
500
1000
1500
2000
2500
GBM
Missense_Mutation
Nonsense_Mutation
Splice_Site
Frame_Shift_Del
In_Frame_Del
Frame_Shift_Ins
Translation_Start_Site
In_Frame_Ins
0
200
400
600
800
1000
1200
LGG
Missense_Mutation
Nonsense_Mutation
Frame_Shift_Del
Frame_Shift_Ins
In_Frame_Del
Splice_Site
0
100
200
300
400
500
600
700
THCA
Missense_Mutation
Nonsense_Mutation
Splice_Site
In_Frame_Del
Frame_Shift_Del
Frame_Shift_Ins
Nonstop_Mutation
In_Frame_Ins
Translation_Start_Site
0
1000
2000
3000
4000
5000
UCEC
Missense_Mutation
Frame_Shift_Del
Nonsense_Mutation
Splice_Site
Frame_Shift_Ins
In_Frame_Del
Translation_Start_Site
0
100
200
300
400
PRAD
Missense_Mutation
Nonsense_Mutation
Frame_Shift_Del
Splice_Site
Frame_Shift_Ins
In_Frame_Del
In_Frame_Ins
Translation_Start_Site
0
200
400
600
800
1000
ESCA

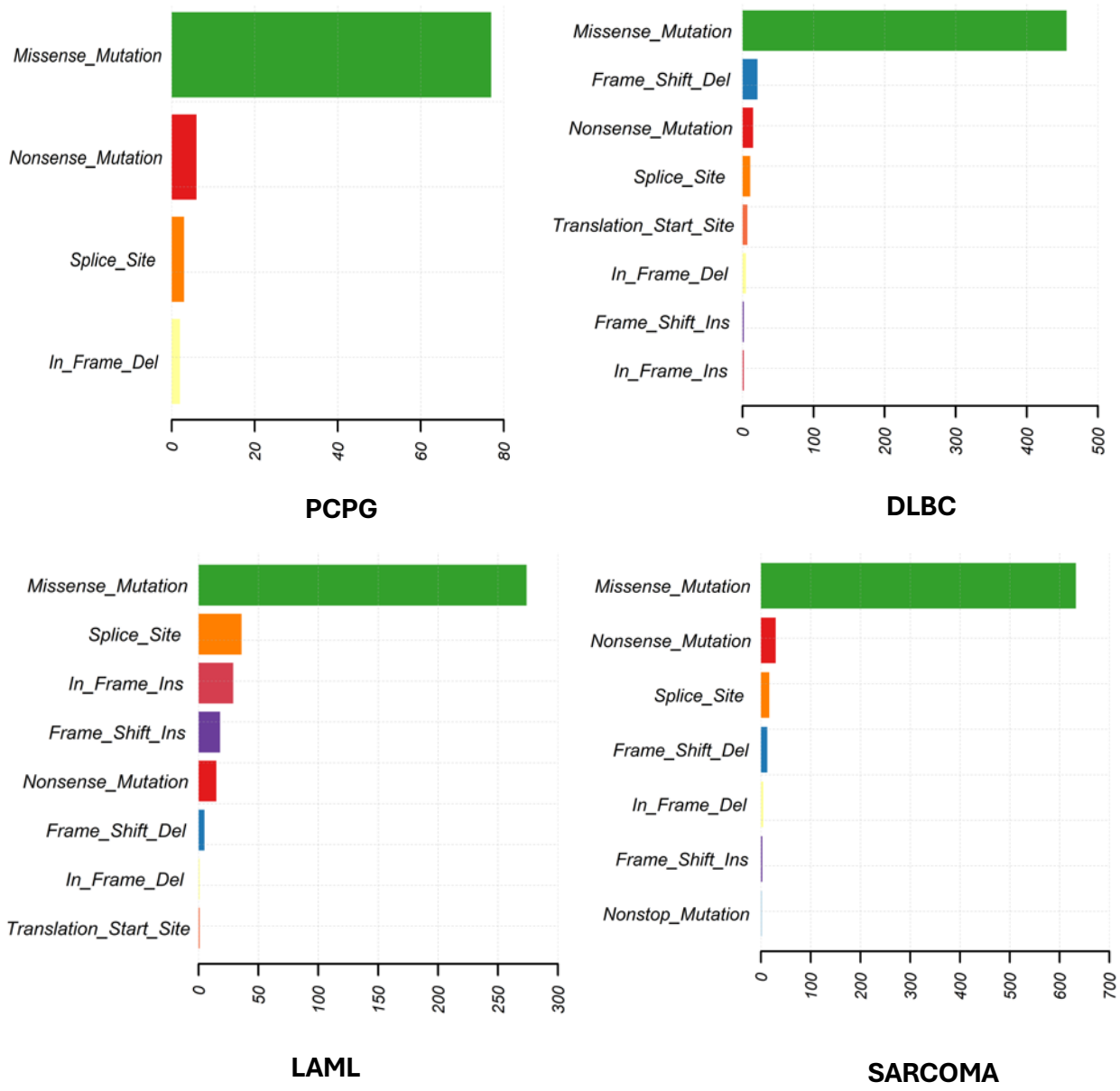
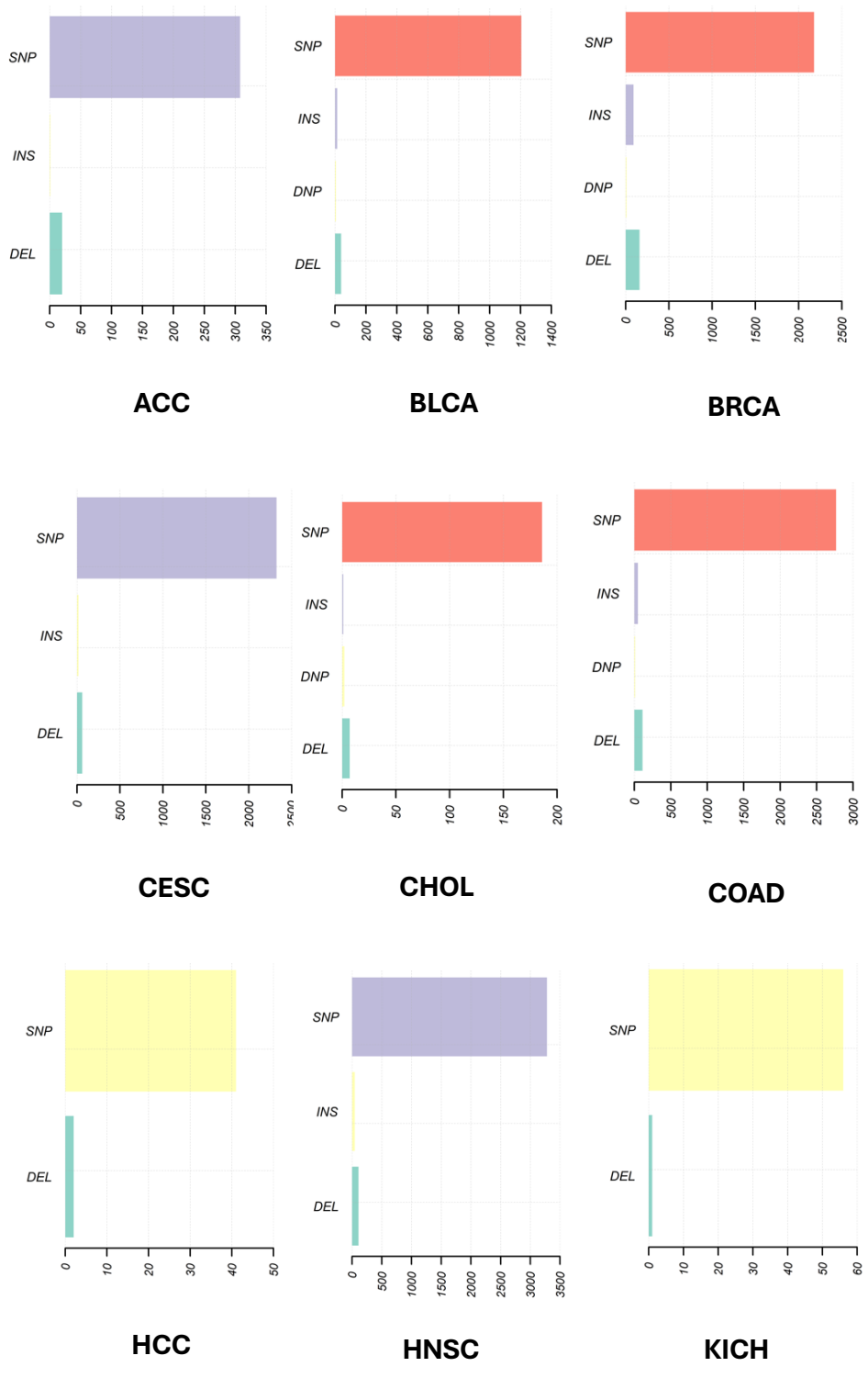
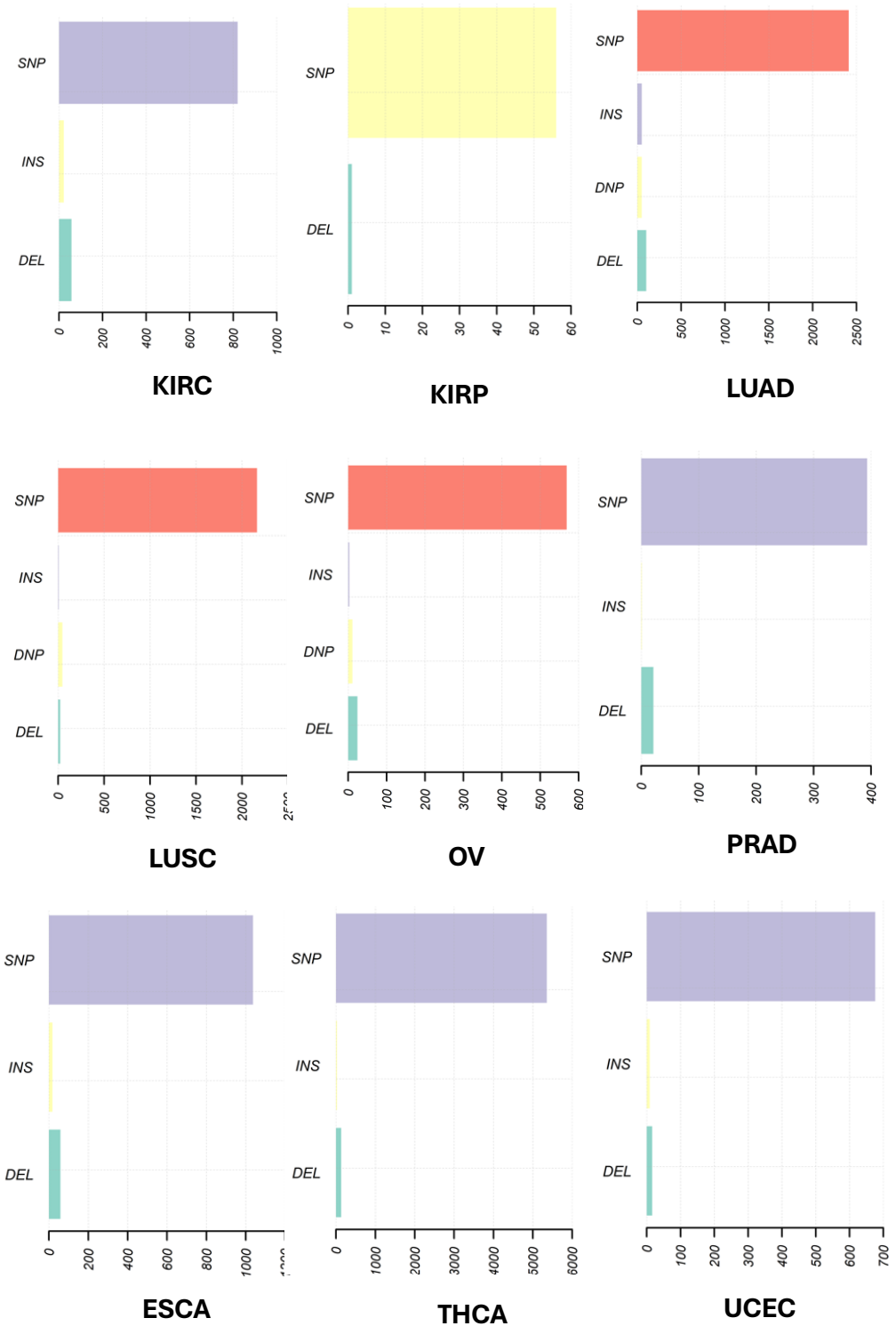


Figure 4.1.1: Bar graphs depicting the number of samples with type of variations in the 24 cancers. Y-axis represents type of variations while X-axis indicates the number of samples.


SNP
INS
DEL
0 50 100 150 200 250 300 350
ACC
SNP
INS
DNP
DEL
0 200 400 600 800 1000 1200 1400
BLCA
SNP
INS
DNP
DEL
0 500 1000 1500 2000 2500
BRCA
SNP
INS
DEL
0 500 1000 1500 2000 2500
CESC
SNP
INS
DNP
DEL
0 50 100 150 200
CHOL
SNP
INS
DNP
DEL
0 500 1000 1500 2000 2500 3000
COAD
SNP
DEL
0 10 20 30 40 50
HCC
SNP
INS
DEL
0 500 1000 1500 2000 2500 3000 3500
HNSC
SNP
DEL
0 10 20 30 40 50 60
KICH


SNP
INS
DEL
0 200 400 600 800 1000
KIRC
SNP
DEL
0 10 20 30 40 50 60
KIRP
SNP
INS
DNP
DEL
0 500 1000 1500 2000 2500
LUAD
SNP
INS
DNP
DEL
0 500 1000 1500 2000
LUSC
SNP
INS
DNP
DEL
0 100 200 300 400 500 600
OV
SNP
INS
DEL
0 100 200 300 400
PRAD
SNP
INS
DEL
0 200 400 600 800 1000
ESCA
SNP
INS
DEL
0 1000 2000 3000 4000 5000 6000
THCA
SNP
INS
DEL
0 100 200 300 400 500 600 700
UCEC

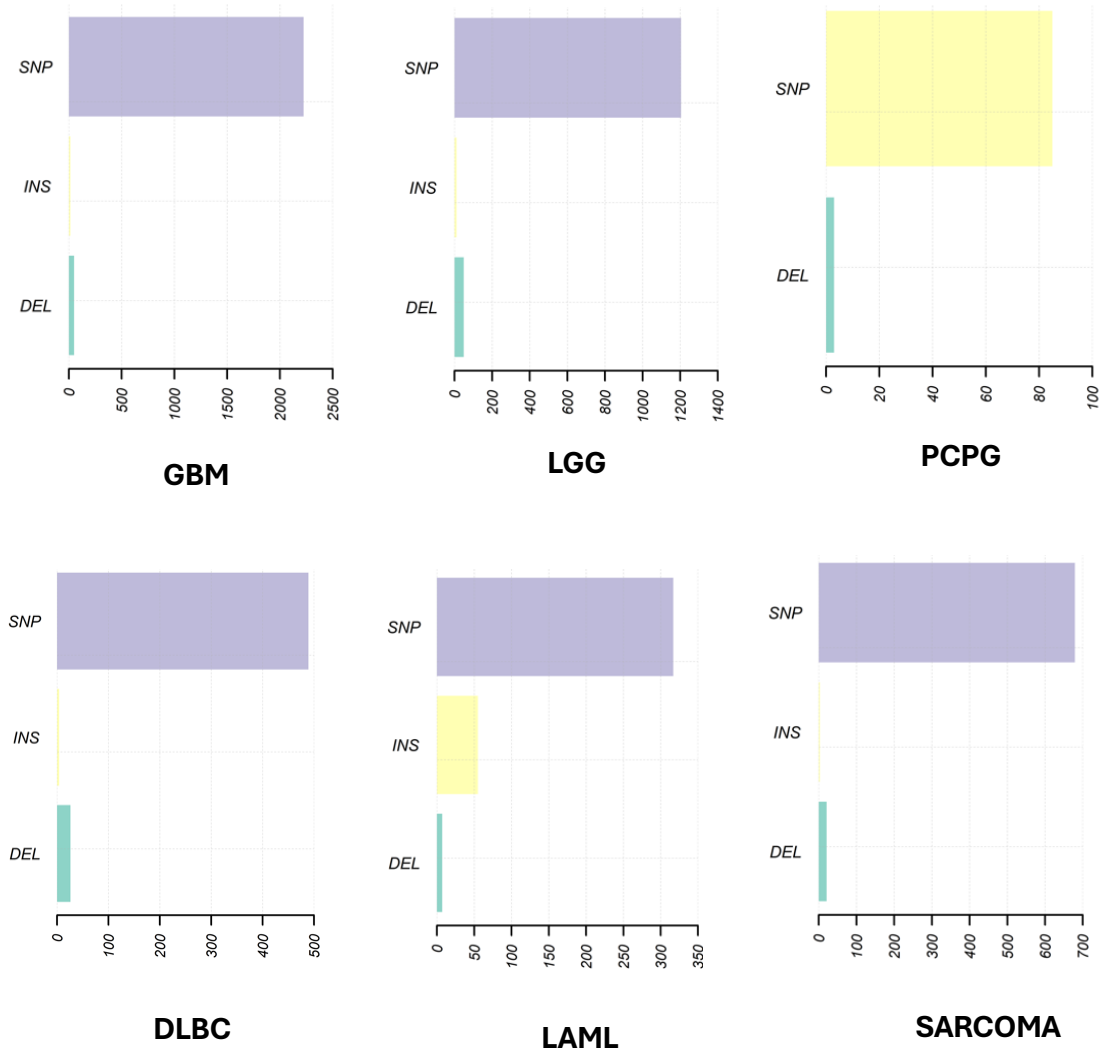
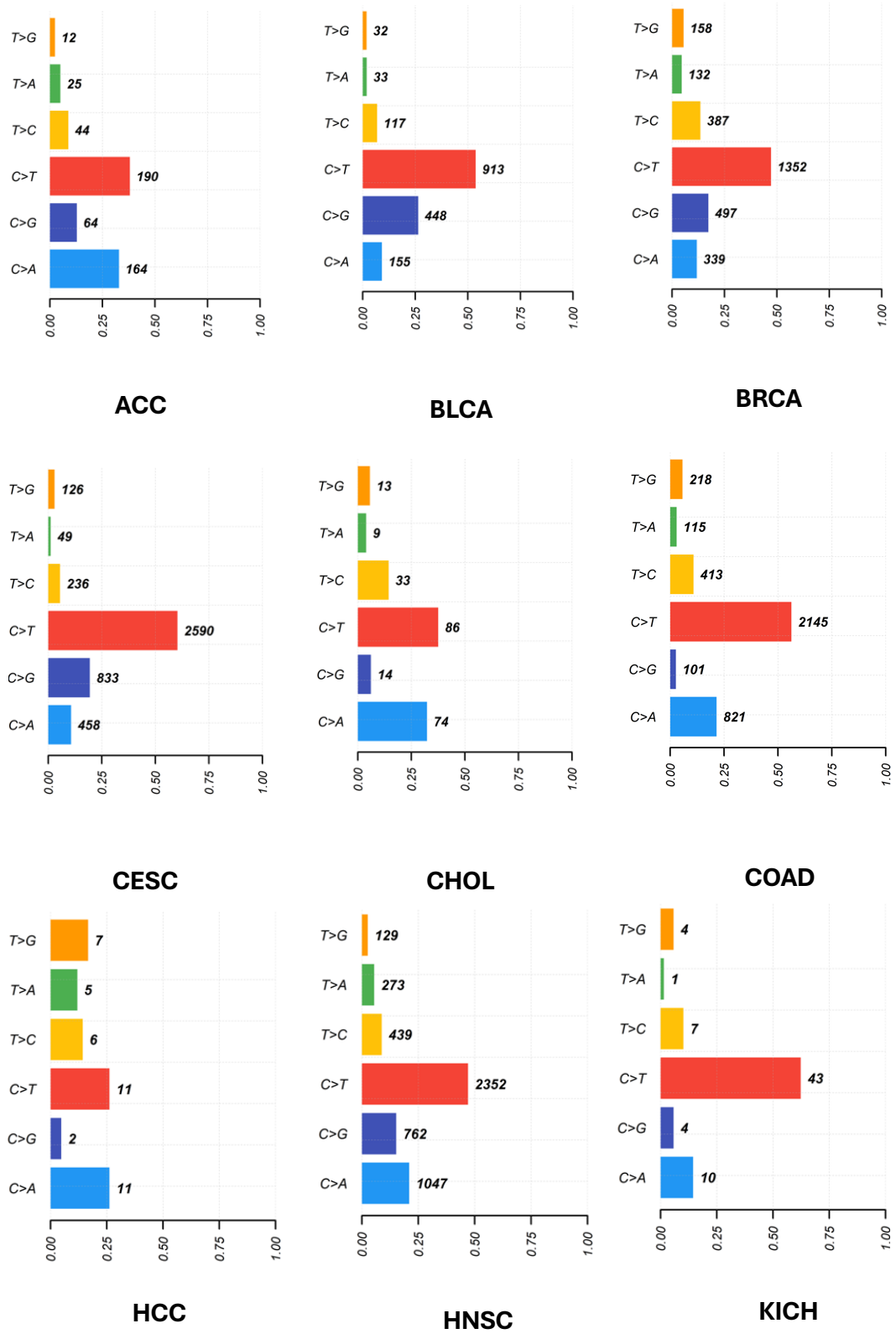
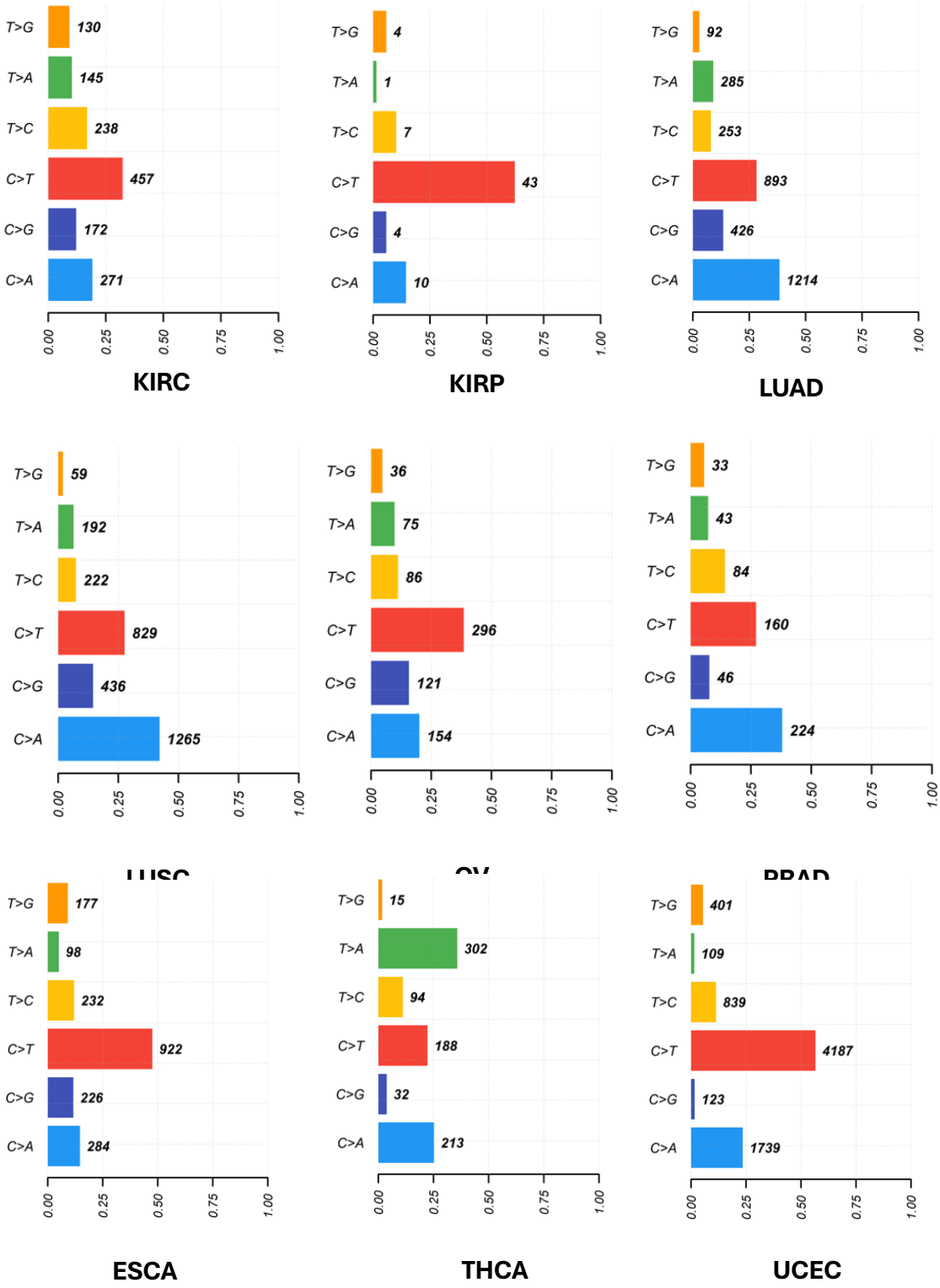


Figure 4.1.2: Bar graphs depicting the number of mutated samples with polymorphisms, insertions and deletions in the 24 cancers. The Y-axis represents type of mutation while X-axis indicates the number of samples.


T>G 12
T>A 25
T>C 44
C>T 190
C>G 64
C>A 164
0.00 0.25 0.50 0.75 1.00
ACC
T>G 32
T>A 33
T>C 117
C>T 913
C>G 448
C>A 155
0.00 0.25 0.50 0.75 1.00
BLCA
T>G 158
T>A 132
T>C 387
C>T 1352
C>G 497
C>A 339
0.00 0.25 0.50 0.75 1.00
BRCA
T>G 126
T>A 49
T>C 236
C>T 2590
C>G 833
C>A 458
0.00 0.25 0.50 0.75 1.00
CESC
T>G 13
T>A 9
T>C 33
C>T 86
C>G 14
C>A 74
0.00 0.25 0.50 0.75 1.00
CHOL
T>G 218
T>A 115
T>C 413
C>T 2145
C>G 101
C>A 821
0.00 0.25 0.50 0.75 1.00
COAD
T>G 7
T>A 5
T>C 6
C>T 11
C>G 2
C>A 11
0.00 0.25 0.50 0.75 1.00
HCC
T>G 129
T>A 273
T>C 439
C>T 2352
C>G 762
C>A 1047
0.00 0.25 0.50 0.75 1.00
HNSC
T>G 4
T>A 1
T>C 7
C>T 43
C>G 4
C>A 10
0.00 0.25 0.50 0.75 1.00
KICH


T>G 130
T>A 145
T>C 238
C>T 457
C>G 172
C>A 271
KIRC
T>G 4
T>A 1
T>C 7
C>T 43
C>G 4
C>A 10
KIRP
T>G 92
T>A 285
T>C 253
C>T 893
C>G 426
C>A 1214
LUAD
T>G 59
T>A 192
T>C 222
C>T 829
C>G 436
C>A 1265
T>G 36
T>A 75
T>C 86
C>T 296
C>G 121
C>A 154
T>G 33
T>A 43
T>C 84
C>T 160
C>G 46
C>A 224
T>G 177
T>A 98
T>C 232
C>T 922
C>G 226
C>A 284
ESCA
T>G 15
T>A 302
T>C 94
C>T 188
C>G 32
C>A 213
THCA
T>G 401
T>A 109
T>C 839
C>T 4187
C>G 123
C>A 1739
UCEC
0.00 0.25 0.50 0.75 1.00

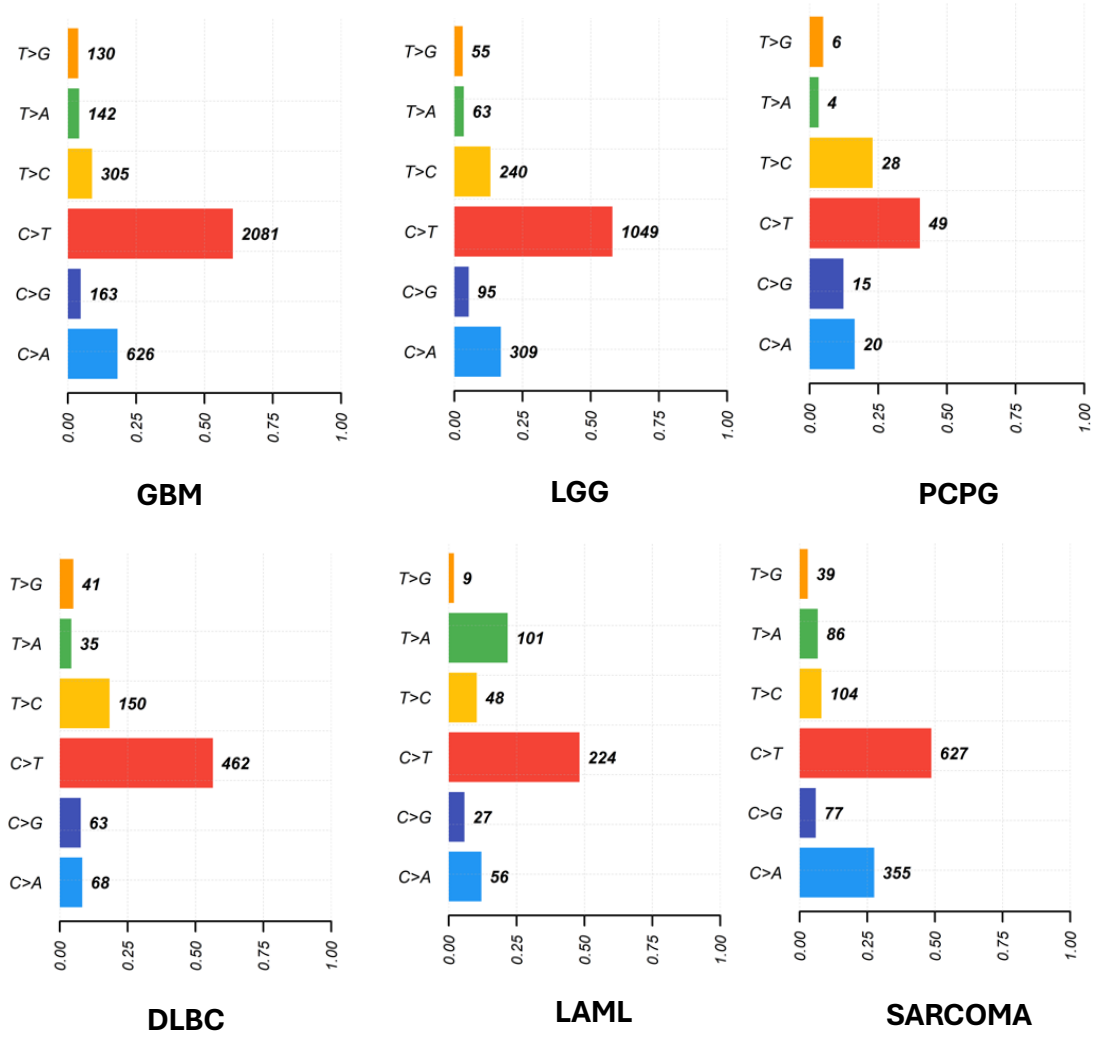
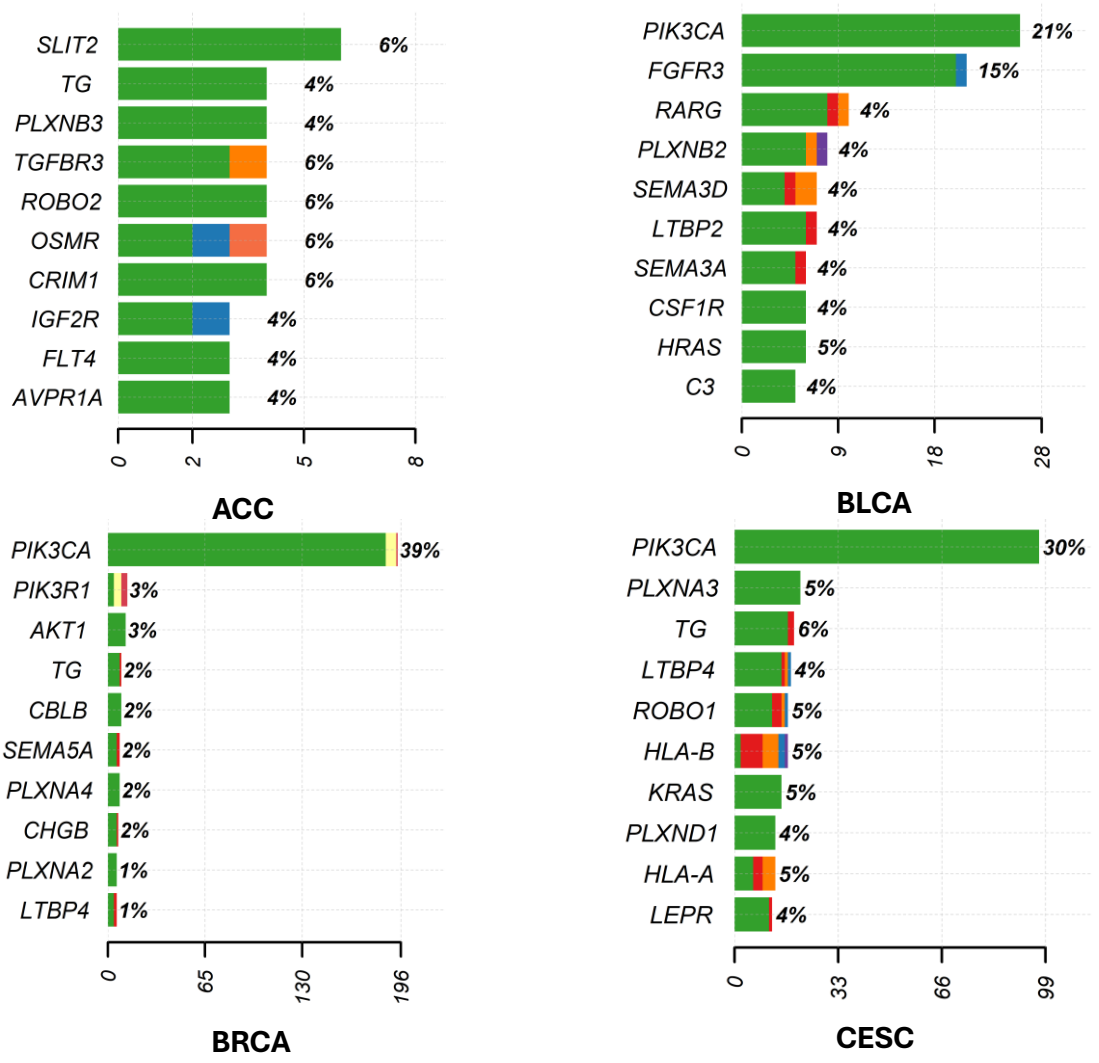
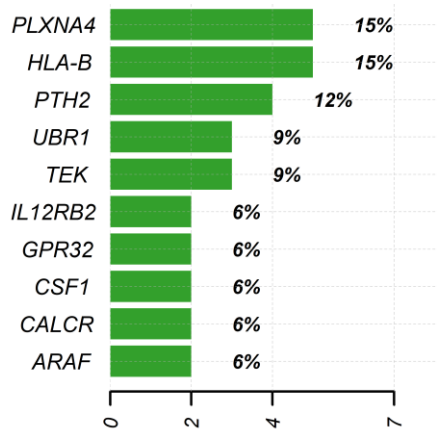


Figure 4.1.3: Bar graphs depicting the type of nucleotide changes in the 24 cancers. The Y-axis represents type of nucleotide changes while X-axis indicates the ratio of mutated samples.

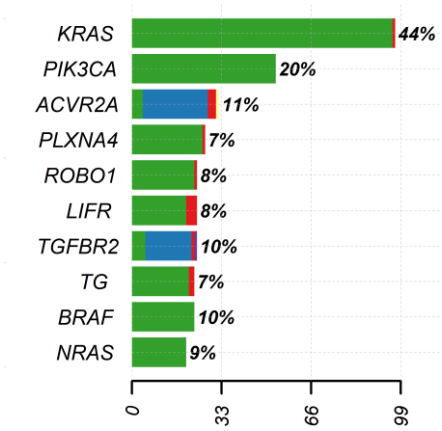
4.1.1.2. Frequently mutated immune-related genes in 24 different cancers

The 24 different cancers (Table 3.1) were examined for frequently mutated immune-related genes. Of the highly mutated genes in each of the studied cancers, TG in 15/24 and PIK3CA in 14/24 cancers respectively was featured among the top 10 highly mutated immune-related genes making them most frequently mutated immune-related genes (Figure 4.1.2). Other than the two genes we obtained a list of 17 more genes that was represented in atleast two of the studied cancers. A total of 19 genes was obtained, the genes could be categorised into three broad categories considering their biological function (Table 4.1.1).

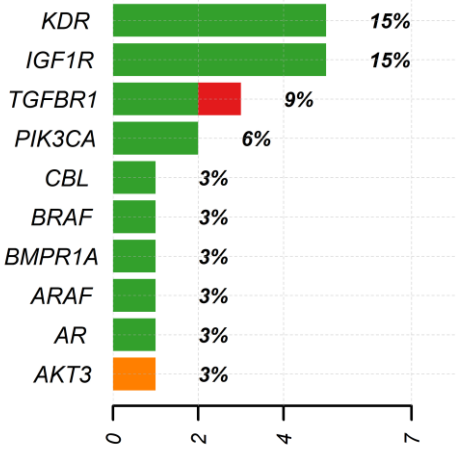




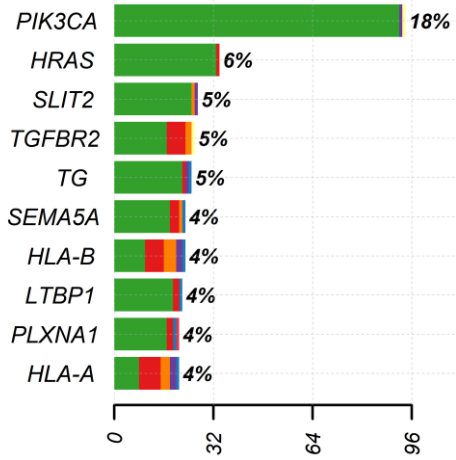
CHOL



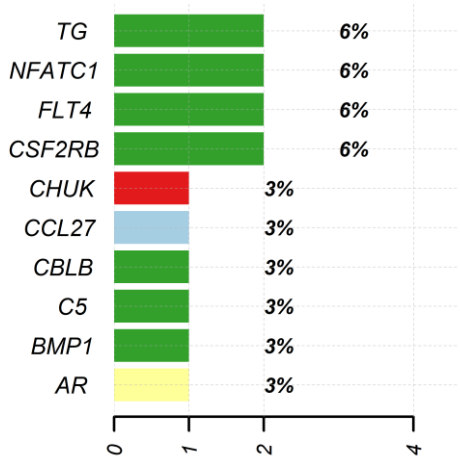
COAD



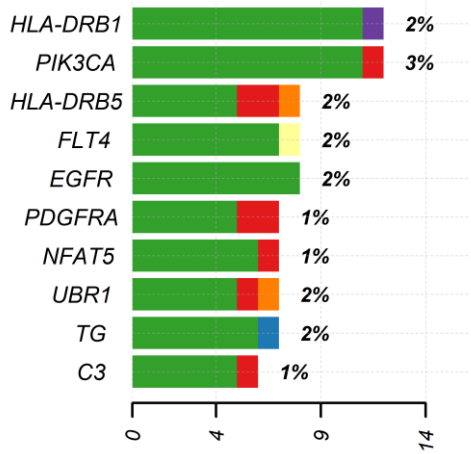
HCC



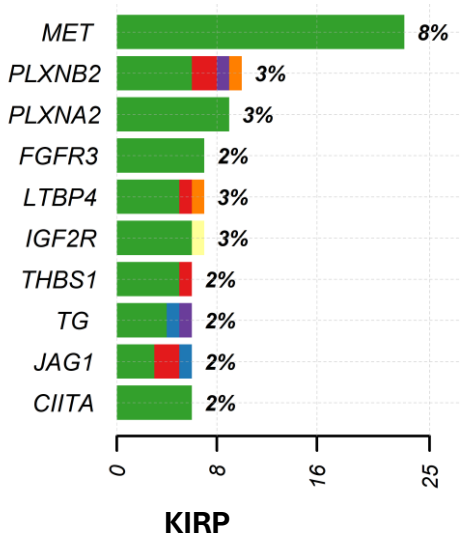
HNSC

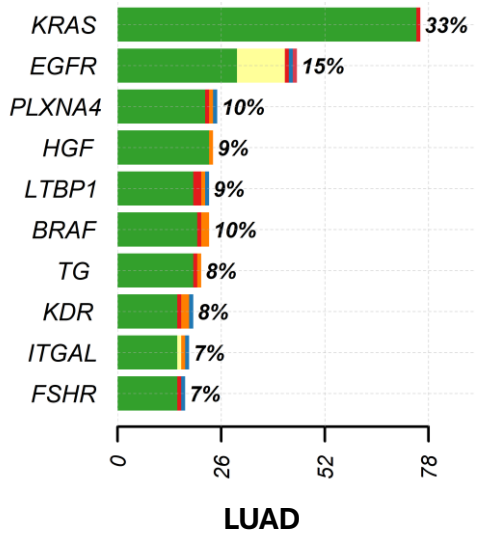


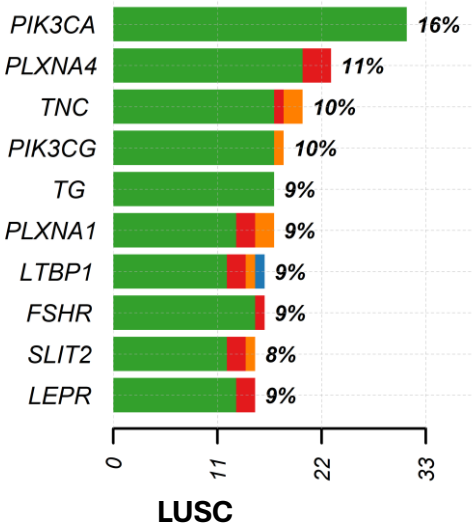
KICH

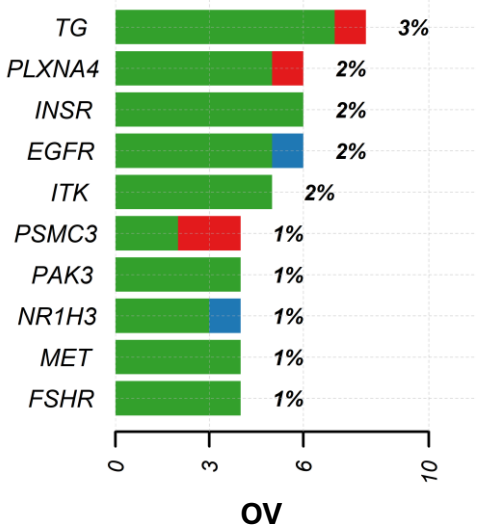


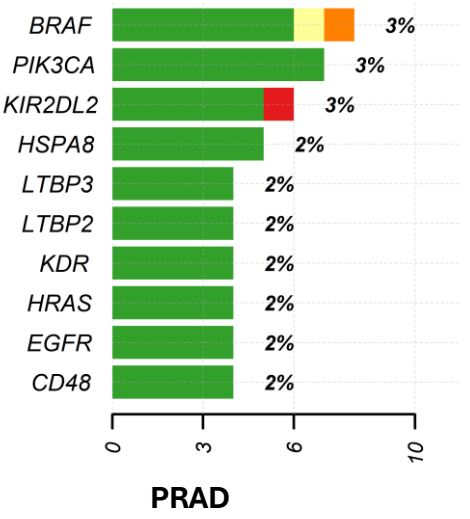
KIRC

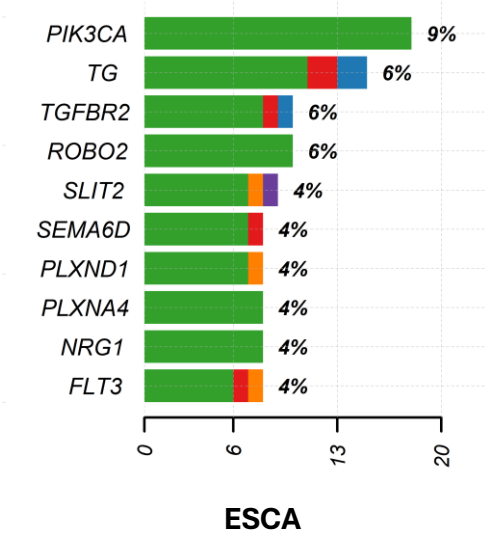

MET 8%
PLXNB2 3%
PLXNA2 3%
FGFR3 2%
LTBP4 3%
IGF2R 3%
THBS1 2%
TG 2%
JAG1 2%
CIITA 2%
0 8 16 25
KIRP

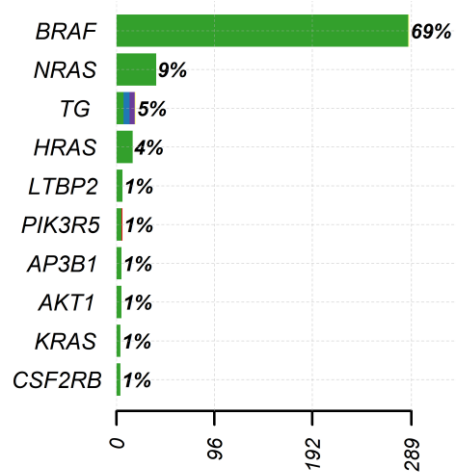

KRAS 33%
EGFR 15%
PLXNA4 10%
HGF 9%
LTBP1 9%
BRAF 10%
TG 8%
KDR 8%
ITGAL 7%
FSHR 7%
0 26 52 78
LUAD


PIK3CA 16%
PLXNA4 11%
TNC 10%
PIK3CG 10%
TG 9%
PLXNA1 9%
LTBP1 9%
FSHR 9%
SLIT2 8%
LEPR 9%
0 11 22 33
LUSC

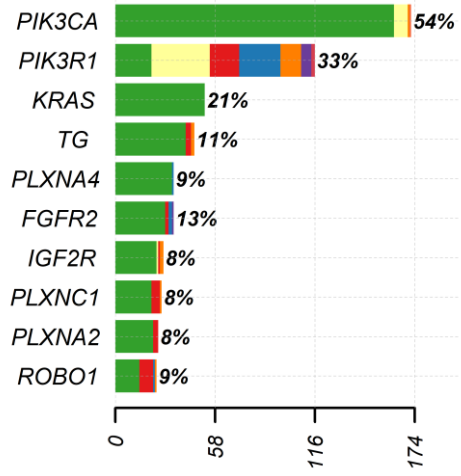

TG 3%
PLXNA4 2%
INSR 2%
EGFR 2%
ITK 2%
PSMC3 1%
PAK3 1%
NR1H3 1%
MET 1%
FSHR 1%
0 3 6 10
OV


BRAF 3%
PIK3CA 3%
KIR2DL2 3%
HSPA8 2%
LTBP3 2%
LTBP2 2%
KDR 2%
HRAS 2%
EGFR 2%
CD48 2%
0 3 6 10
PRAD

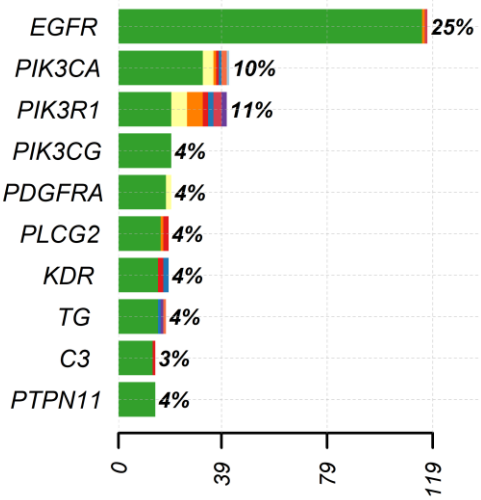

PIK3CA 9%
TG 6%
TGFBR2 6%
ROBO2 6%
SLIT2 4%
SEMA6D 4%
PLXND1 4%
PLXNA4 4%
NRG1 4%
FLT3 4%
0 6 13 20
ESCA



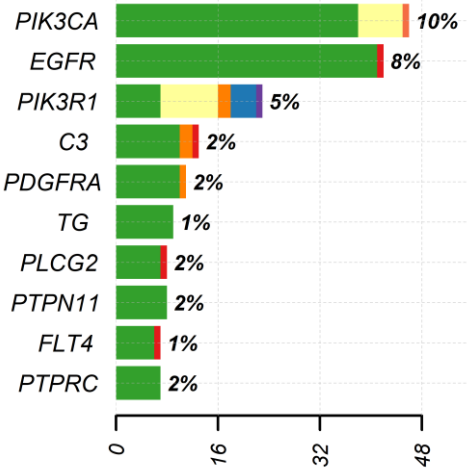
THCA



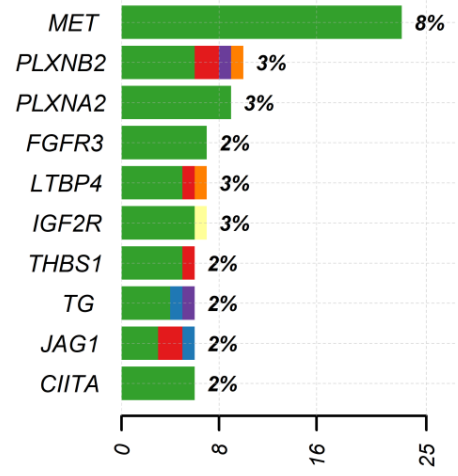
UCEC



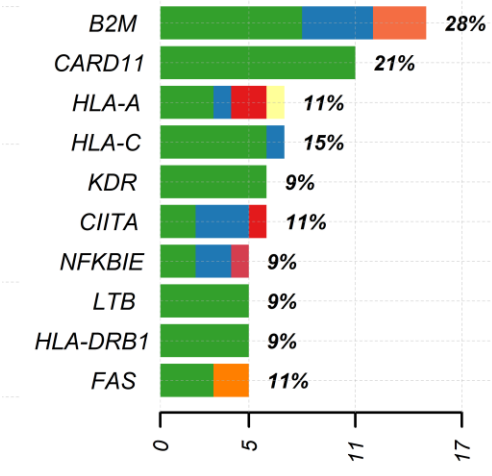
GBM



LGG



PCPG



DLBC

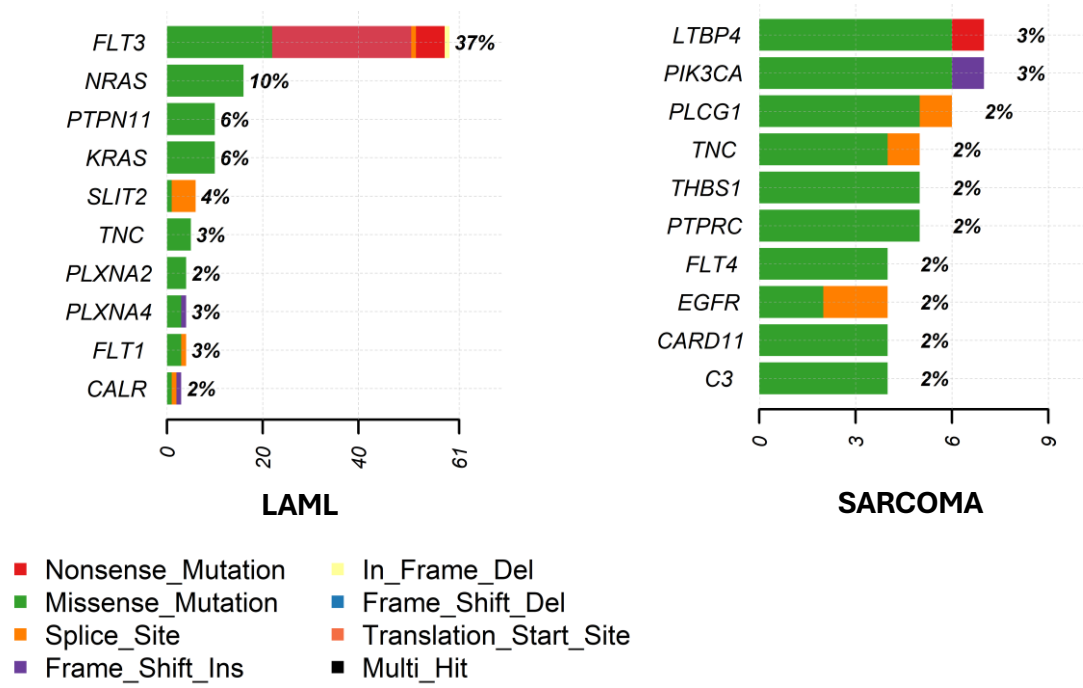


Figure 4.1.4: Top 10 frequently mutated immune-related genes in the 24 different cancers; Y-axis shows gene symbols and X-axis represents the number of samples with mutation in the gene. Each colour in the stacked bars represent different type of mutations. The percent of samples mutated for that gene in the respective cancer is depicted at the end of each bar in the chart.

Cell regulation	**Cell migration**	**Antigen presentation**
IGF2R (3-4)	PLXNA4(2-11)	HLA-DRB1(2-9)
TGFBR2(5-9)	PIK3CA (3-54)	HLA-A (5-11)
LTBP4(3-4)	SLIT2(4-6)	HLA-B (4-5)
EGFR (2-25)	ROBO1(5-8)	CIITA (2)
TG (2-11)	ROBO2(6-9)	
LTBP2(1-2)	PLXNB2(3)	
LTBP1(4-9)	PIK3R1(11-33)	

Table 4.1.1: List of frequently mutated immune related genes divided into three categories according to biological function. The numbers within the brackets represents the range of percentage of samples mutated for the gene in various cancers

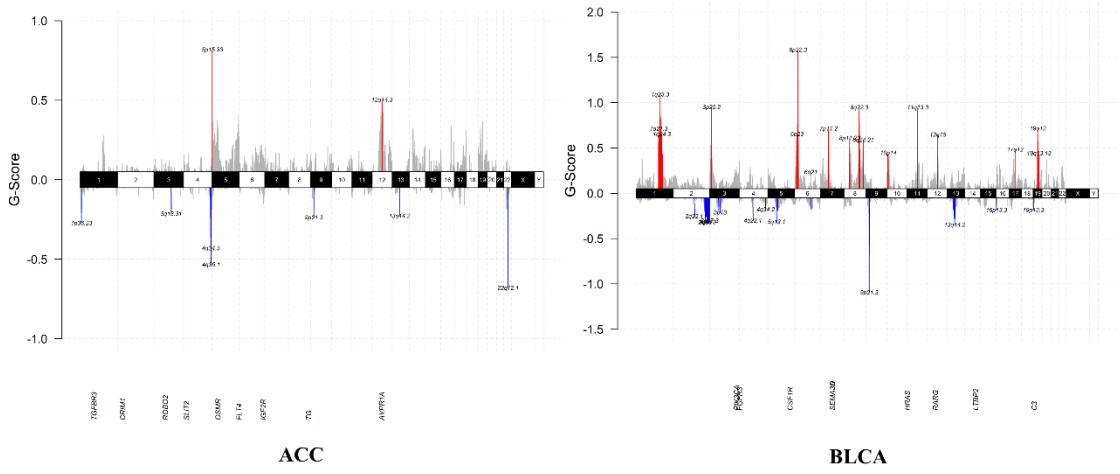
4.1.1.3. CNA in frequently mutated immune related genes

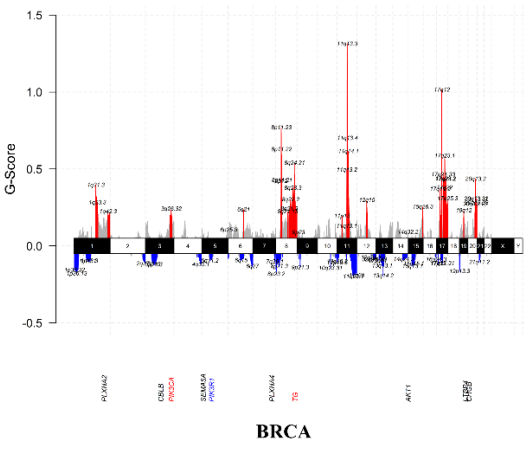
We screened the 24 cancers (Table 3.1) for copy number alterations (CNAs) and mapped them to frequently altered immune-related genes in each cancer. Notably, in this study, PIK3CA and TG genes were highlighted with amplifications in 8 and 7 cancers out of 24 types in the study (Table 4.1.2) (Supplementary Figure1).

Sl. No.	Abbreviation	Amplified	Deleted
1	SARCOMA	PIK3CA	PIK3R1
2	ACC		
3	BLCA	EGFR3	
4	BRCA	PIK3CA, TG	PIK3R1
5	CHOL		
6	COAD	NRAS, TG	
7	CESC	PIK3CA	
8	HCC		
9	HNSC	PIK3CA, TG	
10	KICH		
11	KIRC	PIK3CA, FLT1	
12	KIRP		MET, PLXNA4
13	LUAD	EGFR, TG, KRAS	
14	LUSC		
15	OV	TG	
16	PRAD		
17	ESCA	PIK3CA, TG	ROBO2
18	THCA		
19	UCEC		

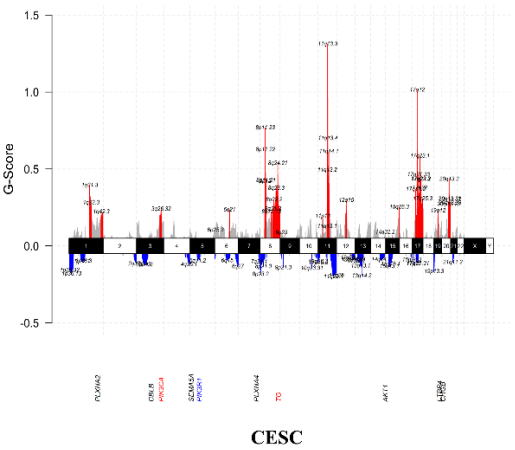
20	GBM	PIK3CA, EGFR, PDGFRA, KDR	
21	LGG	PIK3CA, EGFR, PDGFRA, TG	
22	PCPG		
23	DLBC		FAS, B2M
24	LAML		

Table 4.1.2: Immune-related genes with frequent mutations and CNA both in the particular cancer

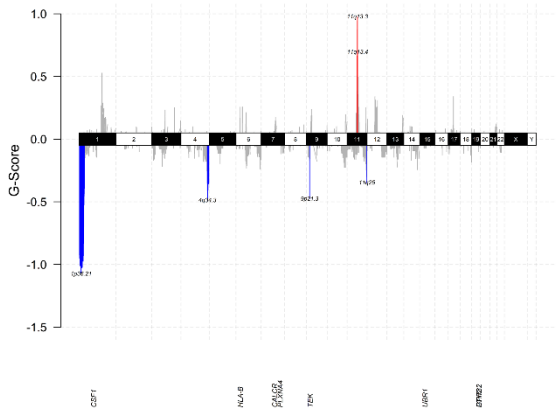




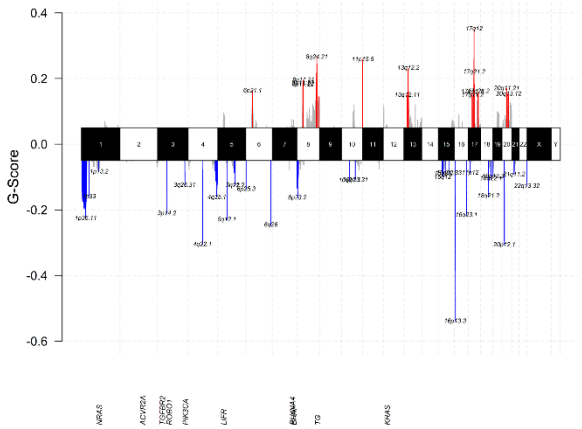
BRCA



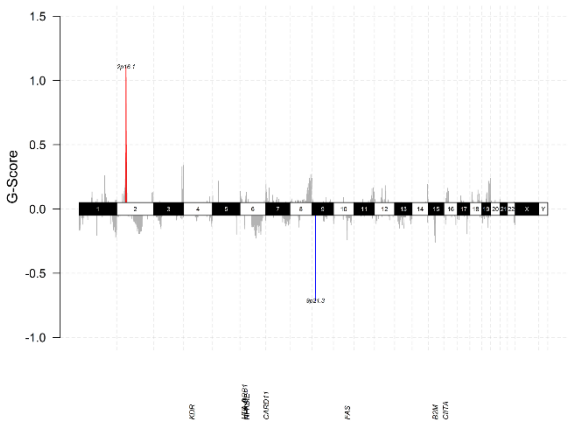
CESC



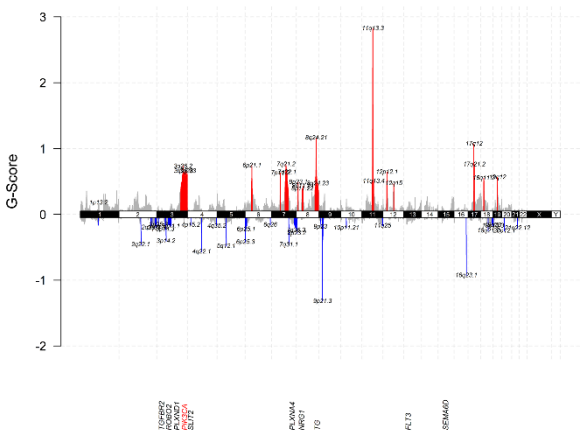
CHOL



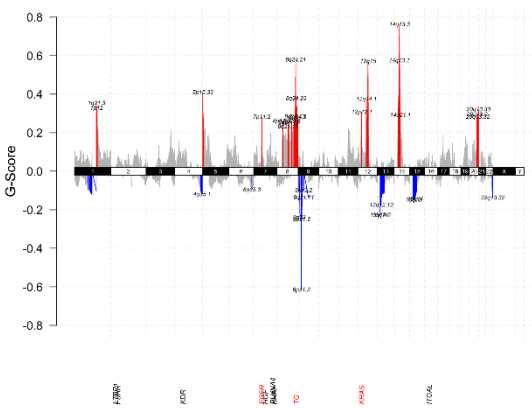
COAD



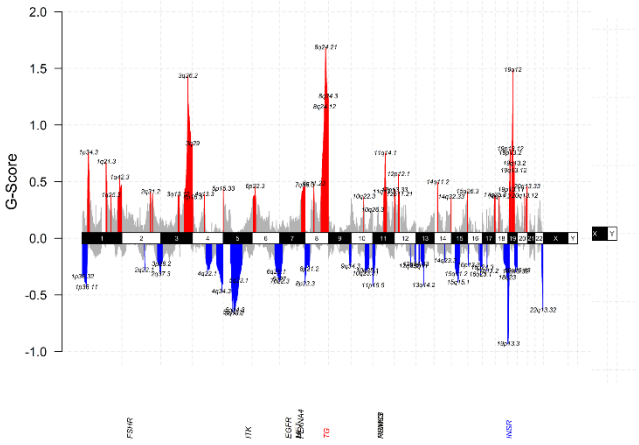
HCC



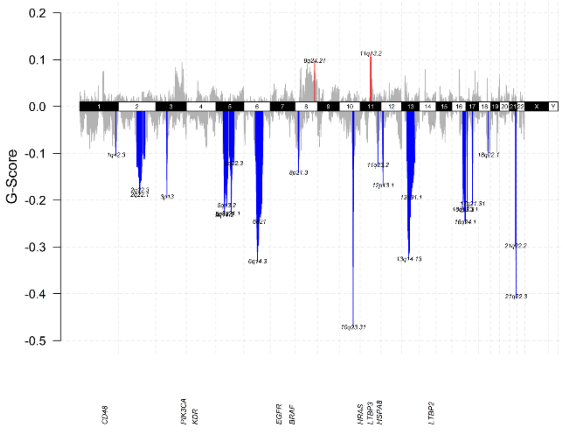
HNSC



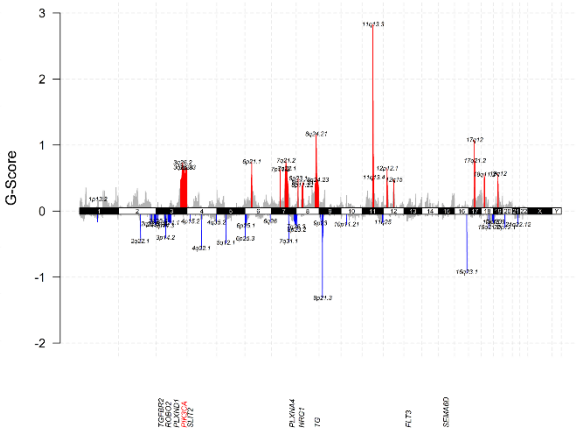
LUAD



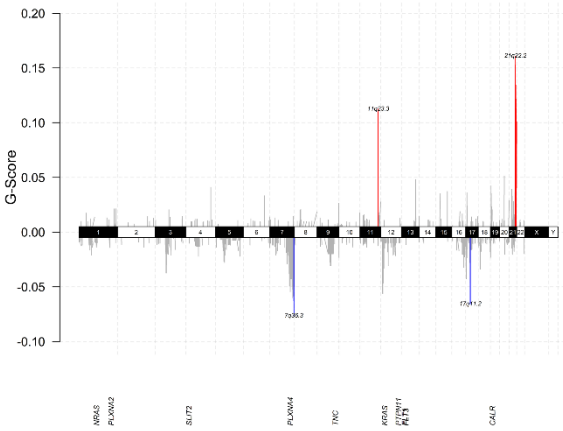
OV



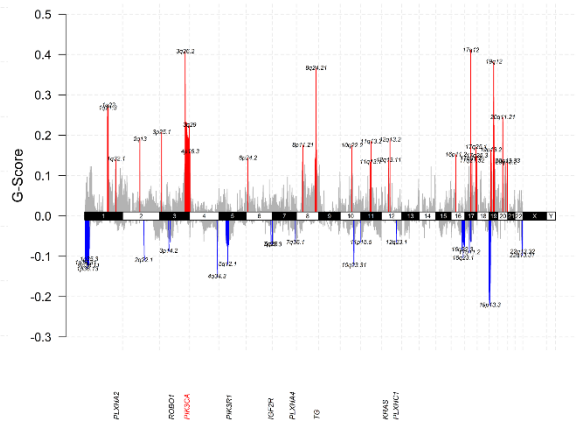
PRAD



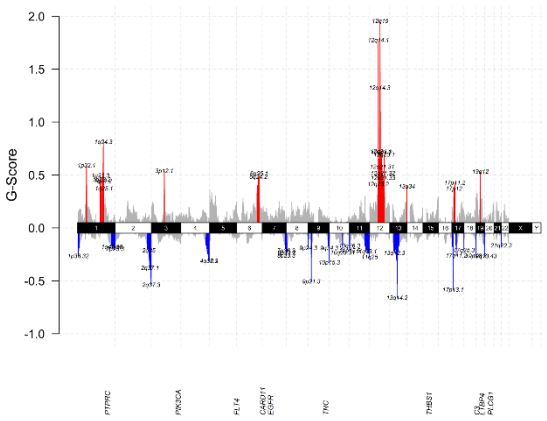
ESCA



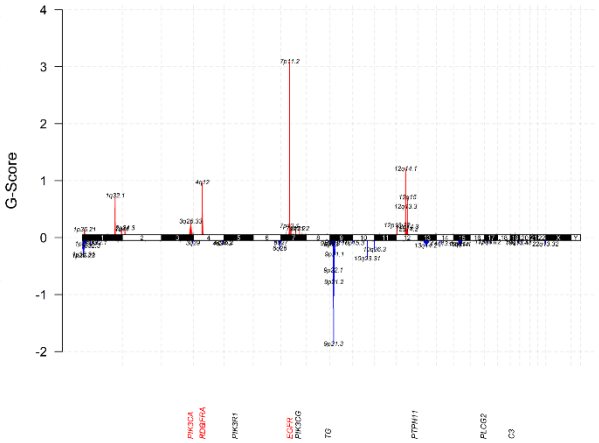
LAML



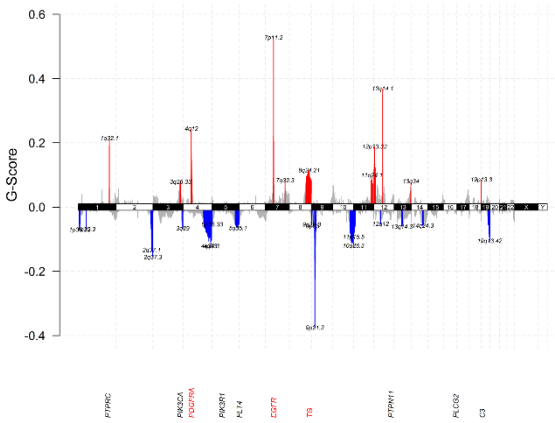
UCEC



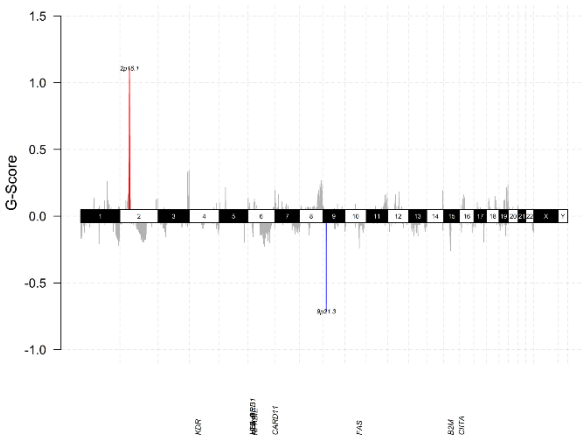
SARCOMA



GBM



LGG



DLBC

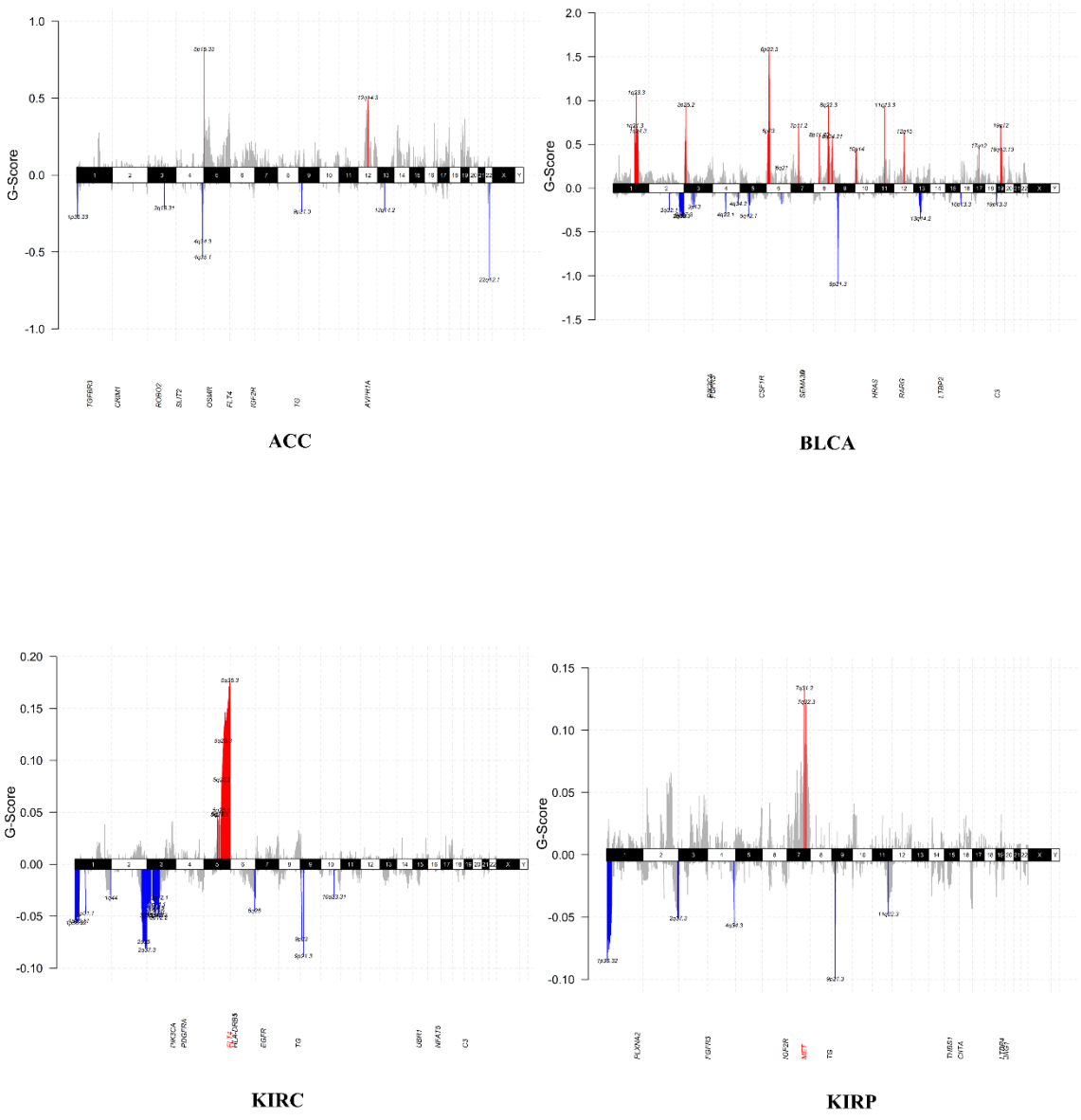


Figure 4.1.5: Copy number alterations in frequently mutated genes of the respective cancers. The peaks marked in red are amplifications and the ones in blue are deletions. Y axis gives us G-score(It refers to a statistical metric derived by the GISTIC algorithm that indicates the significance and frequency of a certain CNA over a group of samples,

taking into account both the magnitude of the copy number change and how frequently it occurs in the studied data) and X axis represent the chromosome numbers of human genome.

4.1.1.4. G-quadruplexes (G4) in immune-related genes

The mapping of both putative G4 locations and immune-related genes onto the human genome revealed a noteworthy overlap at several genomic locations (Figure 4.1.3). This observation underscores the abundance of G4 structures within genes associated with our immune system.

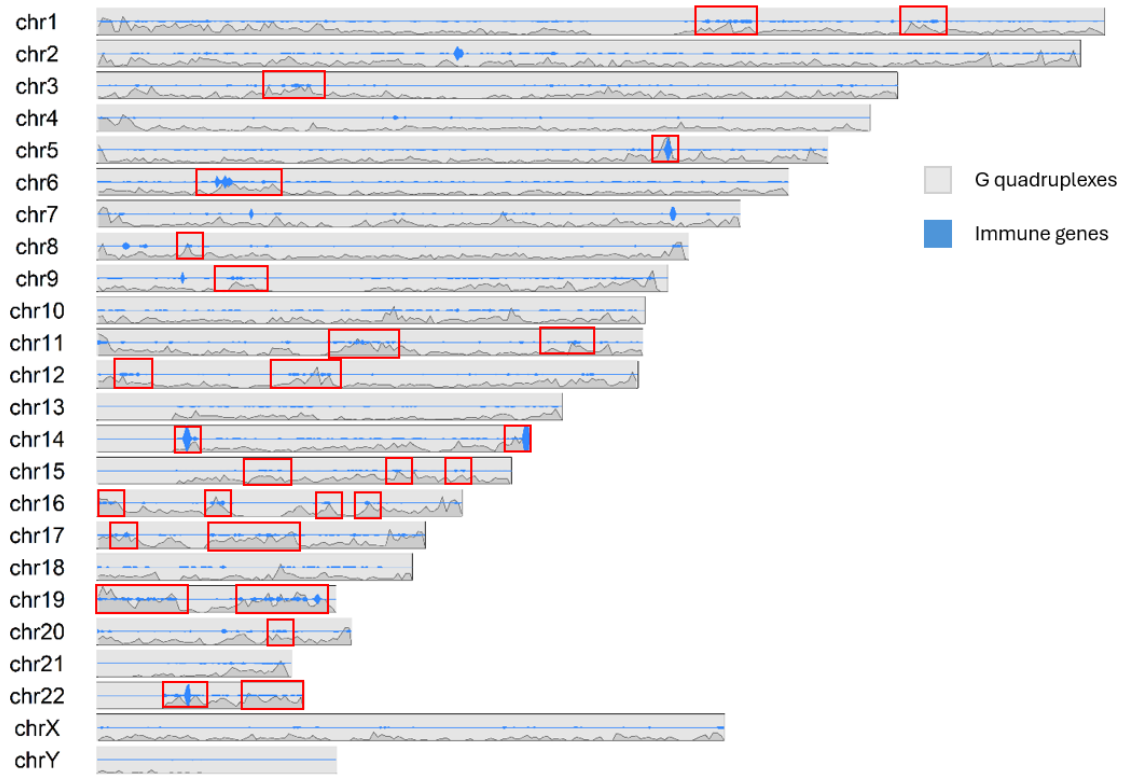


Figure 4.1.6: Putative G4 locations and immune related genes along the human genome. Red boxes in the diagram represents overlap between G4 locations and immune genes Chromosome numbers were mentioned on the left side of the diagram

4.1.1.5. G4 density in the gene and promoters of immune related genes, housekeeping genes and proto-oncogenes

G4 density along the gene and in promoters of proto-oncogenes were compared to immune related genes and housekeeping genes, as abundance of G4 structures in proto-oncogenes has been established. Housekeeping genes have a relatively simple regulation therefore it was expected that G4 density might be low in them as the structures were established to be involved in various regulatory functions. The density of G4 structures along the genes of immune-related genes was observed to be comparatively higher than in proto-oncogenes and housekeeping genes (Figure 4.1.6). In contrast, the promoters of proto-oncogenes and housekeeping genes harboured more G4 structures than those in immune-related genes (Figure 4.1.7). Here we used one-way ANOVA to rule out the possibility that the difference between the groups is not just by random coincidence and the Dunnett's test establish that the difference between proto-oncogenes and the other two groups namely- immune genes and housekeeping genes are statistically significant.

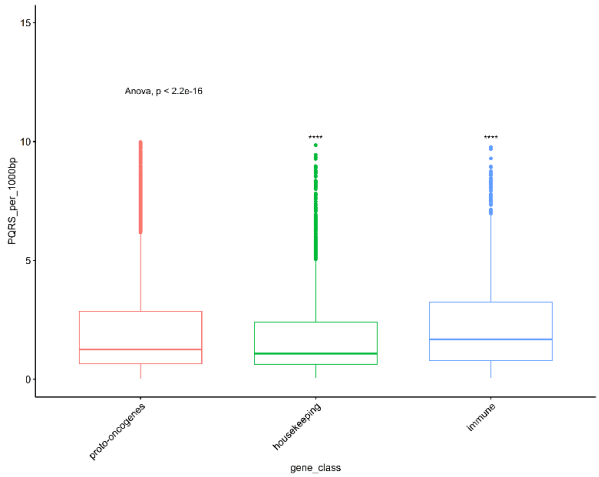


Figure 4.1.7: Box plot with ANOVA ($p<2.2e-16$) for comparing G4 density (G4/1000bp) in genes of immune system, housekeeping genes and proto-oncogenes. X-axis represents the gene classes and Y-axis represent putative G4 (PQRS) per 1000 base pairs along the gene. The statistical significance denoted are results from Dunnett's test.

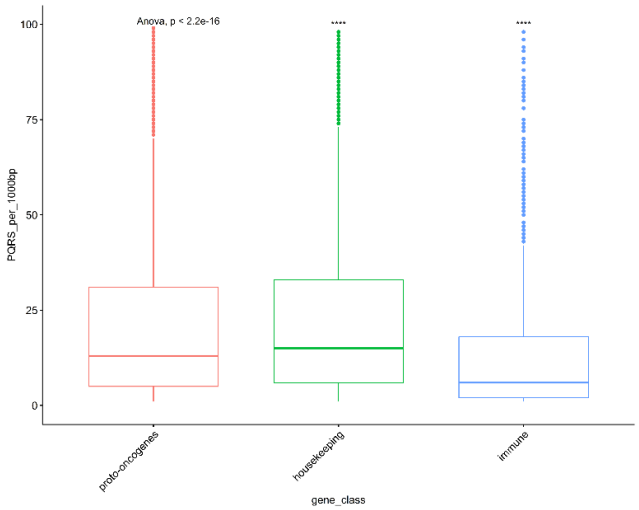


Figure 4.1.8: Box plot with ANOVA ($p<2.2e-16$) for comparing G4 density (G4/1000bp) in promoters of immune genes, housekeeping genes and proto-oncogenes. X-axis represents the gene classes and Y-axis represent putative G4 (PQRS) per 1000 base pairs along the gene. The statistical significance denoted are results from Dunnett's test.

4.1.1.6. G4 density in genes and promoters of different immune classes compared to proto-oncogenes

The density of putative G-quadruplexes (PQRS) along the genes and in the promoter region of proto-oncogenes was compared to different classes of immune related genes. The density of PQRS is significantly higher than proto-oncogenes along the length of genes involved in the T-cell receptor (TCR) and B-cell receptor (BCR) pathways (Figure 4.1.8). On the other hand, the density of PQRS was notably lower in the promoters of cytokines, TGF-β and TNF family genes, TCR, and BCR pathway genes compared to proto-oncogenes (Figure 4.1.9). We performed a one-way ANOVA to rule out the possibility that the differences between the groups occurred by random chance. Following this, Dunnett's test was used to identify which groups were significantly different from proto-oncogenes.

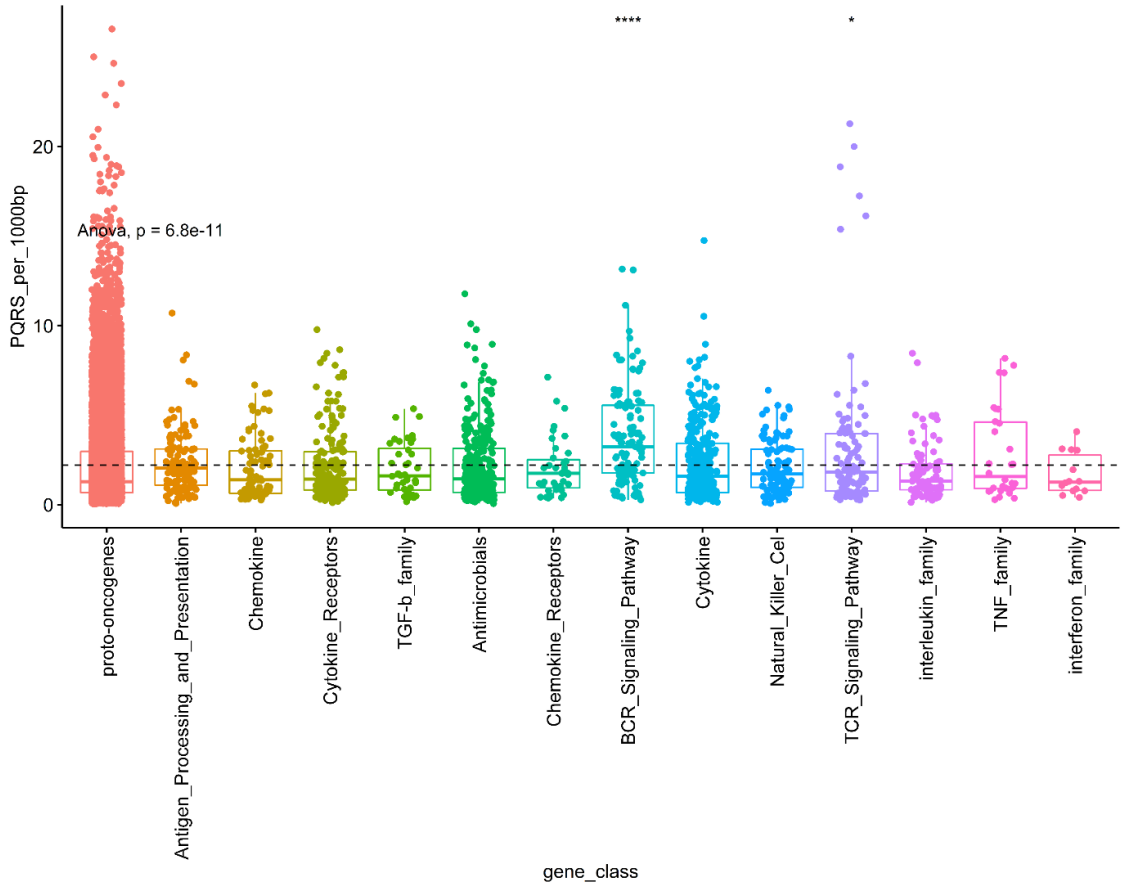


Figure 4.1.9: Box plot with ANOVA (p<2.2e-16) for comparing G4 density (G4/1000bp) in promoter of different immune classes to proto-oncogenes. X-axis represents the gene classes and Y-axis represent putative G4 (PQRS) per 1000 base pairs along the gene. The statistical significance denoted are results from Dunnett's test.

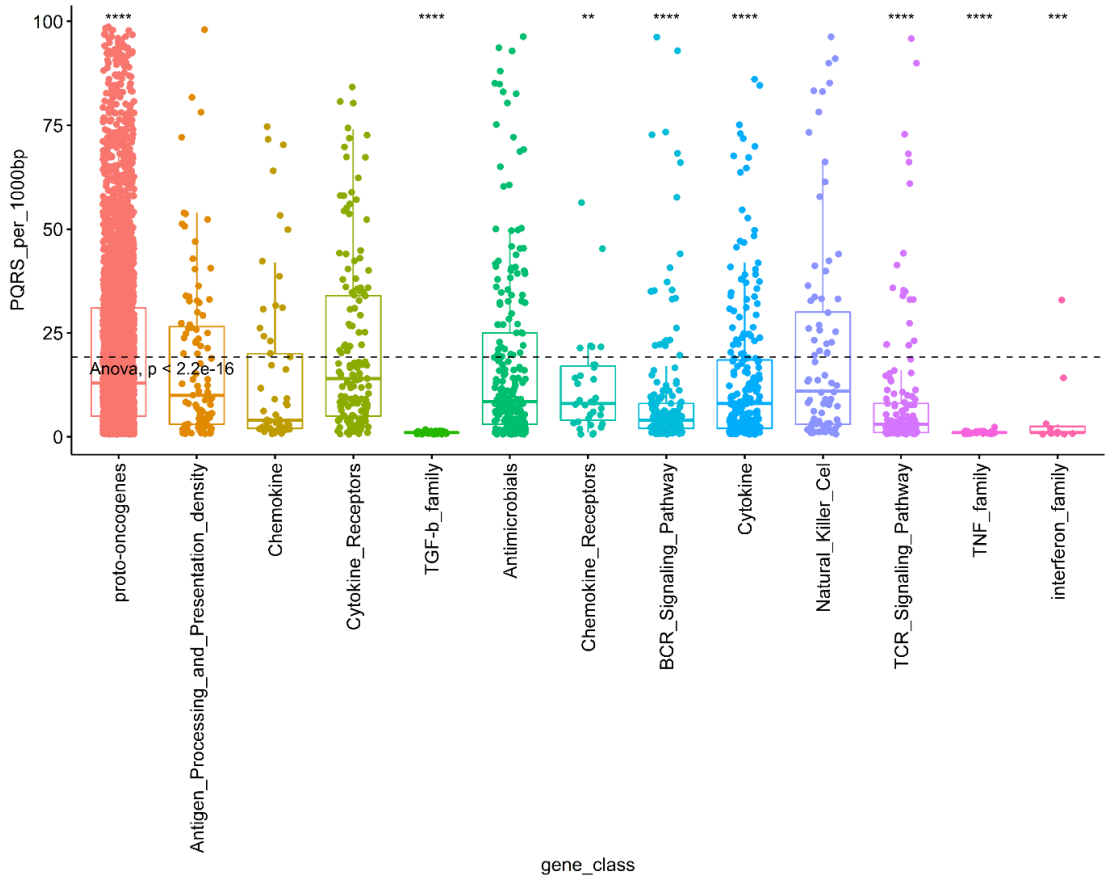


Figure 4.1.10: Box plot with ANOVA (p<2.2e-16) for comparing G4 density (G4/1000bp) in promoter of different immune classes to proto-oncogenes. X-axis represents the gene classes and Y-axis represent putative G4 (PQRS) per 1000 base pairs along the gene. The statistical significance denoted are results from Dunnett's test.

4.1.1.7. Mutations in G4 locations of the immune related genes

Putative G4 (PQRS) locations and mutations both were mapped onto genes and promoter region of genes related to immune system. Our analysis revealed that 943 immune-related genes contained PQRS along their gene sequences while 549 did not. Additionally, we found that 648 promoters associated with immune-related genes harboured PQRS structures, whereas 844 did not (Table 4.1.5). This observation highlights that putative PQRS were abundant along the DNA sequences of immune-related genes, in contrast to their lower abundance within promoters of these genes.

Out of all the PQRS locations present along the DNA sequences of immune-related genes, only 274 of them harboured mutations out of a total of 358,060 PQRS. In the case of PQRS located within the promoters of immune-related genes, only 729 out of 18,494 PQRS had mutations (Table 4.1.6).

	Genes	**Promoter**
Immune genes with G4	943(63%)	648(43%)
Immune genes without G4	549(36%)	844(56%)

Table 4.1.3: Number of immune related genes with and without G4 both along the gene and in promoter

	Gene	Promoter
G4 with mutation	274(0.076%)	729(3.94%)
G4 without mutation	357786(99.9%)	17765(96.05%)

Table 4.1.4: Number of G4 with and without mutations in promoters and along the gene of immune related genes

4.1.2. Discussion

The paradoxical behaviour of our immune system from cancer suppressive to cancer promoting is aided by mutations of the cancer genome. Mutations in immune related genes assist immune evasion in several ways (109). Our data depicts that mutation frequency in immune related genes was not uniform among different types of cancers studied (Figure 4.1.1). Heterogeneity across cancers and within a tumor is well documented and is attributed to the heterogeneity within the TME (110, 111). C>T was the most prevalent mutation in immune related genes of majority cancers studied, while in KICH, KIRC, KIRP, and LUSC had C>A as the most frequent mutation (Figure 4.1.1). For LUSC C>A mutation was assigned as a signature of cigarette smoke exposure whereas such a connection for kidney cancers is yet to be established (112, 113).

Indeed, single nucleotide polymorphisms (SNPs) within immune-related genes can exert significant effects on their functions and efficacy. Mutations in immune-related genes are anticipated to have negative effects, impacting immune pathways and immune cell infiltration within the tumor microenvironment. For instance, in the context of colorectal cancer, mutations in KRAS have been established to suppress immune pathways and immune cell infiltration (114) Similarly, mutations in components of the complement system can induce hypoxia, which in turn can contribute to poor survival outcomes (115). These examples underscore the critical role played by mutations in immune-related genes in cancer progression.

Genes positively selected during immunoediting can aid cancer cells to elude immune surveillance (22). Hence, it's reasonable to anticipate that these genes should experience frequent mutations and genomic instability. Among the genes related to the immune system, it was observed that PIK3CA stood out as one of the most frequently mutated genes across the various types of cancers examined (116). In addition to PIK3CA, TG emerged as another gene frequently mutated in 15 out of the 24 cancers examined (Figure 4.1.2). Additionally, we extended our research scope to include a list of 19 immune related

genes, all of which exhibited a high frequency of mutations. These genes could be grouped into three distinct categories based on their biological functions, warranting further in-depth study (Table 4.1.1).

Apart from mutations, CNA represent another source of genomic instability. These changes can significantly impact immune outcomes and responses in the context of various diseases, including cancer. In the study, PIK3CA and TG were once again emphasized, exhibiting amplification in 8 and 7 out of 24 cancers, respectively (Table 4.1.2). It has been previously demonstrated that in cervical cancer CNA involving driver genes can influence immune infiltration suggesting that there lies an intricate interplay between genomic alterations and immune responses in cancer development and progression (117). In hepatocellular carcinoma, tumors characterized by broad CNA tend to exhibit features associated with immune evasion. Conversely, tumors with lower burdens of CNA often display a more conducive to immune activity and infiltration (118). Indeed, changes in the nucleotide sequence, including mutations and copy number alterations, can result from various factors. These include the damaging effects of mutagens or errors during DNA replication (119). In some cases, these factors can be facilitated by non-canonical DNA structures, and G-quadruplexes serve as an illustrative example of this phenomenon(120).

It has been documented in the literature that these non-canonical DNA structures, such as G-quadruplexes(G4), have the potential to induce genome instability. They can do so by instigating mutations and promoting recombination events (121). In addition, they have a well-established connection to the immune system. They have been found in the promoters and genes of inflammatory mediators (12), have a role in facilitating class-switching (122) and holds the potential to activate innate immune genes through interactions with G4 ligands (123). This highlights their significance in immune processes. Hence, we mapped putative G4 locations and immune gene locations on the human genome, this revealed that the locations overlap at various points along the

genome (Figure 4.1.3). It is well-documented that G4 density in cancer genes is notably high (124). G4 were associated with RNA regulation therefore we expected housekeeping genes might have a lower density of G4 as they remain constitutively expressed and have a relatively simple regulatory system. G4 density in immune related genes were higher than proto-oncogenes and housekeeping genes (Figure 4.1.4) but the results were opposite for G4 density in promoters of the three gene classes (Figure 4.1.5). We conducted a comparison of G-quadruplex (G4) density in various immune gene classes with the G4 density observed in proto-oncogenes. Our analysis revealed that in specific immune classes, such as the TCR and BCR signalling pathways, G4 density along the gene sequences was even higher than that observed in proto-oncogenes (Figure 4.1.6). This suggest G4 structures might influence complex regulation at mRNA level in the immune classes. However, in the promoters of immune classes like cytokines, TGF-β and TNF family genes, and TCR and BCR pathway genes, the G4 density was lower compared to that in proto-oncogenes (Figure 4.1.7). Although a substantial number of immune-related genes had G4 within both their gene sequences and promoters (Table 4.1.5), it was observed that mutations in the G4 locations of these genes were not particularly common (Table 4.1.6). Consequently, we can infer that G4 structures did not appear to have a significant influence on the occurrence of mutations in immune-related genes. A considerable amount of research has been dedicated to developing effective G4 ligands for cancer therapy, but thus far, these efforts have not yielded significant success. However, a more comprehensive analysis of G4 structures within the immune system could provide valuable insights and potentially lead to the development of G4-based therapeutics in cancer.

In summary, mutational profile of immune related genes is not uniform throughout different cancers. Among the immune-related genes, TG and PIK3CA stood out as the two most frequently mutated genes. Including these two genes we obtained a list of 19 genes that were frequently mutated among different cancers in the study. These 19 genes could be categorised into three biological processes namely cell regulation, cell

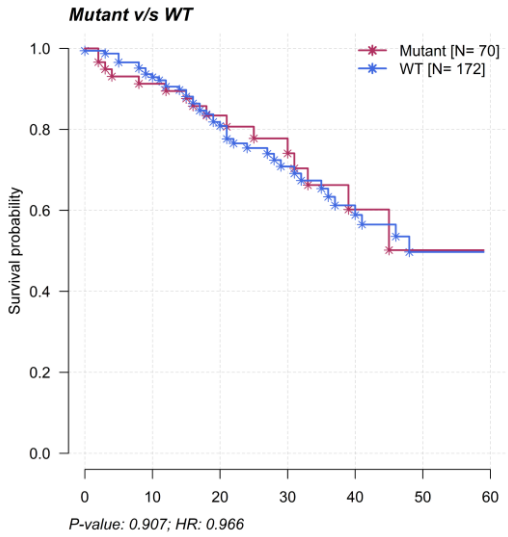
migration, and antigen presentation. Given that both PIK3CA and TG exhibited frequent mutations, we undertook a thorough examination of these genes on how mutations in them influence survival and alter the tumor immune microenvironment. PIK3CA has been previously established as a cancer driver gene, whereas TG is not. Pan-cancer studies typically classify genes as recurrently mutated based on scores calculated using various methods. In contrast, we identified frequently mutated genes based on the number of times they were reported as such for a specific cancer type. TG mutations were observed in only 2–11% of the samples, which may explain why it was not classified among frequently mutated genes since our study focused exclusively on immune-related genes TG came into focus(125-127). We considered all the four genes in antigen processing and presentation category for same analysis. Our initial expectation was that mutations in immune-related genes might be influenced by G-quadruplexes (G4), given the established connection of G4 structures to immune responses. However, our analysis did not reveal a significant prevalence of mutations in immune response related genes within putative G4 (PQRS) locations. This suggests that the presence of G4 structures may not be a primary driver of mutations in these genes.

4.2. Functional in silico analysis of immune related genes with genomic instability

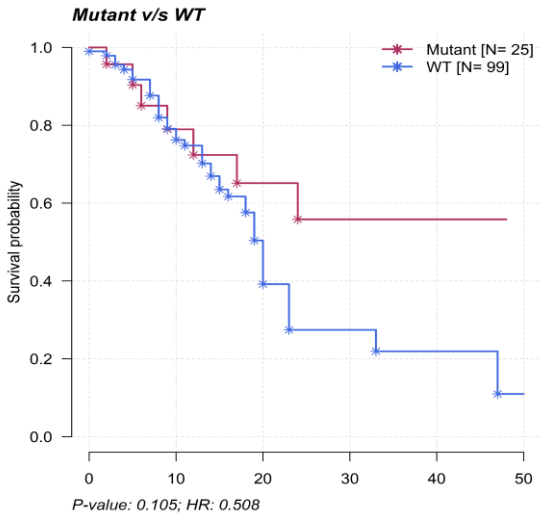
4.2.1. Results:

4.2.1.1. Kaplan Meier survival curves for PIK3CA mutations

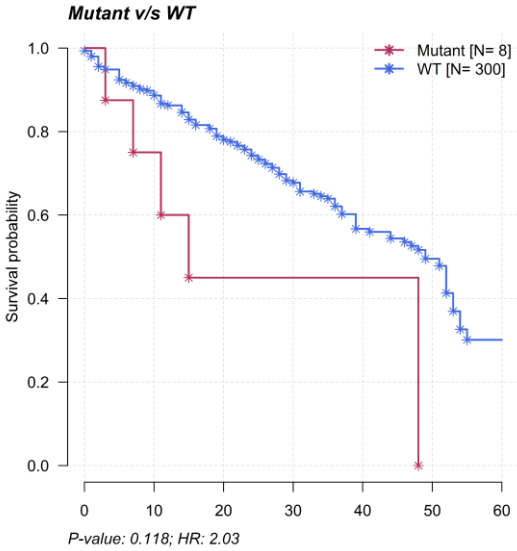
The implication of mutations in PIK3CA on survival was analysed using Kaplan Meier survival curves. The analysis showed that the influence of PIK3CA mutations on survival varied across the studied cancers. In LGG, KIRC and GBM odds of having a 5-year survival time was lower with mutant PIK3CA; however, it was not statistically significant. Higher survival was seen to be associated with mutant alleles in UCEC, suggesting mutations were influencing better survival. Similar results were found in BLCA but failed to reach statistical significance (Figure 4.2.1).



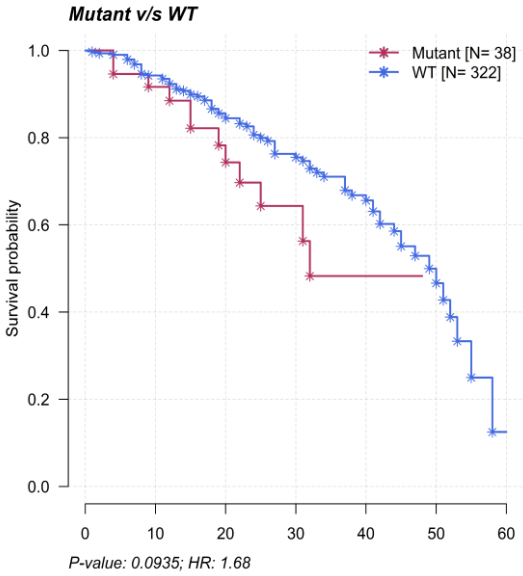
CESC



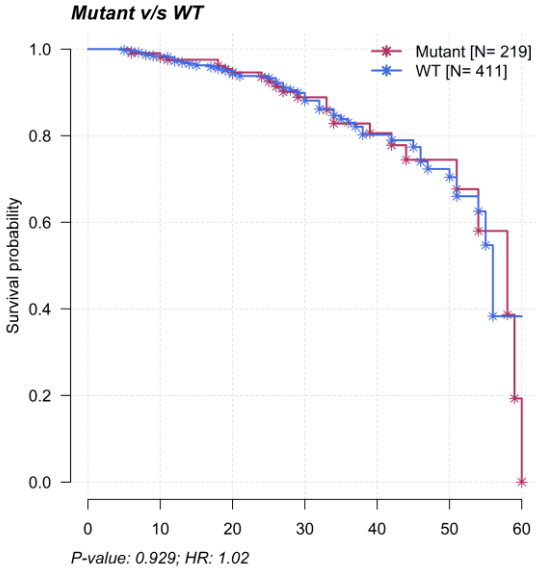
BLCA



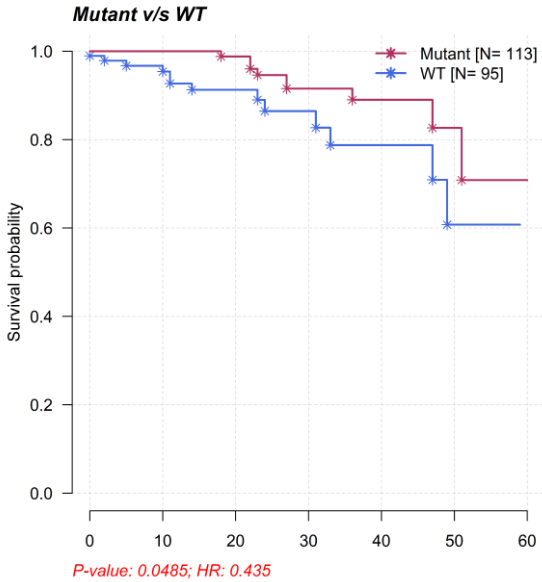
KIRC



LGG



BRCA



UCEC

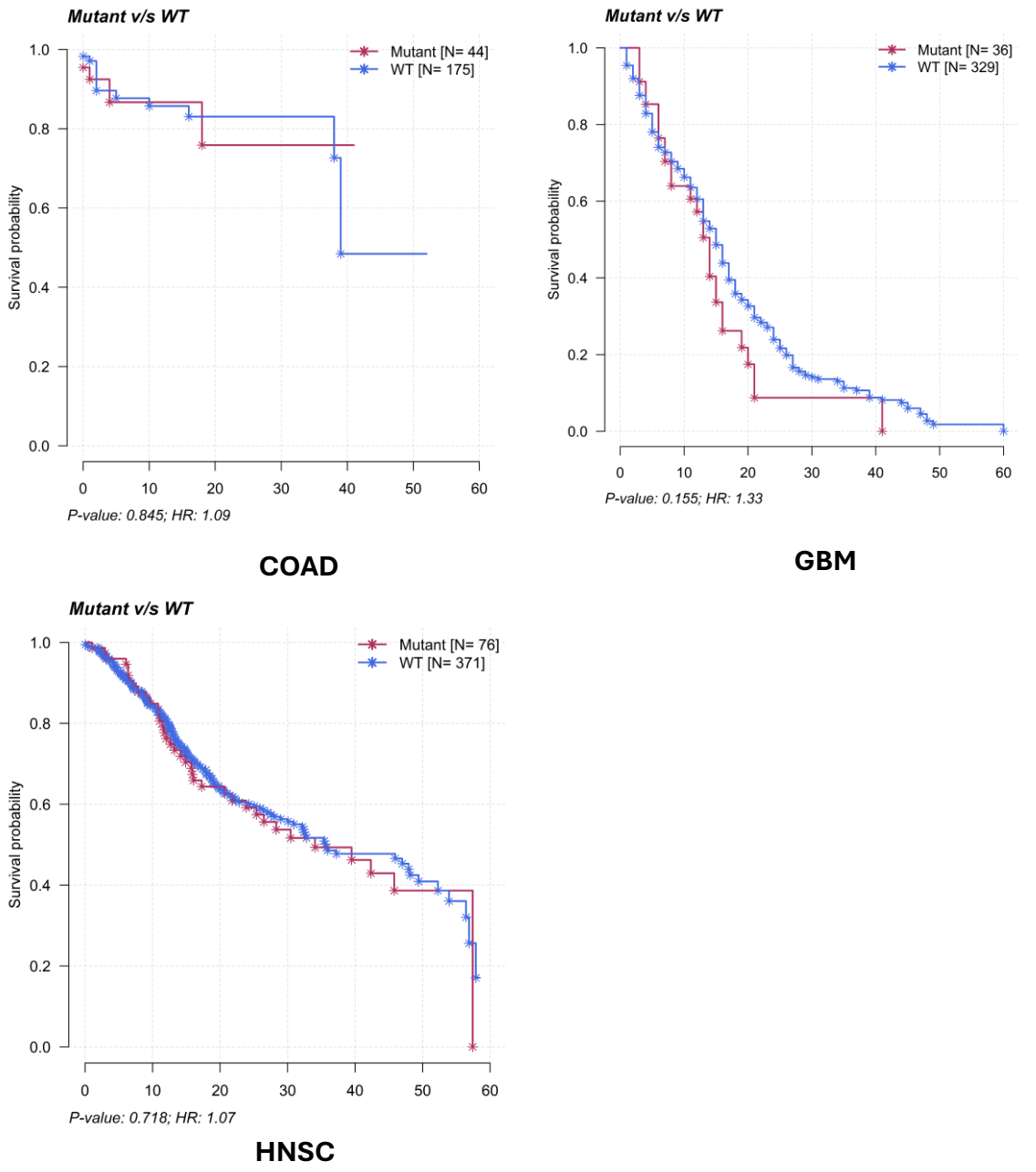


Figure 4.2.1: 5year Kaplan Meier survival curves for mutation in PIK3CA in different cancers were generated by maftools. Survival curves for mutant (Mutant) is depicted in red while for wildtype (WT) it was blue, the X-axis represent time (in months) whereas Y-axis represents survival probability.

4.2.1.2. Cox- regression analysis of mutated and non-mutated PIK3CA dataset of UCEC and LGG

Both univariate and multivariate cox-proportional hazard analysis considering age for LGG and UCEC using expression data was performed to evaluate the association between survival and mutation. The results for univariate analysis in mutated PIK3CA dataset of UCEC supported Kaplan Meier analysis with hazard ratio of 0.4962 (p-value 0.0419) (Table 4.2.1(a)). In LGG, the result for multivariate analysis showed that with advancement in age PIK3CA expression has a negative influence on survival (Table 4.2.1(b)).

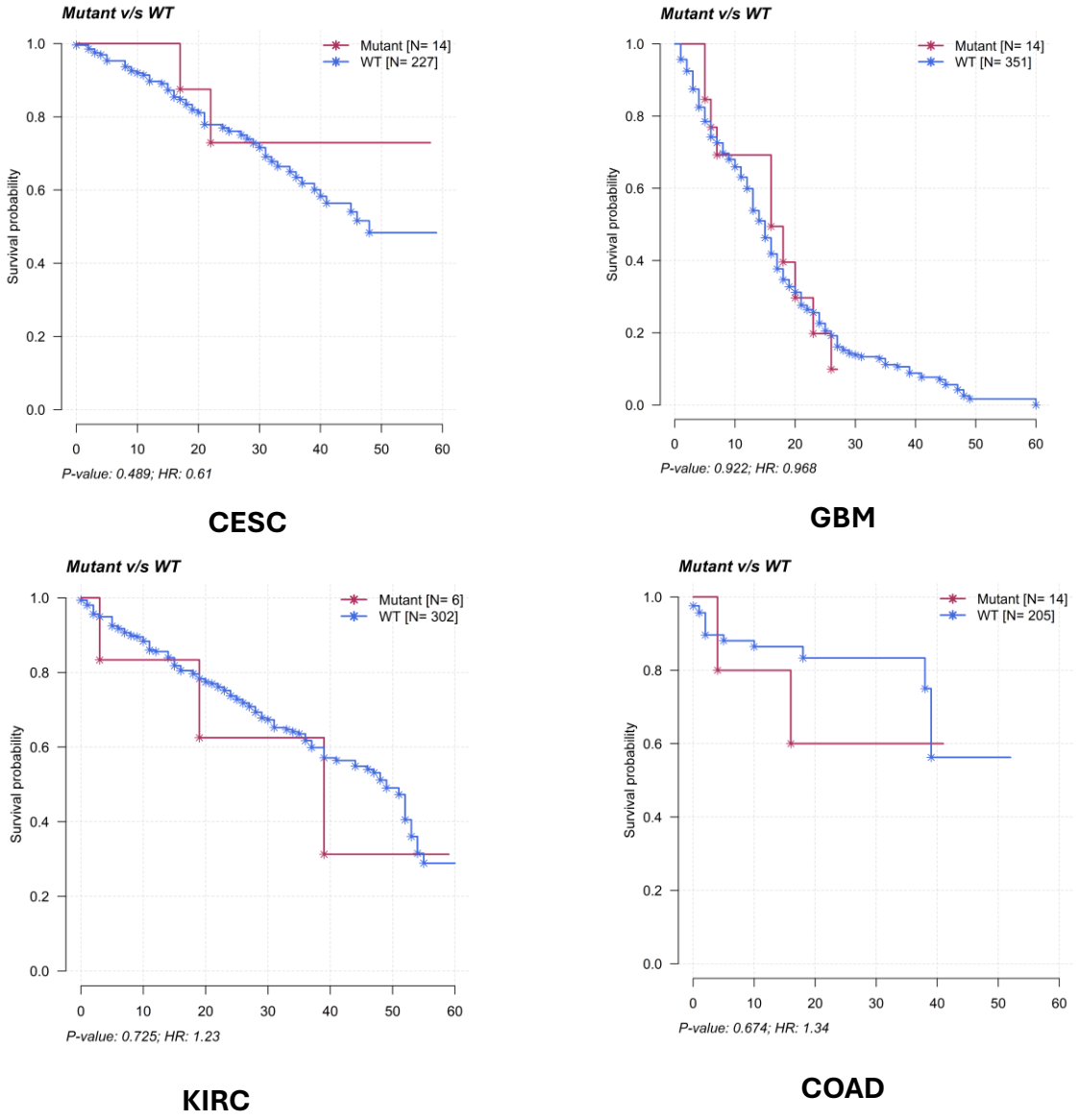
(a)Univariate				
	Mutated		Not Mutated	
	HR	P-value	HR	P-value
UCEC	0.4962	0.0419	0.8686	0.516
LGG	1.2977	0.441	1.09151	0.389

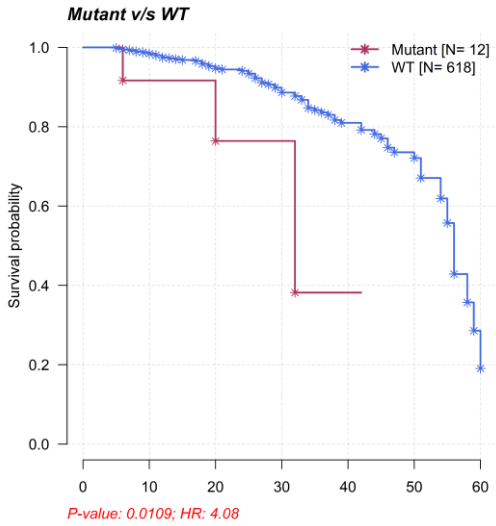
(b)Multivariate					
		Mutated		Not Mutated	
		HR	P-value	HR	P-value
UCEC	PIK3CA	0.96813	0.2835	1.04349	0.0449
	Age	0.55972	0.0952	0.89347	0.6086
LGG	PIK3CA	2.41202	0.04105	1.03686	0.710
	Age	1.13668	0.00151	1.05486	1.84e-12

Table 4.2.1: Cox-proportional hazard analysis for both mutated and non-mutated datasets of UCEC and LGG (a) Univariate (b) Multivariate

4.2.1.3. Kaplan Meier survival curves for TG mutations

The impact of TG mutations on survival was analysed using Kaplan-Meier survival curves and it varied among the different cancers examined. In ~~both HNSC and~~ BRCA, the presence of mutant TG was associated with a decreased likelihood of achieving a 5-year survival time. However, it is important to note that this result reached statistical significance only for BRCA (Figure 4.2.2).





BRCA

Figure 4.2.2: 5year Kaplan Meier survival curves for mutation in TG in different cancers were generated by maftools. Survival curves for mutant (Mutant) is depicted in red while for wildtype (WT) it was blue, the X-axis represent time (in months) whereas Y-axis represents survival probability.

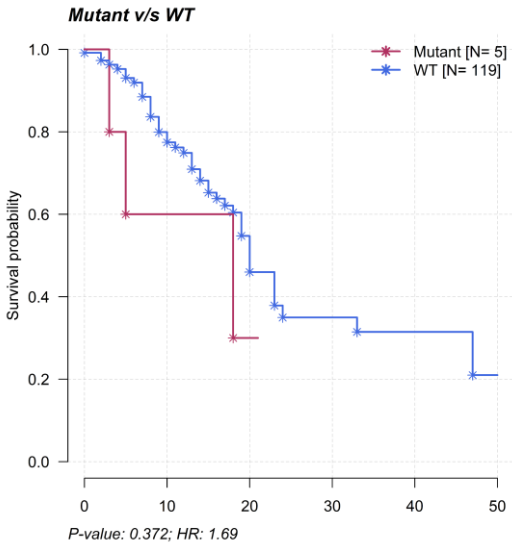
4.2.1.4. Survival data for mutations of the 4 antigen presentation genes

The impact of mutations in the four antigen-presenting genes, namely HLA-A, HLA-B, HLA-DRB1, and CIITA, on survival was studied using Kaplan Mier survival curves. In CESC, COAD and HNSC mutations in HLA-A promotes better survival in 5 years survival time. In BRCA odds of having a 5-year survival time was higher with mutant HLA-B. Mutant alleles of CIITA were associated with better survival in BRCA, CESC, LGG, UCEC. HLA-DRB1 mutants showed better survival for BRCA. However, due to the small number of samples with mutations and clinical information, we were unable to achieve statistical significance in some cases; those that did, had a very small number of samples. As a result, we examined the genes for their impact on survival using expression. (Table 4.2.2) (Figure).

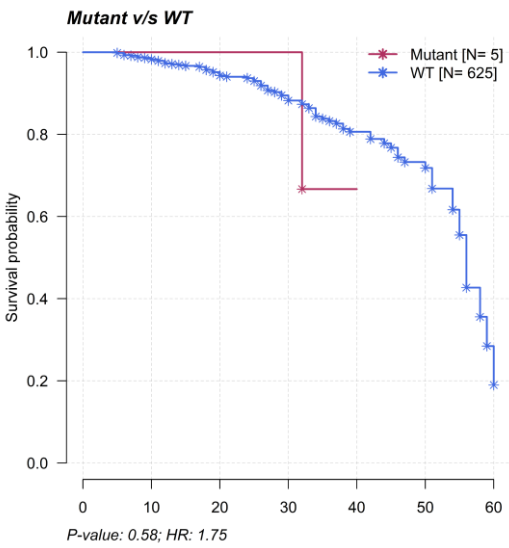
Genes		BLCA	BRCA	CESC	COAD	GBM	HNSC	KIRC	LGG	UCEC
HLA-A	HR	1.69	0.58	0.0687	0.355		0.372	2.23		
	P-value	0.372	1.75	3.70E-08	3.70E-08		1.28E-08	0.251		
	Mutated sample	5/119	5/365	10/231	3/216		3/28	3/305		
HLA-B	HR	1.69	0.619	0.573			1.7	0.938		
	P-value	0.372	8.23E-07	0.436			0.618	0.931		
	Mutated sample	5/119	5/625	13/228			3/29	4/304		
HLA-DRB1	HR		0.387					0.0126		
	P-value		0.338					2.73		
	Mutated sample		10/625					7/301		
CIITA	HR	1.41	0.538	0.533	1.32	0.619	1.56	1.73	0.429	0.258
	P-value	0.779	3.02E-07	3.02E-07	0.728	0.819	0.676	0.573	3.02E-07	3.71E-07
	Mutated sample	1/123	5/625	2/239	8/211	9/356	2 /29	1/307	1/359	8/200

Table 4.2.2: Data from Kaplan Meier plots for survival considering mutations in the different cancers for the four antigen presenting genes considered in the study

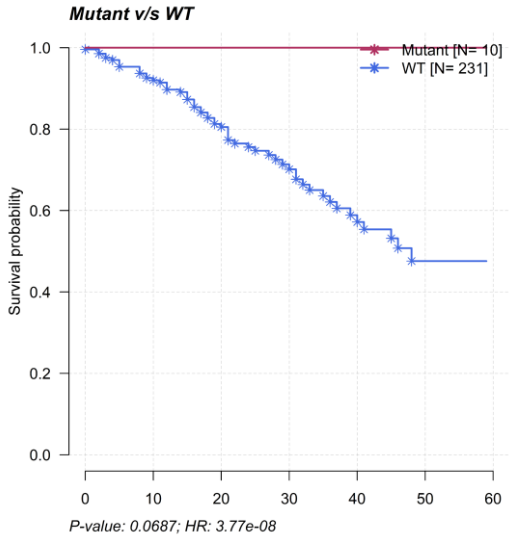
1. HLA-A



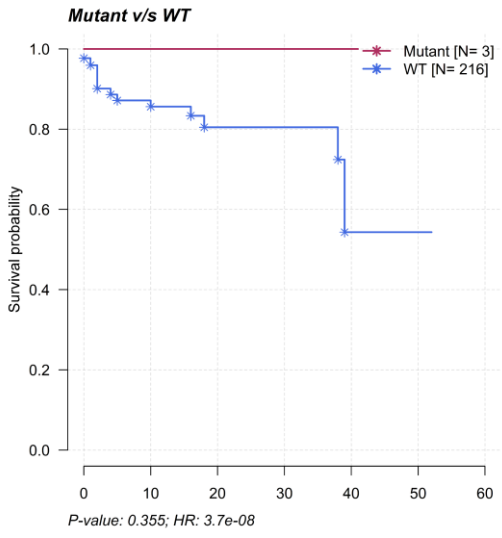
BLCA



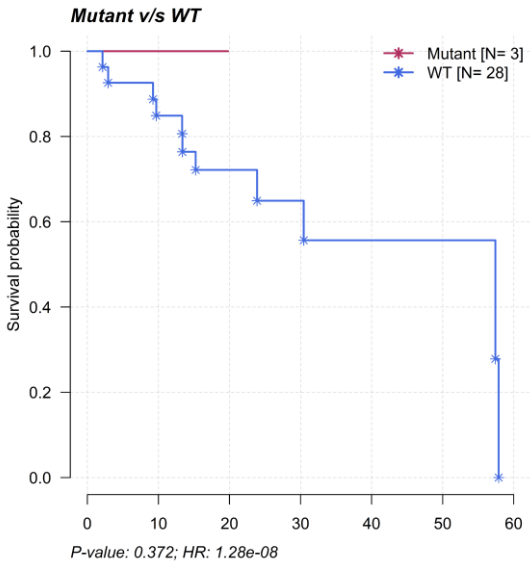
BRCA



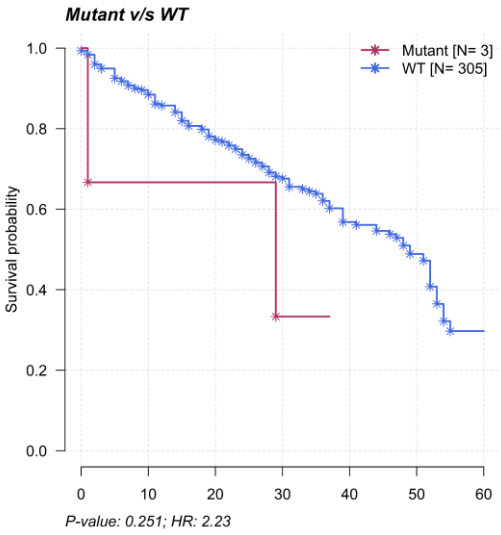
CESC



COAD

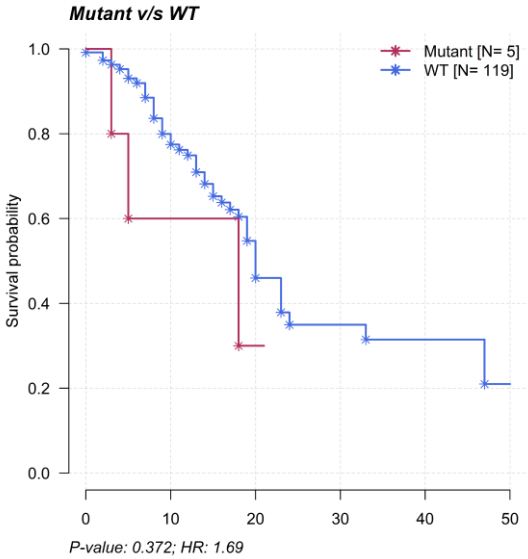


HNSC

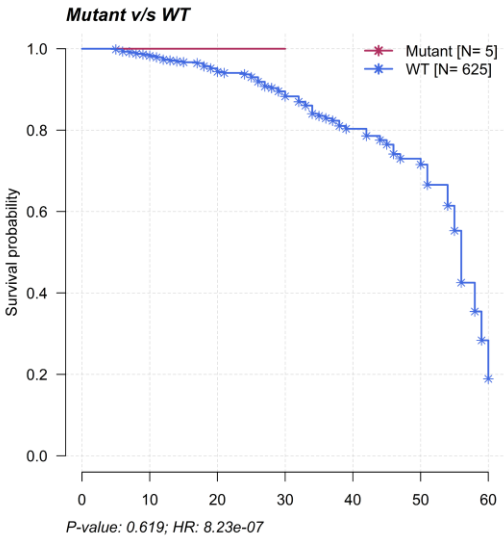


KIRC

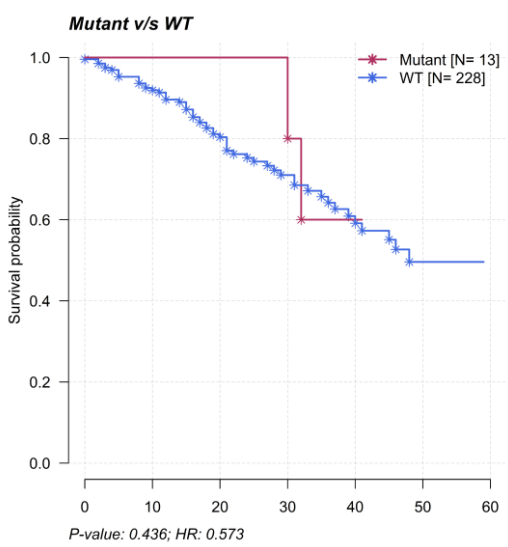
2. HLA-B



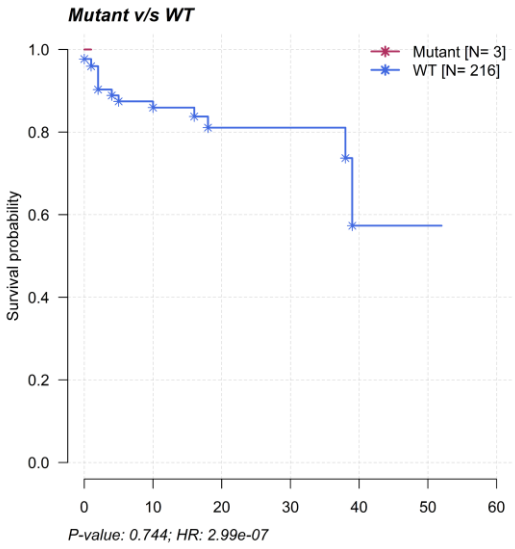
BLCA



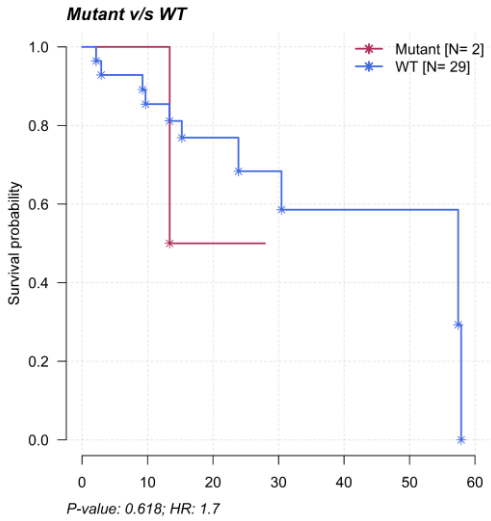
BRCA



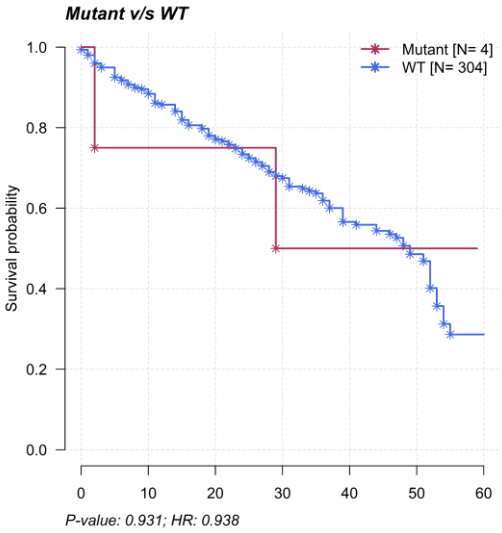
CESC



COAD

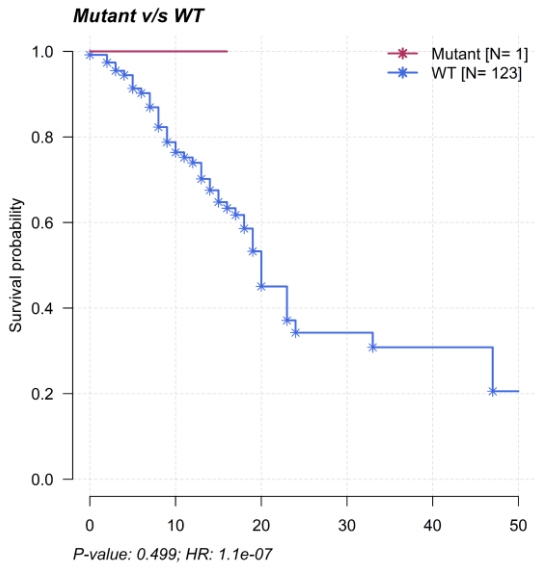


HNSC

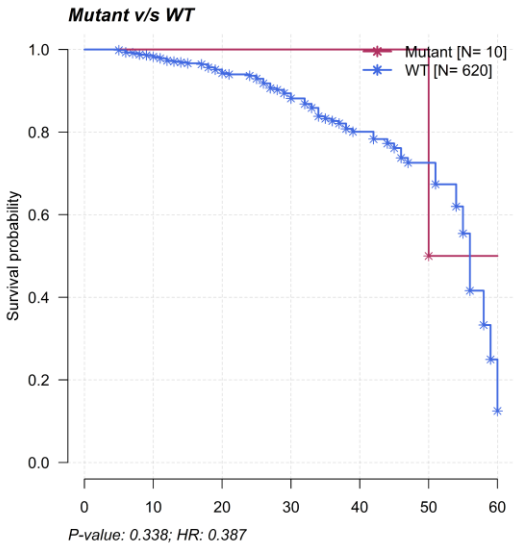


KIRC

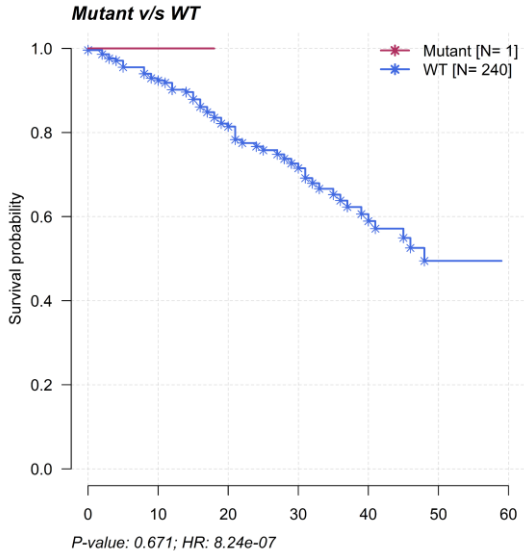
3. HLA-DRB1



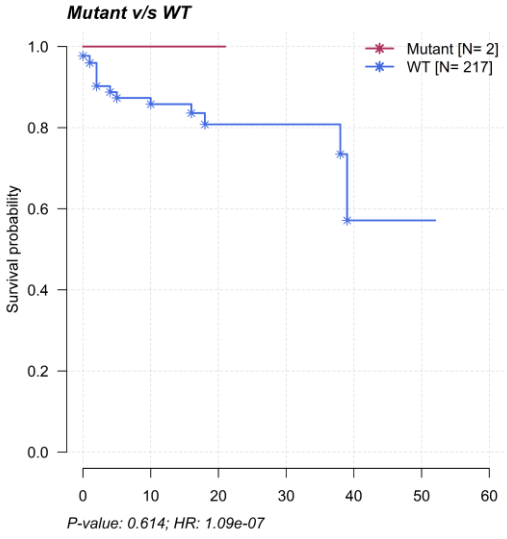
BLCA



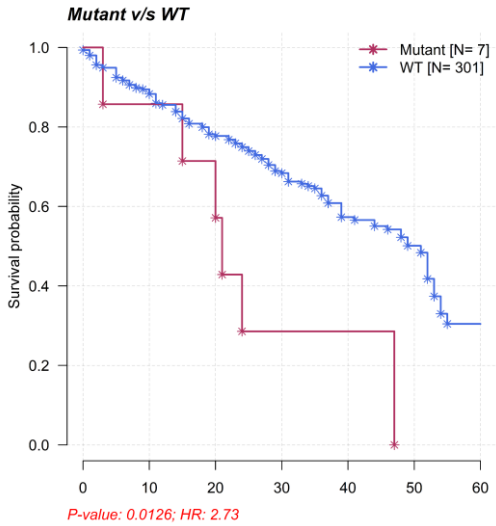
BRCA



CESC

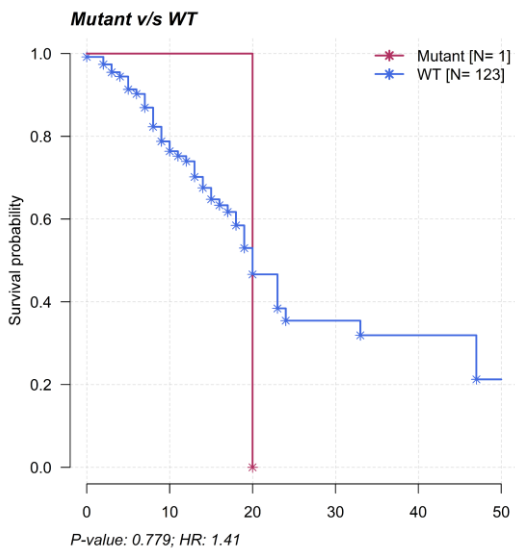


COAD

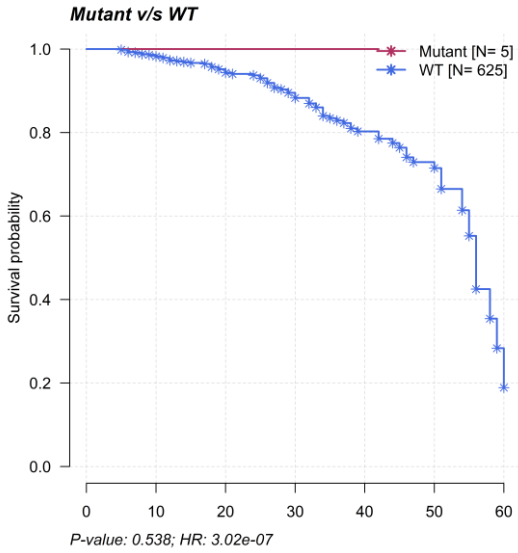


KIRC

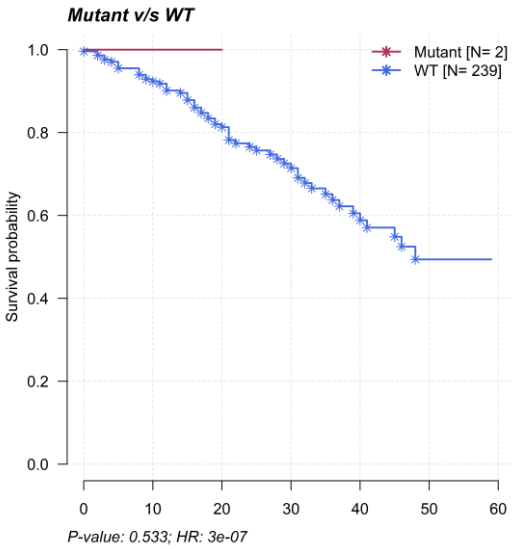
4. CIITA



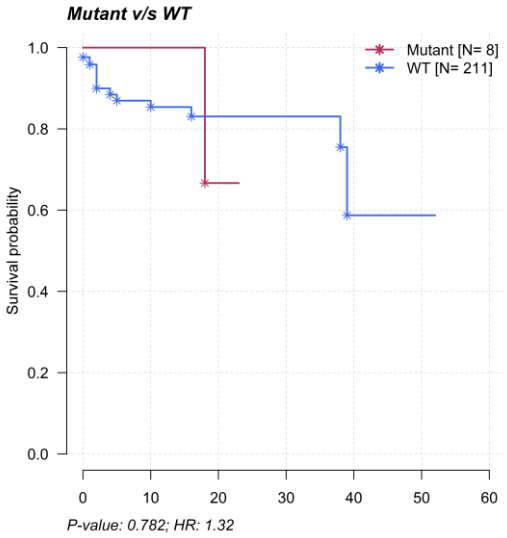
BLCA



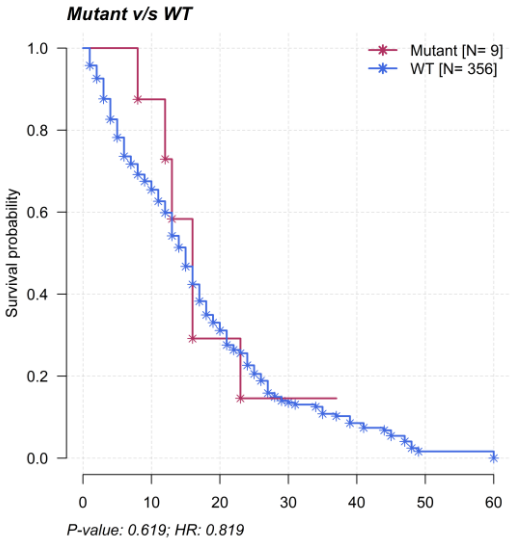
BRCA



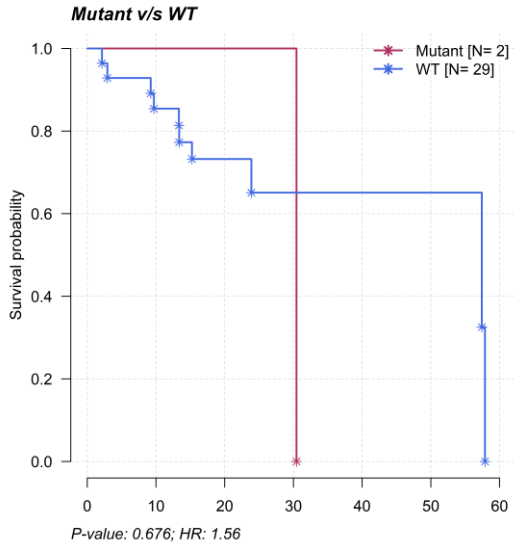
CESC



COAD



GBM



HNSC

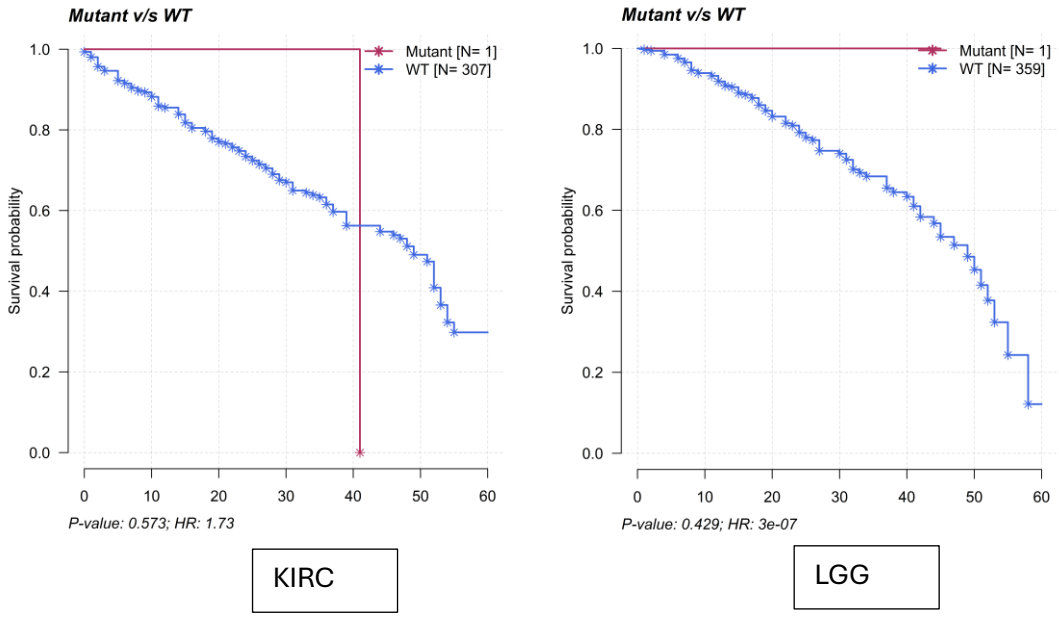


Figure 4.2.3: Kaplan Meier survival curves for mutated versus wild type samples of cancers in 5 year survival time in 4 antigen presenting genes (1. HLA-A, 2. HLA-B, 3. HLA-DRB1, 4. CIITA) in our study. Survival curves for mutant (Mutant) is depicted in red while for wildtype (WT) it was blue, the X-axis represent time (in months) whereas Y-axis represents survival probability

4.2.1.5. Survival analysis based on the expression of the four antigen-presenting genes

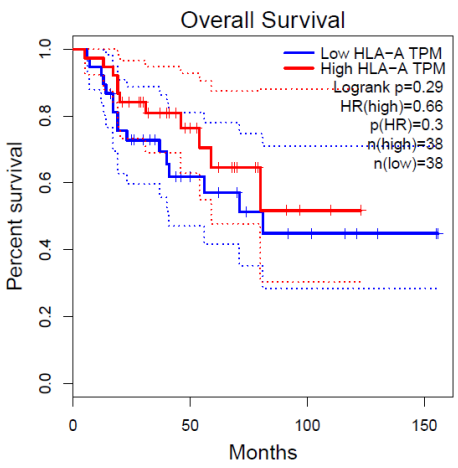
The effect of altered expression of the four antigen-presenting genes, HLA-A, HLA-B, HLA-DRB1, and CIITA, on survival was investigated using Kaplan Mier survival curves (Supplementary Figure3). In LGG, high expression levels of the four antigen-presenting genes examined in our study were related to poor survival. In SARC, lower HLA-A, HLA-B, and CIITA expression of were associated with poorer survival outcomes. Lower HLA-A and CIITA expression correlated to poorer survival outcomes in KIRP and HNSC respectively whereas higher expression of HLA-B was associated with poorer survival outcomes in LUAD. In BRCA and LUAD, lower levels of HLA-DRB1 were related to poor survival outcomes (Table 4.2.3).

Sl. No	Cancer	HLA-A(high)		HLA-B(high)		HLA-DRB1(high)		CIITA (high)	
		HR	p-value	HR	p-value	HR	p-value	HR	p-value
1	ACC	0.66	0.3	0.81	0.59	0.52	0.1	0.83	0.63
2	BLCA	0.91	0.55	0.97	0.84	1	0.78	0.85	0.29
3	BRCA	0.8	0.16	0.94	0.69	0.73	0.055	0.79	0.16
4	CESC	0.87	0.54	0.93	0.76	0.72	0.17	0.77	0.27
5	CHOL	1.3	0.56	1.3	0.56	1.3	0.56	0.98	0.97
6	COAD	0.7	0.15	0.82	0.41	0.97	0.91	0.83	0.46
7	DLBC	2.2	0.33	1.4	0.6	2.9	0.2	1.9	0.38
8	ESCA	1.3	0.21	1.4	0.16	0.77	0.27	1.1	0.69
9	GBM	0.94	0.73	1.2	0.2	0.98	0.92	1.2	0.4
10	HCC	0.97	0.88	0.93	0.7	0.84	0.31	0.89	0.5
11	HNSC	1.2	0.1	1.1	0.71	0.95	0.71	0.72	0.016
12	KICH	1.3	0.72	1.3	0.7	0.29	0.12		0.91
13	KIRC	0.95	0.72	0.91	0.55	0.84	0.26	0.85	0.28
14	KIRP	0.55	0.056	0.7	0.35	0.77	0.4	0.71	0.27

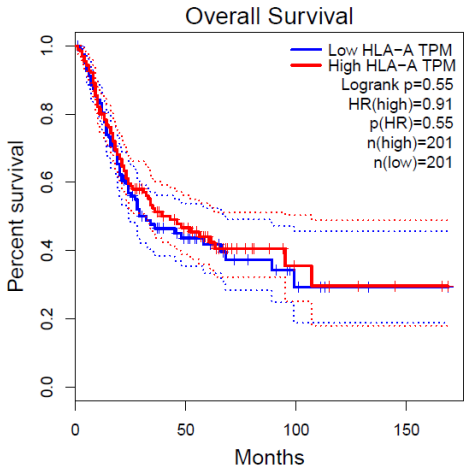
15	LAML	1.3	0.38	0.8	0.15	1.1	0.78	1.4	0.21
16	LGG	2.1	0.00013	2	0.00019	1.6	0.01	2.7	1.80E-07
17	LUAD	1	0.96	1.3	0.057	0.68	0.011	0.76	0.067
18	LUSC	1.1	0.44	1.1	0.86	1.1	0.38	1	0.93
19	OV	0.85	0.2	0.81	0.094	0.97	0.81	1.1	0.69
20	PCPG	1.1	0.92	0.56	0.51	0.51	0.45	0.23	0.18
21	PRAD	1.4	0.62	0.48	0.28	0.7	0.59	0.75	0.66
22	SARC	0.66	0.042	0.64	0.03	0.71	0.099	0.68	0.059
23	THCA	0.35	0.075	0.75	0.58	0.95	0.91	1.5	0.45
24	UCEC	0.79	0.49	0.94	0.86	0.5	0.06	0.47	0.044

Table 4.2.3: Influence of change in expression levels of the 4 antigen presenting genes in our study on survival. The hazard ratios in the table represent the impact of higher expression levels of the respective genes.

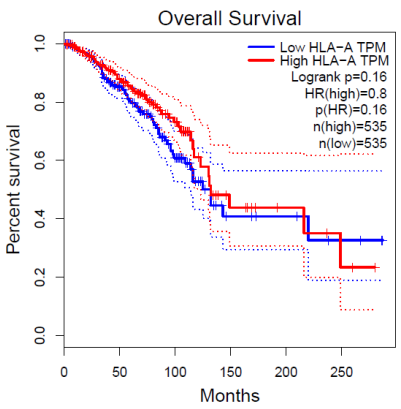
1. HLA-A



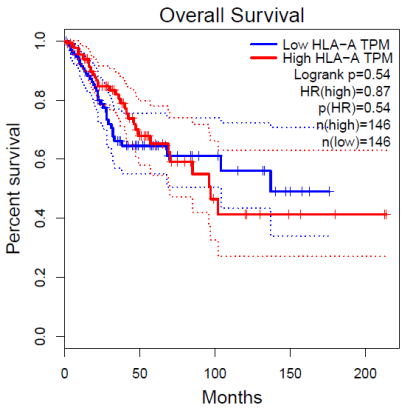
ACC



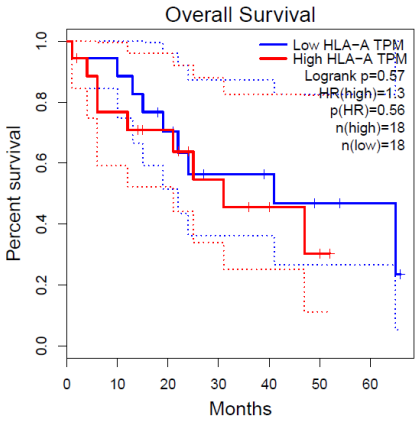
BLCA



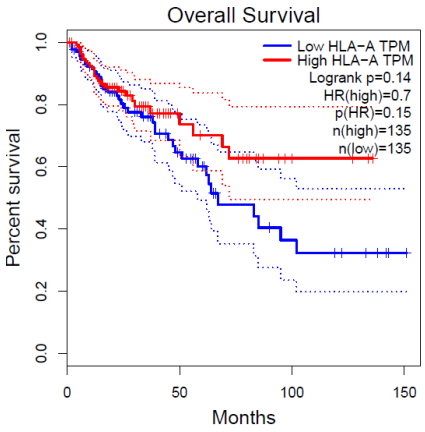
BRCA



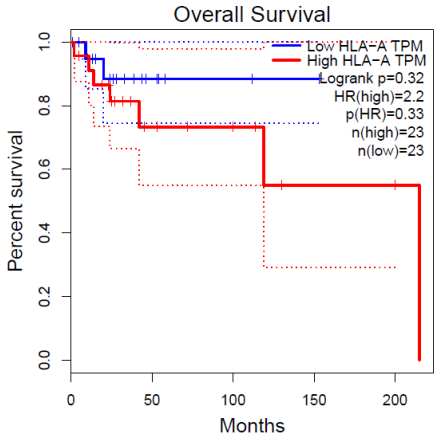
CESC



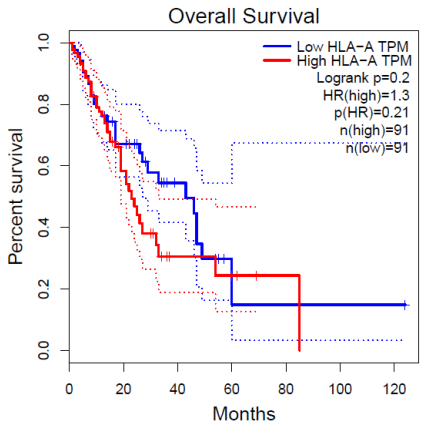
CHOL



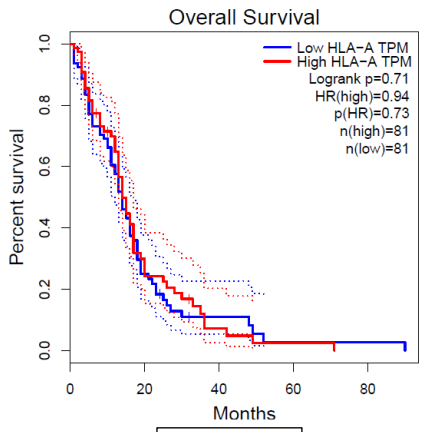
COAD



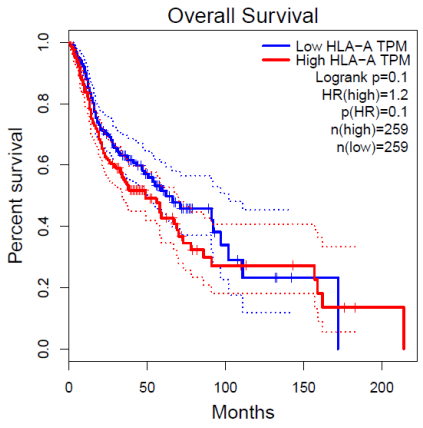
DLBC



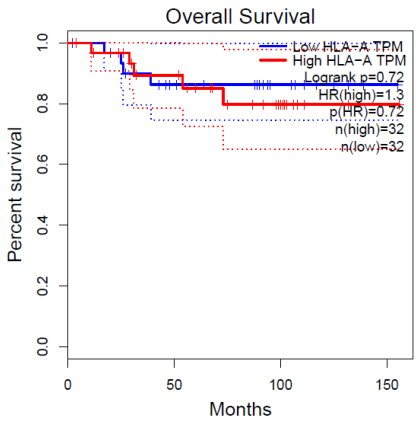
ESCA



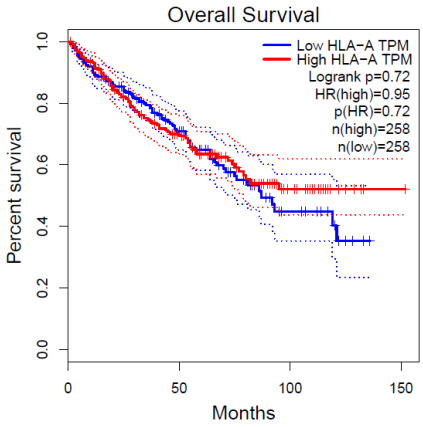
GBM



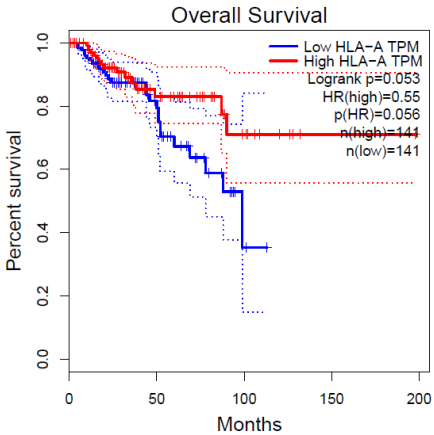
HNSC



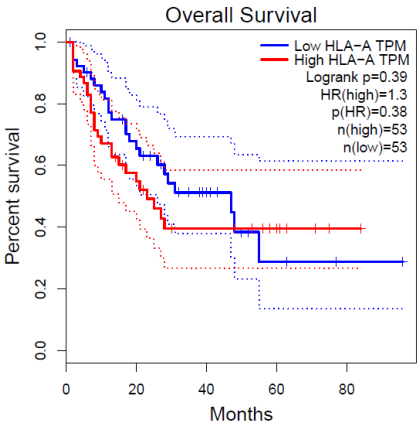
KICH



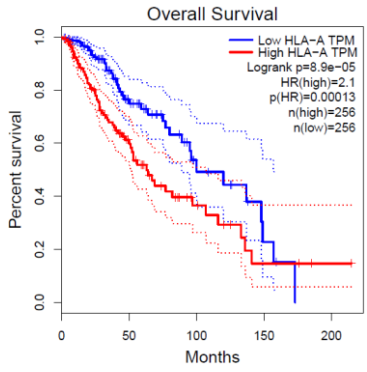
KIRC



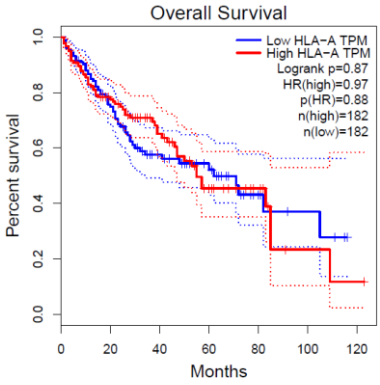
KIRP



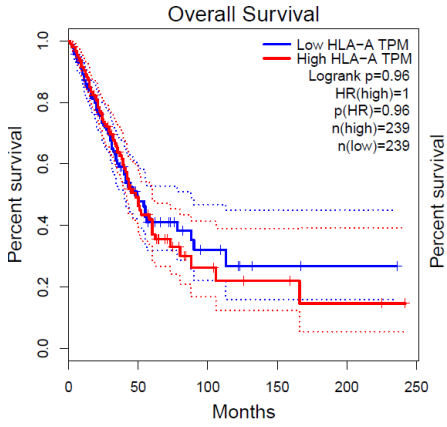
LAML



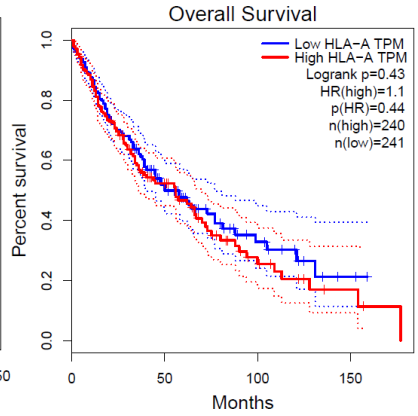
LGG



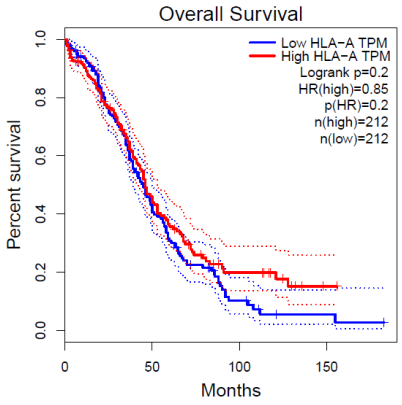
LIHC



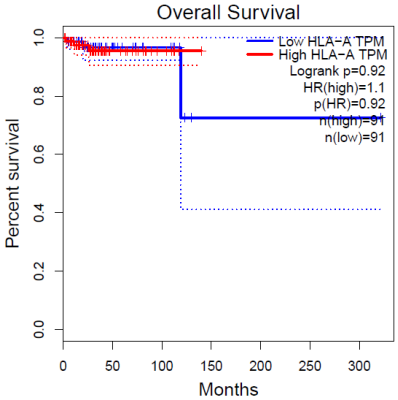
LUAD



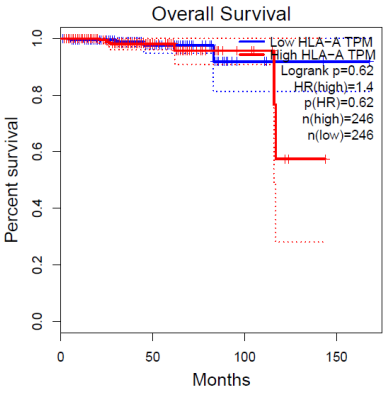
LUSC



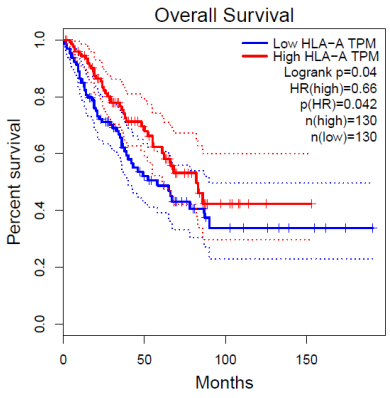
OV



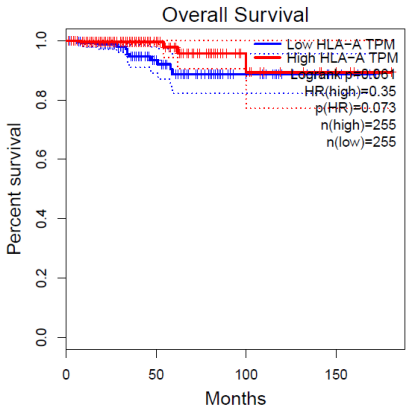
PCPG



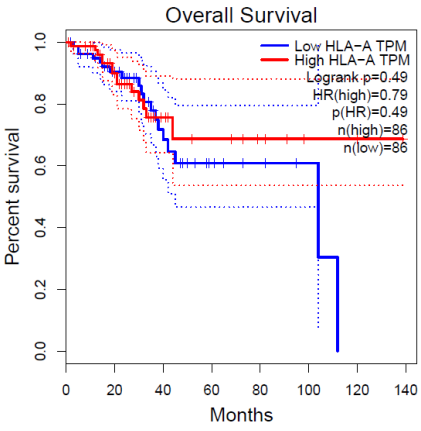
PRAD



SARC

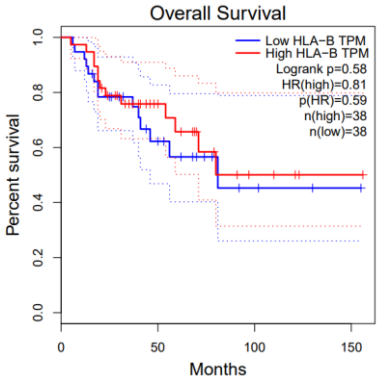


THCA

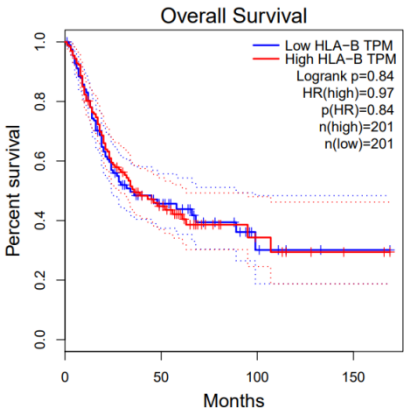


UCEC

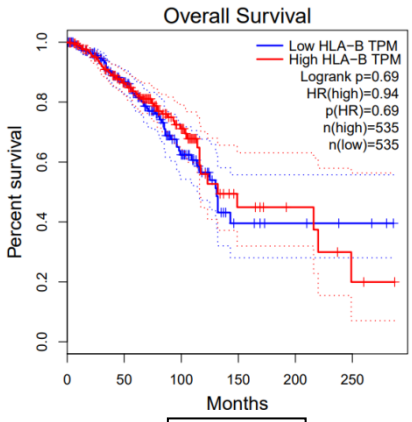
2. HLA-B



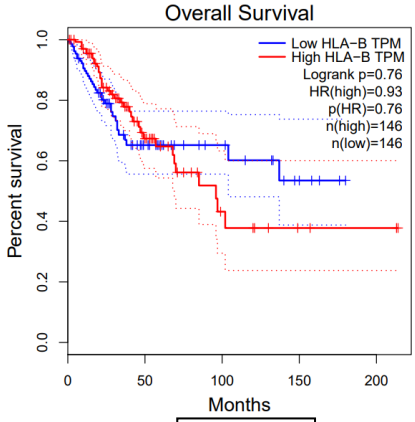
ACC



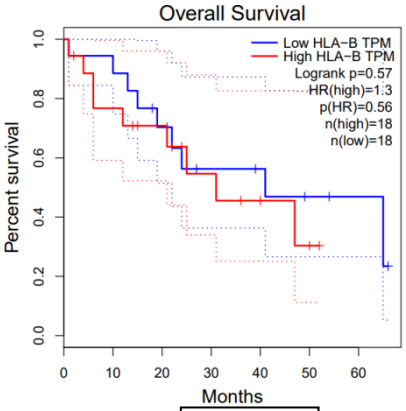
BLCA



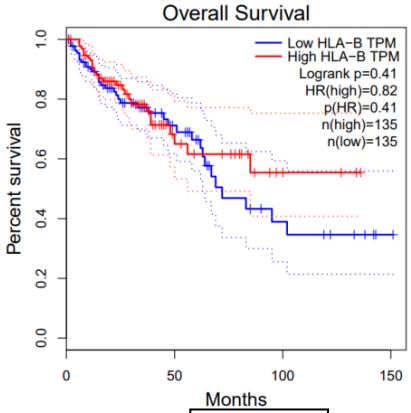
BRCA



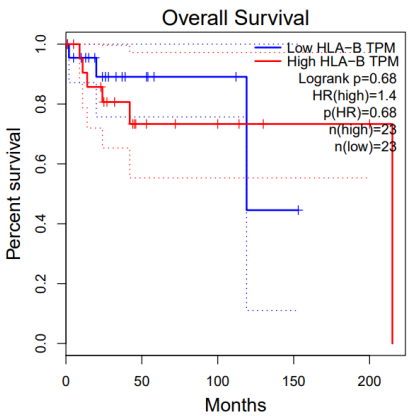
CESC



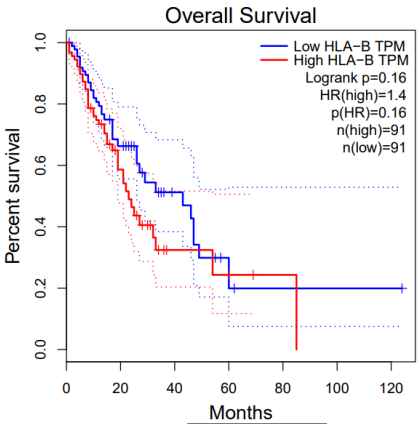
CHOL



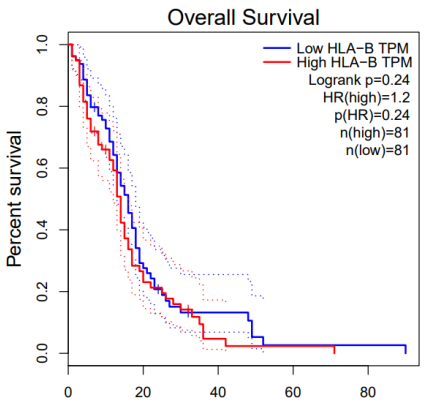
COAD

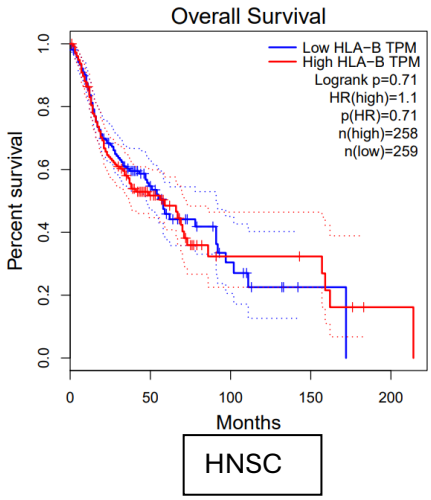


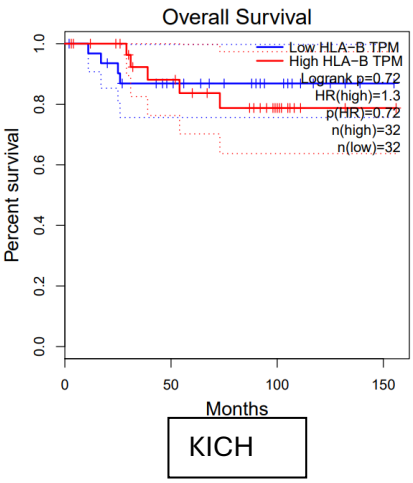
DLBC

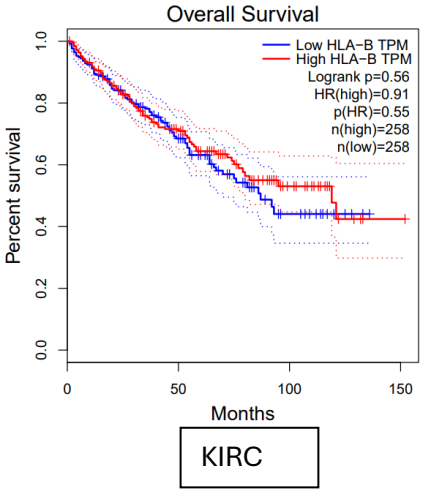


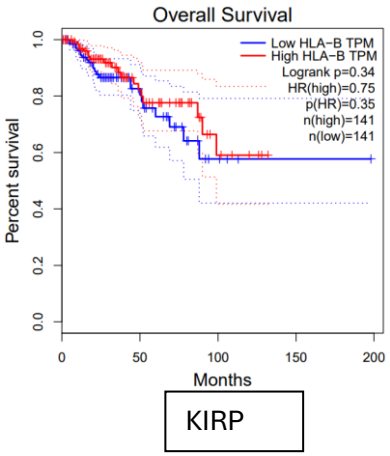
ESCA

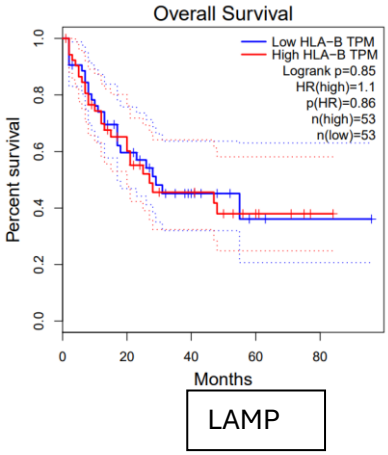

Overall Survival
Low HLA−B TPM
High HLA−B TPM
Logrank p=0.24
HR(high)=1.2
p(HR)=0.24
n(high)=81
n(low)=81
Percent survival
Months
GMB

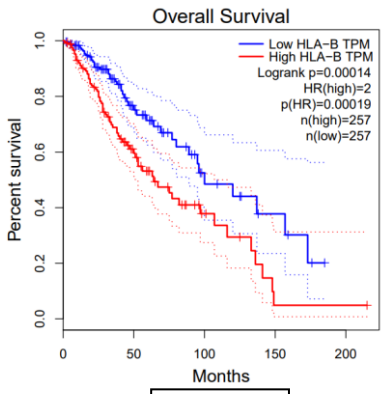

Overall Survival
Low HLA−B TPM
High HLA−B TPM
Logrank p=0.71
HR(high)=1.1
p(HR)=0.71
n(high)=258
n(low)=259
Percent survival
Months
HNSC


Overall Survival
Low HLA−B TPM
High HLA−B TPM
Logrank p=0.72
HR(high)=1.3
p(HR)=0.72
n(high)=32
n(low)=32
Percent survival
Months
KICH

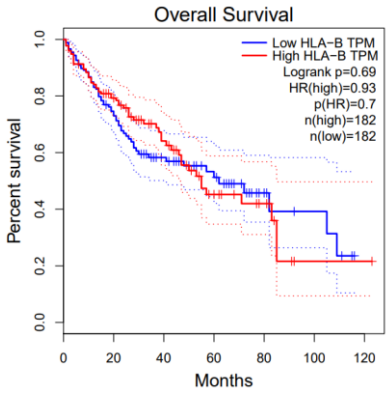

Overall Survival
Low HLA−B TPM
High HLA−B TPM
Logrank p=0.56
HR(high)=0.91
p(HR)=0.55
n(high)=258
n(low)=258
Percent survival
Months
KIRC


Overall Survival
Low HLA−B TPM
High HLA−B TPM
Logrank p=0.34
HR(high)=0.75
p(HR)=0.35
n(high)=141
n(low)=141
Percent survival
Months
KIRP

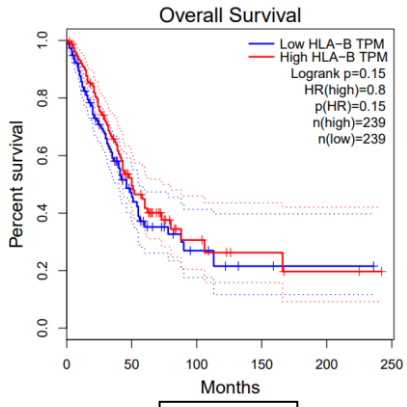

Overall Survival
Low HLA−B TPM
High HLA−B TPM
Logrank p=0.85
HR(high)=1.1
p(HR)=0.86
n(high)=53
n(low)=53
Percent survival
Months
LAMP



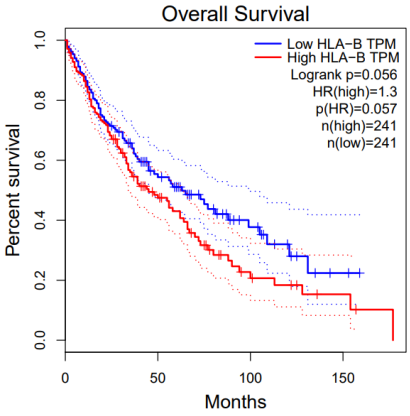
LGG



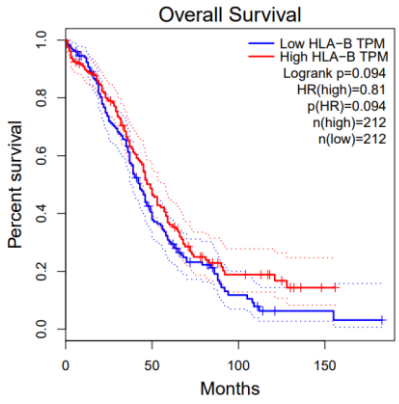
LICH



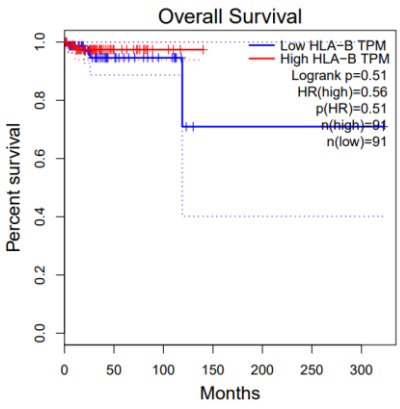
LUAD



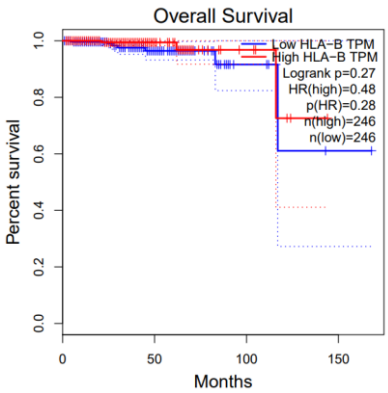
LUSC



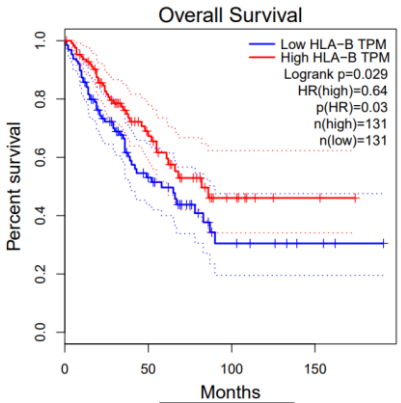
OV



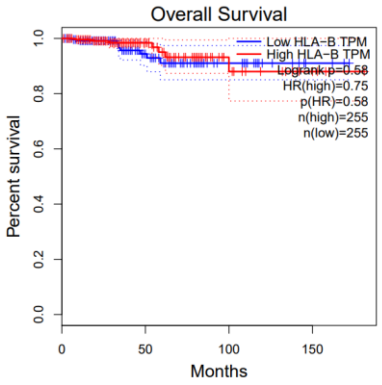
PCPG



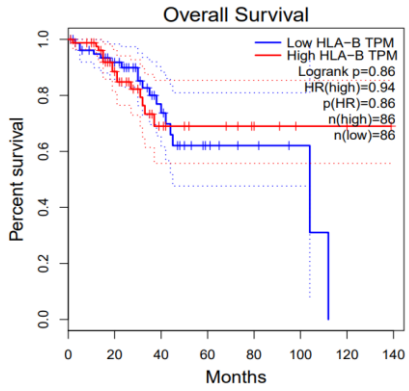
PRBD



SARC

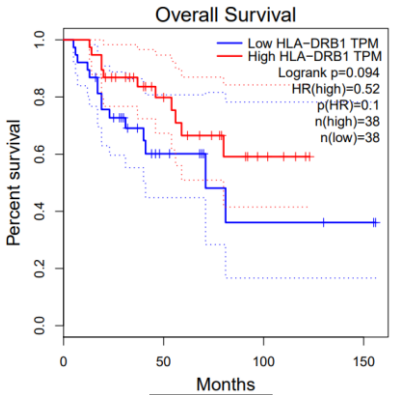


THCA

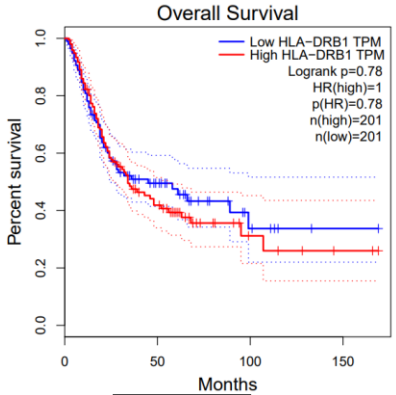


UCEC

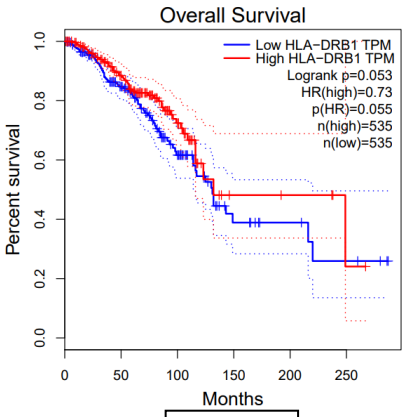
3. HLA-DRB1



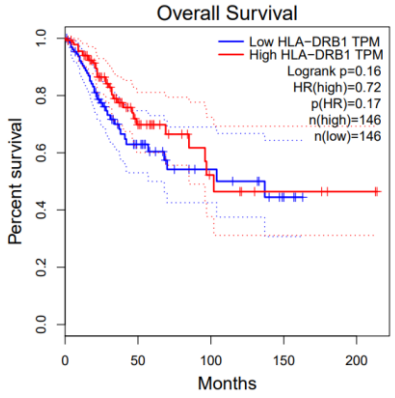
ACC



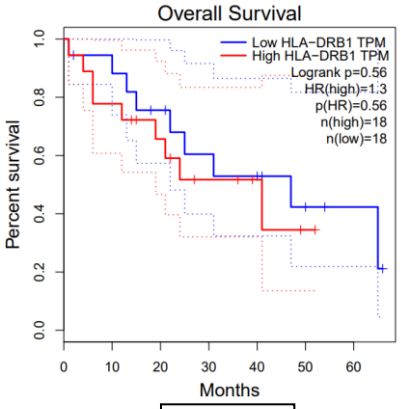
BLCA



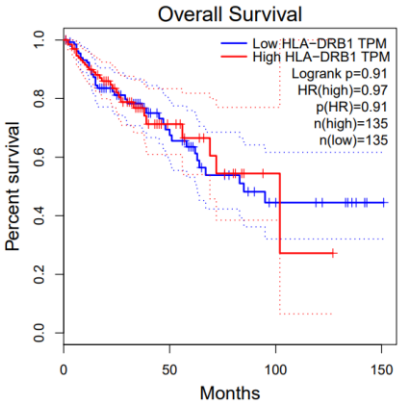
BRCA



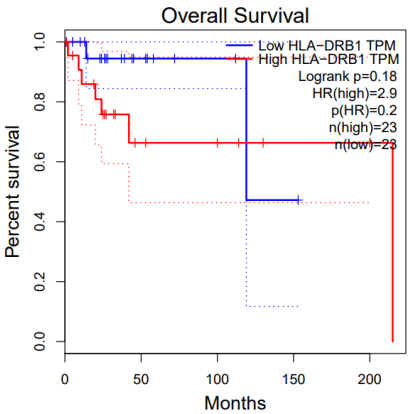
CESC



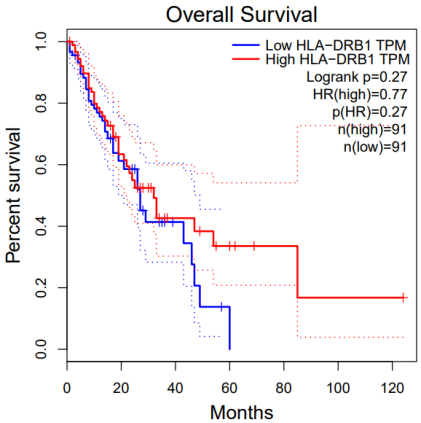
CHOL



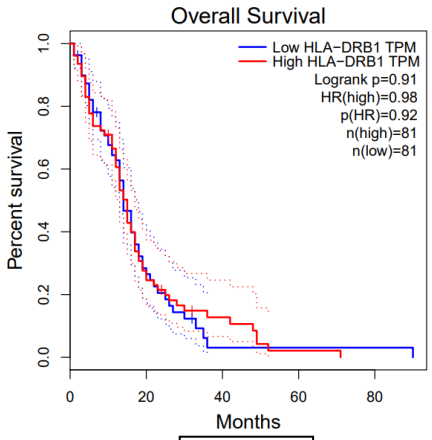
COAD



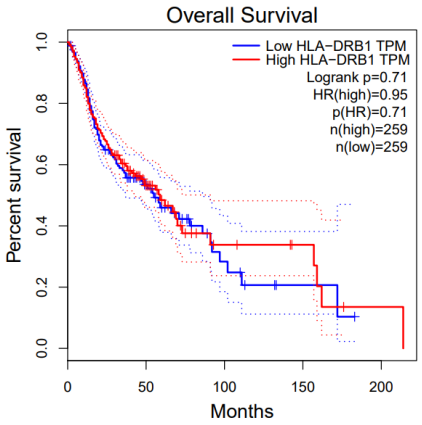
DLBC



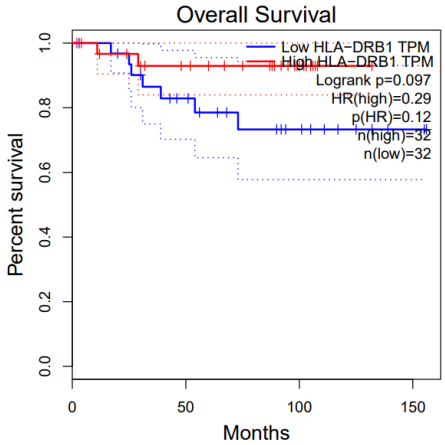
ESCA



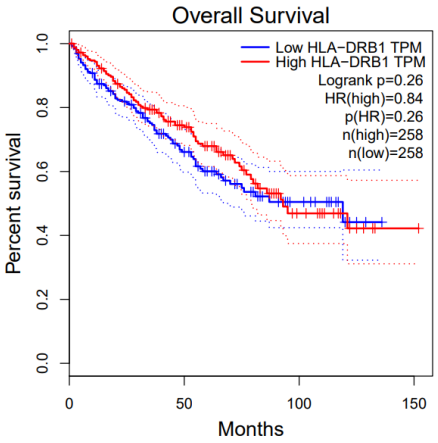
GBM



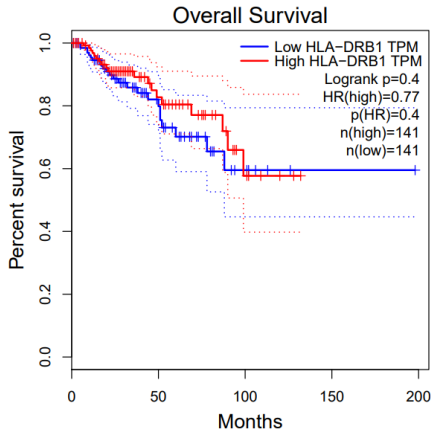
HNSC



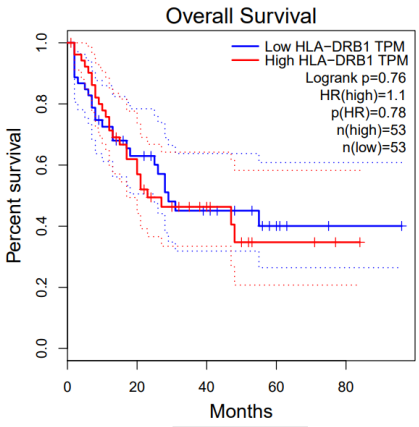
KICH



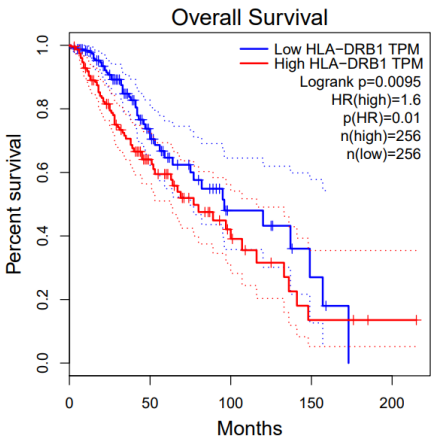
KIRC



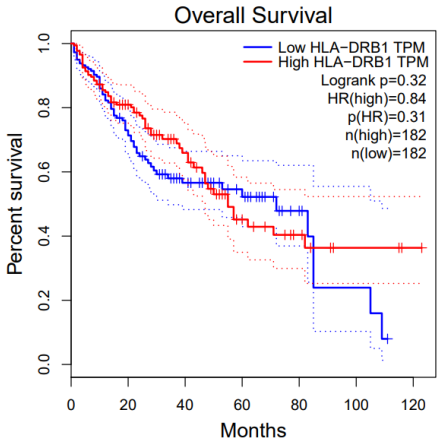
KIRP



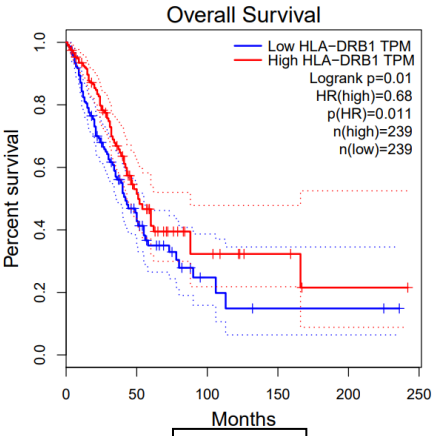
LAML



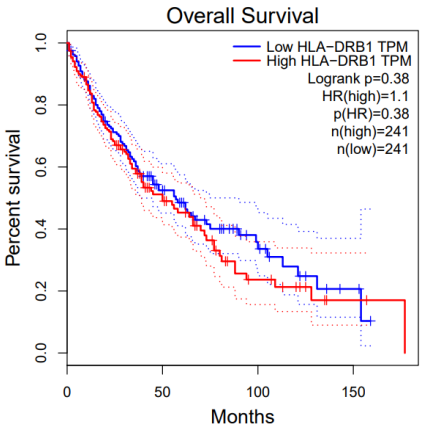
LGG



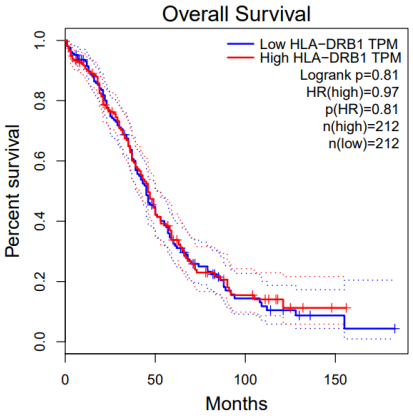
LIHC



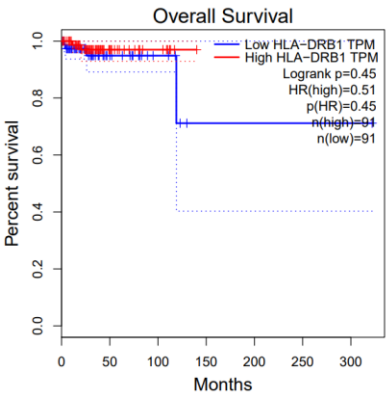
LUAD



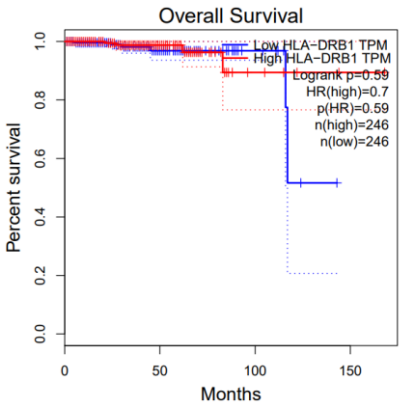
LUSC



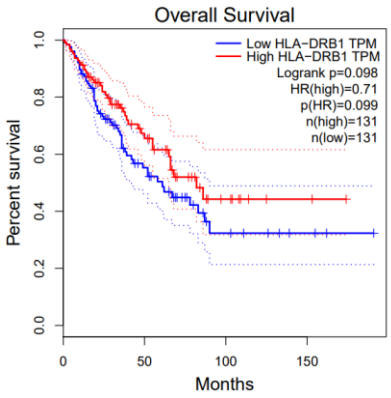
OV



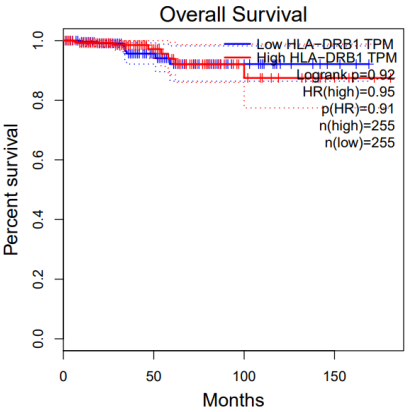
PCPG



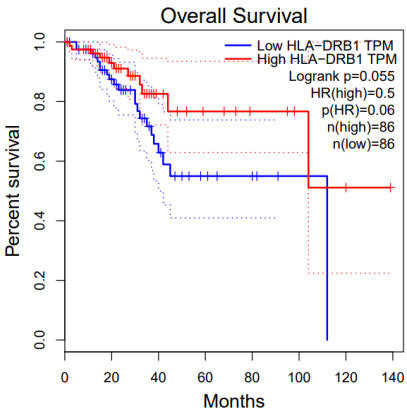
PRAD



SARC

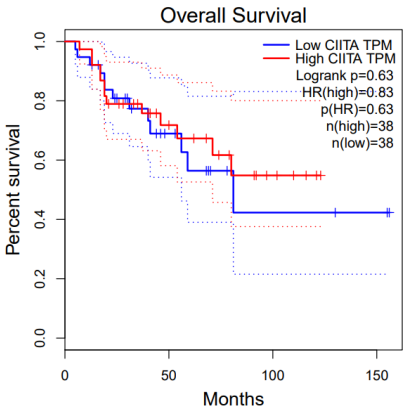


THCA

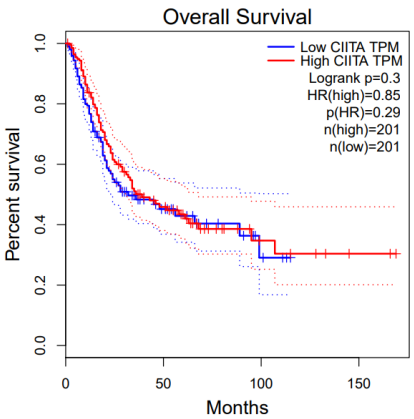


UCEC

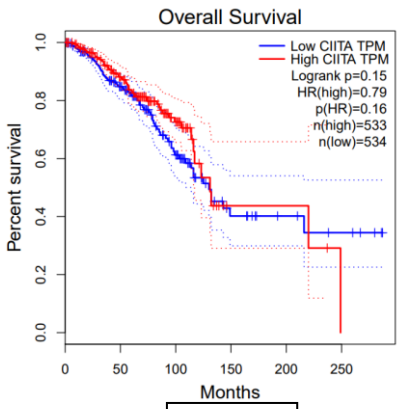
4. CIITA



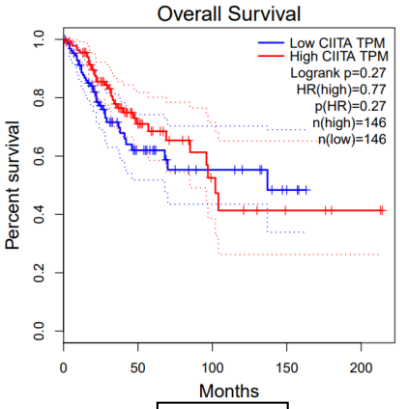
ACC



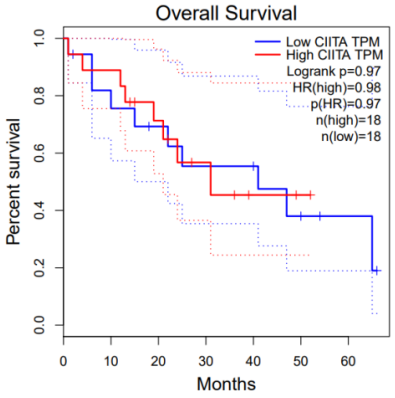
BLCA



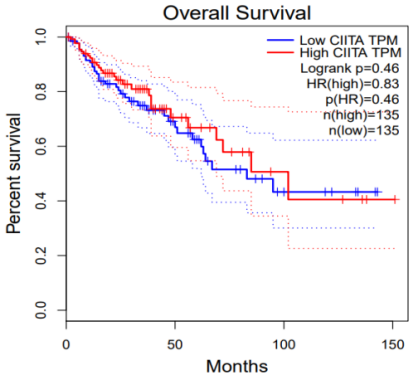
BRCA



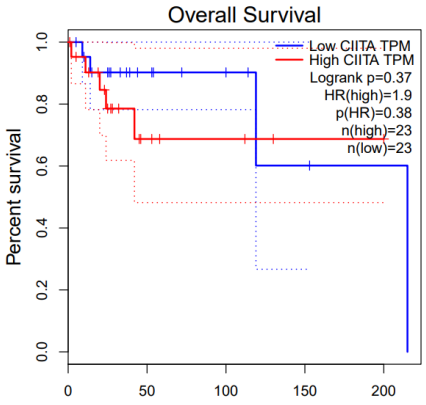
CESC



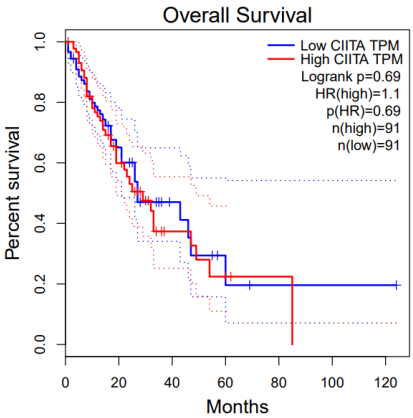
CHOL



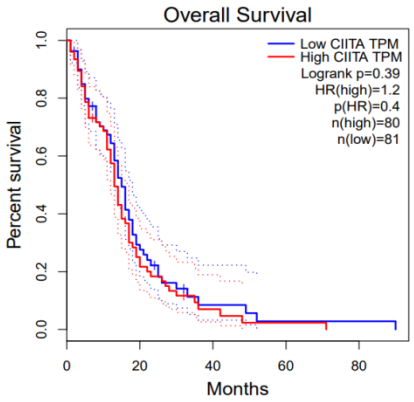
COAD



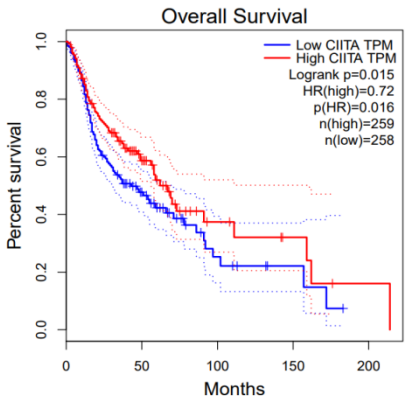
DLBC



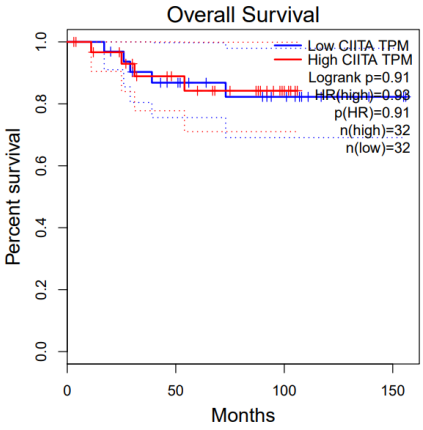
ESCA



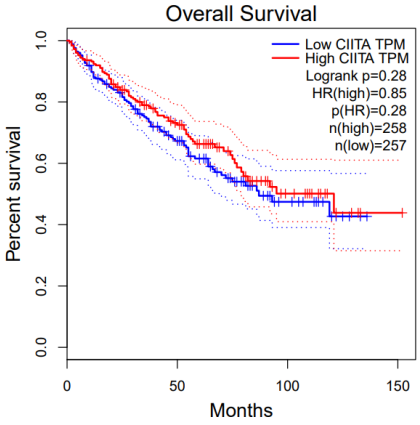
GBM



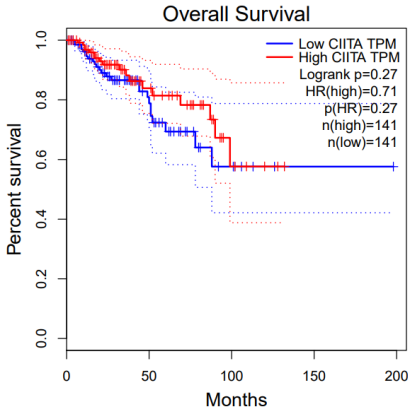
HNSC



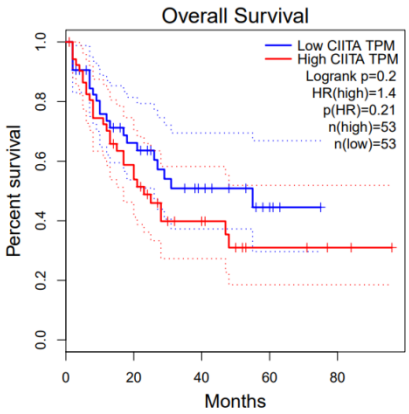
KICH



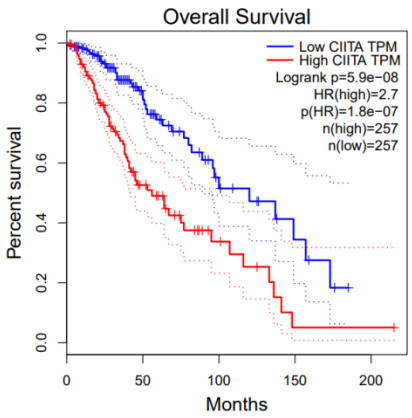
KIRC



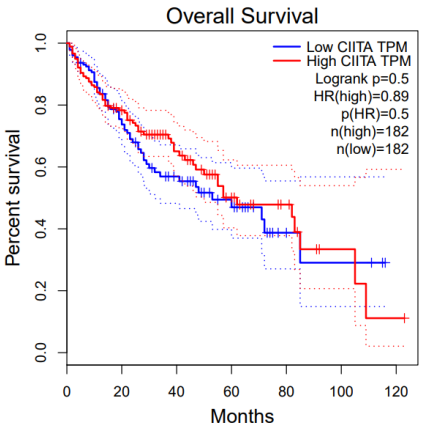
KIRP



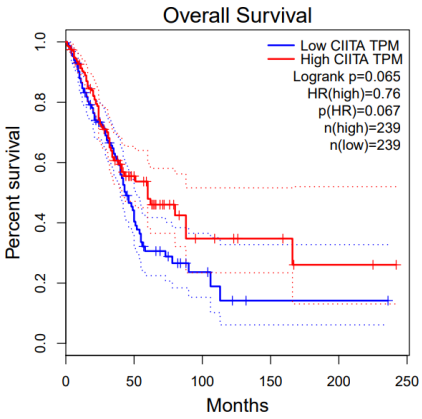
LAML



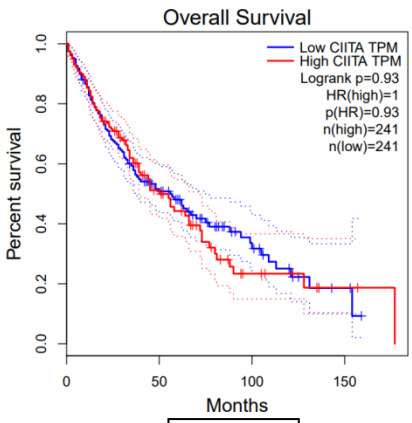
LLG



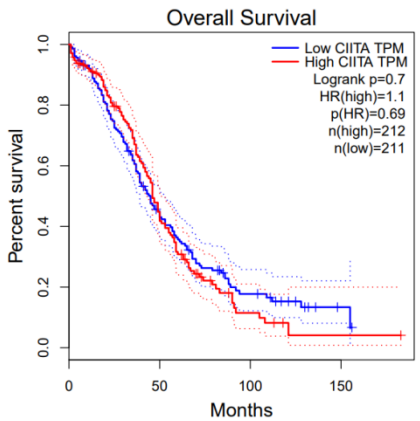
LIHC



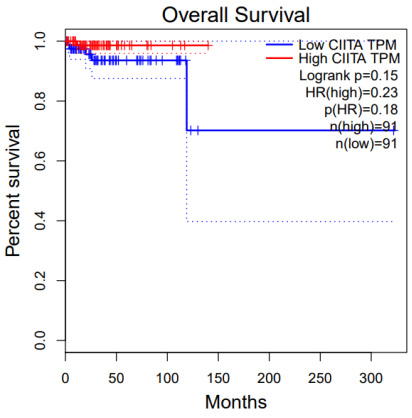
LUAD



LUSC



OV



PCPG

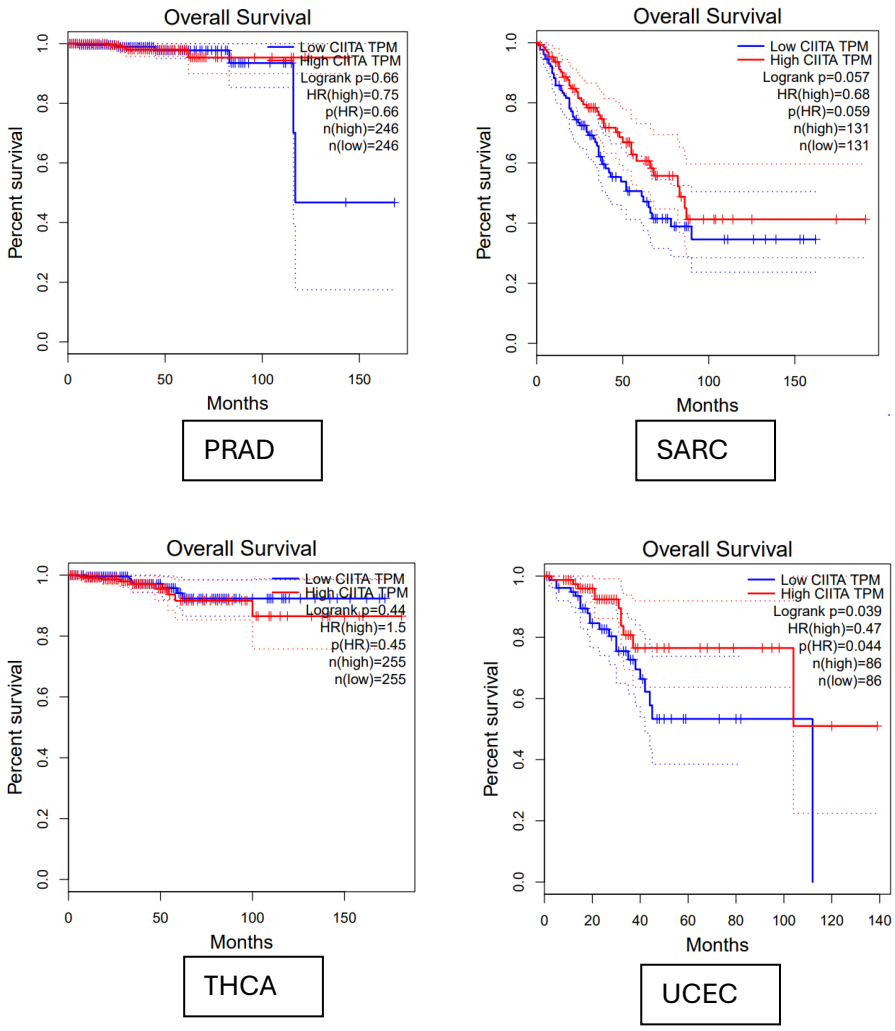


Figure 4.2.4: Kaplan Meier survival curves for 4 antigen presenting genes in our study (1.HLA-A, 2.HLA-B, 3. HLA-DRB1, 4. CIITA) considering expression. Survival curves with high expression is depicted in red while for low expression it was blue, the X-axis represent time (in months) whereas Y-axis represents percent survival.

4.2.2. Discussion:

It may be anticipated that genes under positive selection, by immunoediting will tend to confer immune escape and mutations play a major role in this escape (22). Among the immune-related genes, PIK3CA was one of the most frequently mutated genes. Mutations in PIK3CA were related to poor survival in LGG; UCEC was an exception with positive correlation between PIK3CA mutations and survival. Such antioncogenic mutations in PIK3CA were previously reported in breast cancer (128). Similarly, mutated ZFHX3 in non-small cell lung cancer (129), CTNNB1 and SMARCA4 in stomach adenocarcinoma (130) are other such examples of protective mutations, however, the exact mechanism of conferred protection remains to be elucidated. In LGG, mutant PIK3CA was associated with a decreased survival rate but did not achieve statistical significance (Figure 4.2.1). We could conclude that PIK3CA mutations has the potential to influence survival. An investigation evaluating differentially expressed genes in PIK3CA mutated versus non-mutated samples, over-represented pathways, and immune cell infiltration in the mutated samples could provide insights into how mutations can impact immune response and survival outcomes in LGG and UCEC.

Mutations in the TG (Thyroglobulin) gene were originally known to influence thyroid cancer, and it has been well established in literature. Information on influence of TG alterations in different cancers were limited. Our research has revealed that TG mutations can also have a negative impact on survival in BRCA (Figure 4.2.2). This suggests that the role of TG mutations in cancer may be more complex and have broader implications than previously understood, potentially affecting different cancers or pathways. Further investigation would be valuable to understand the specific mechanisms and implications of TG mutations in breast cancer. Thyroid carcinoma may originate from dyshormonogenic goiters formed as a result of mutations in TG (131). A study conducted in Japan has indicated a high prevalence of thyroid malignancy among patients with thyroglobulin mutations (132). Our study reveals mutation in TG can influence survival

and prognosis in BRCA. A more detailed analysis of the influence of TG mutations in BRCA could indeed provide valuable insights for disease diagnosis and therapeutic strategies. Understanding how TG mutations impact different pathways and immune cells in TIME can help us understand the contribution of TG in altered immune scenario and disease progression in BRCA.

It is important to note that in the survival analysis with the mutations in four genes related to antigen presentation in our study, none of the individual genes showed acceptable results (Table 4.2.2). Following this, survival analysis considering expression levels in cancers were performed. It is quite intriguing that all four genes related to antigen presentation in our study exhibited similar results in the survival analysis for LGG (Table 4.2.3), indicating that higher expression of these genes was associated with poorer survival outcomes. This consistency in the findings suggests that the genes may collectively play a significant role in LGG prognosis and potentially have a detrimental impact on patient survival when expressed at higher levels. Given the consistent and significant impact of these four antigen presentation genes on survival outcomes in LGG patients, it was a prudent decision to study the effect of these genes further. Altered gene expression of antigen processing and presentation is an established technique to evade immune destruction by tumors. It was observed that there lies a strong connection between gene expression signatures of effector lymphocyte and HLA mutations, supporting the concept that somatic alterations in HLA genes can serve as a viable immunological escape strategy (133). CIITA has earned the distinction of being the "master regulator" of MHC-class II expression, because the variation in the expression levels of MHC-II genes is predominantly attributed to the differences in levels of CIITA(134, 135). In diffuse large B-cell lymphoma (DLBCL) patients, low CIITA expression is a unique independent predictor of poor survival. Loss of CIITA provides an immunological escape and treatment resistance mechanism by reducing MHC class II expression(136). Levels of HLA-DRB1 could influence the prognosis of cutaneous melanoma patients by altering the TME (79). HLA-A and HLA-B expression correlates

to T cell activation and recognizes immune-activated breast tumors with a good prognosis(137). Tumors lacking HLA class I, HLA-E, and HLA-G expression are associated with better overall survival and disease-free survival(138). The better prognosis of colorectal cancer patients may be due to the elimination of HLA-negative cells by NK cells(139). It is certainly possible that alterations in these four antigen presentation genes, when considered together as a group, could collectively influence disease pathogenesis. Gene interactions and networks play complex roles in biological processes, and studying these genes as a group may provide insights into their combined effects on immune responses and disease progression. Indeed, further study into the influence of these four genes may unveil the potential mechanisms and connections through which these genes exert their influence on the disease pathogenesis.

This research on the influence of PIK3CA, TG, and four antigen-presenting genes on survival indicates their potential as therapeutic targets. A detailed examination of the influence of these genes on pathways and the immune microenvironment will provide us with essential information on the impact of their alterations. Understanding these mechanisms is essential for developing more effective treatments and strategies to manage the disease. It's a promising avenue for gaining a deeper understanding of disease biology.

4.3. Investigating the crucial role of immune related genes with genomic instability in different type of cancers

4.3.1. Results:

4.3.1.1. Over-represented pathways in PIK3CA mutated datasets

Pathway enrichment analysis of differentially expressed genes in the PIK3CA-mutated versus non-mutated dataset of UCEC (mutated=113, non-mutated=95) identified the phototransduction and complement and coagulation pathways as the significantly enriched pathways (Figure 4.3.1(a)). In LGG (mutated=36, non-mutated=322), seven pathways were over-represented. Complement and coagulation pathway plus cytokine and cytokine receptor interaction pathway, focal adhesion and ECM receptor interaction, PIK3/Akt pathway, protein digestion and absorption pathways were enriched the enriched pathways in PIK3CA mutated dataset of LGG (Figure 4.3.1(b)).

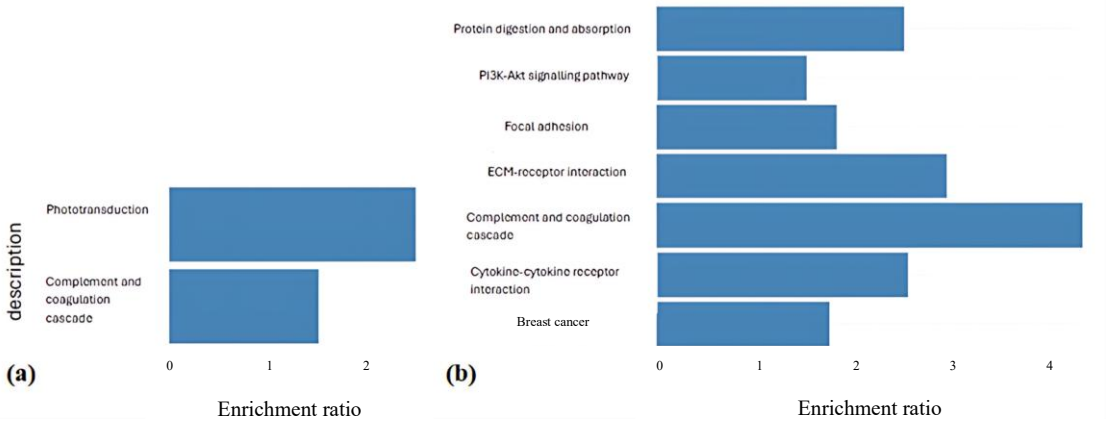


Figure 4.3.1: Pathways over-represented by differentially expressed genes in PIK3CA mutated (a) UCEC, (b) LGG dataset, Y-axis has a description of the overrepresented pathways and X-axis has the enrichment ratio obtained from the analysis (p-value>0.05)

4.3.1.2. Differentially expressed genes in complement and coagulation pathway

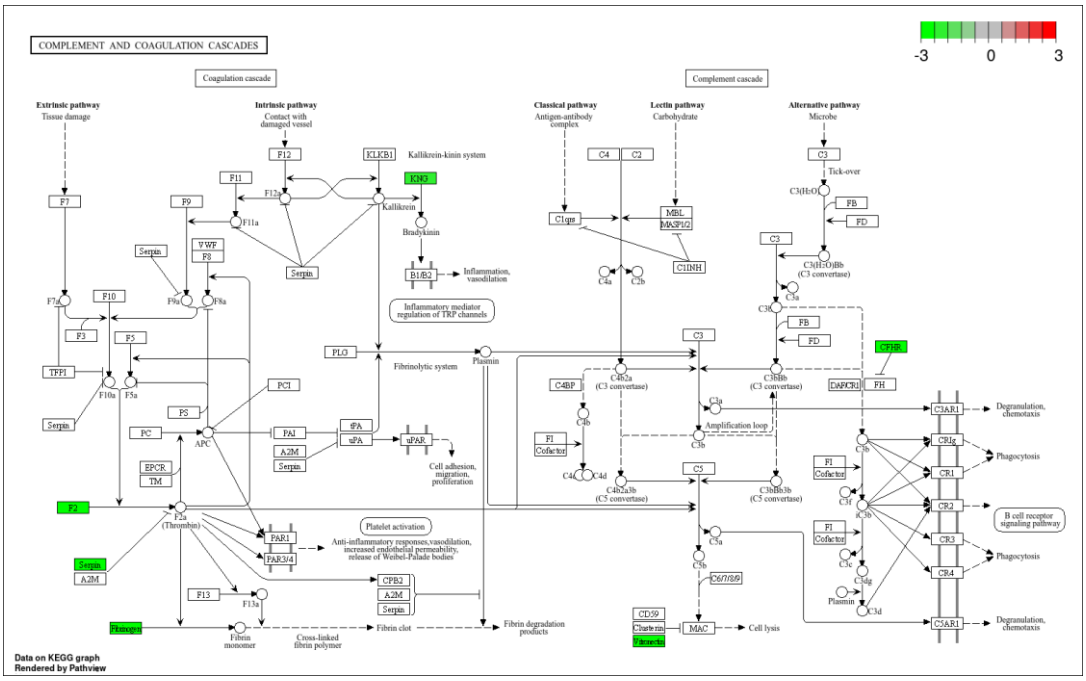
Among the over-represented pathways in the PIK3CA mutated datasets, the complement and coagulation pathway were common to both UCEC and LGG. However, the differentially expressed genes were mostly distinct, with the exception of Fibrinogen beta chain and Serpin family D member 1. Serpin family D member 1 was downregulated in UCEC but upregulated in LGG, while Fibrinogen beta chain was consistently downregulated in both cancers (Table 4.3.1). In the PIK3CA mutated dataset of UCEC, other downregulated genes of the complement and coagulation cascade included Kininogen 1, Coagulation Factor II (thrombin), Fibrinogen alpha chain, Fibrinogen gamma chain, Vitronectin, and Complement Factor H Related 5 (Table 4.3.1(A)) (Figure 4.3.2(a)). In LGG, the upregulated genes within the complement and coagulation cascade were Coagulation Factor XIII A Polypeptide, Bradykinin Receptor B2, Complement C6, Complement C7, Complement C3b/C4b Receptor 1, Complement Component 4 Binding Protein Beta, Complement Factor H, and Complement Factor H Related 3 (Table 4.3.1(B)) (Figure 4.3.2(b)).

(A) UCEC						
Sl. No	**Gene Symbol**	**Gene Name**	**EntrezID**	**logFC**	**PValue**	**FDR**
1	KNG1	Kininogen 1	3827	-2.310	1.36E-06	2.19E-04
2	F2	Coagulation factor II, thrombin	2147	-2.670	5.23E-09	2.19E-06
3	FGA	Fibrinogen alpha chain	2243	-1.695	1.33E-03	3.60E-02
4	FGB	Fibrinogen beta chain	2244	-1.653	1.08E-03	3.18E-02
5	FGG	Fibrinogen gamma chain	2266	-2.866	1.45E-04	7.74E-03
6	SERPIND1	Serpin family D member 1	3053	-2.859	1.15E-10	8.44E-08
7	VTN	Vitronectin	7448	-2.506	4.74E-13	8.08E-10
8	CFHR5	Complement Factor H Related 5	81494	-2.934	1.02E-03	3.05E-02
9	C3P1	Complement Component 3 Precursor Pseudogene	388503	-1.980	0.0019	0.044

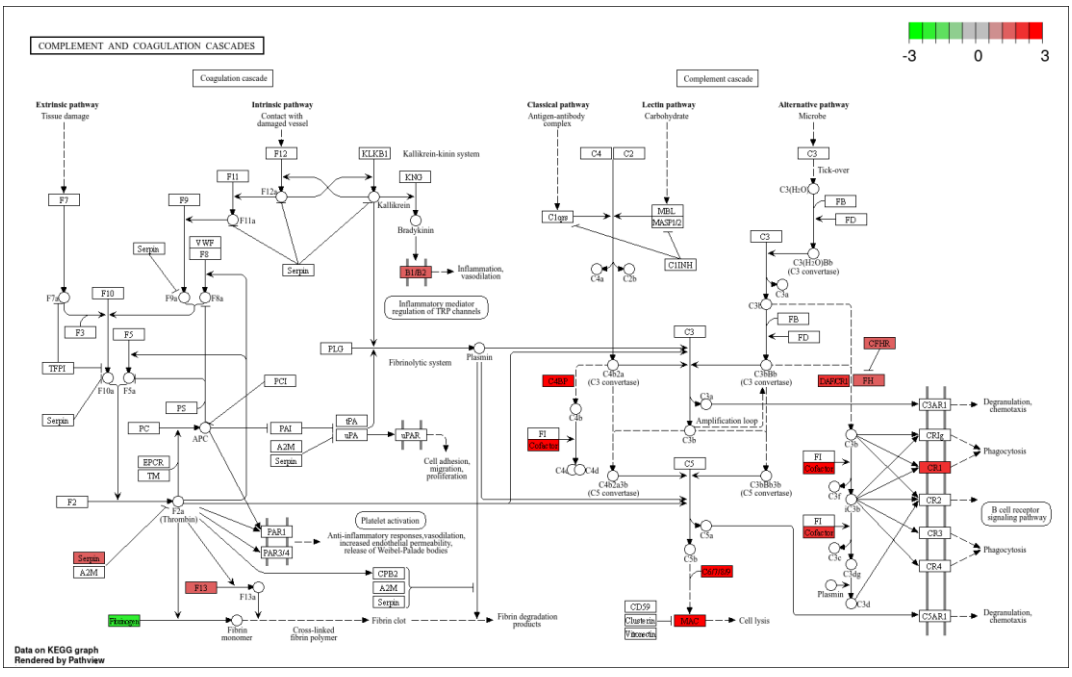
(B) LGG						
Sl. No	**Gene Symbol**	**Gene Name**	**EntrezID**	**logFC**	**PValue**	**FDR**
1	SERPIND1	Serpin family D member 1	3053	1.576	1.09E-06	7.89E-05
2	FGB	Fibrinogen beta chain	2244	-1.832	3.34E-03	3.31E-02
3	F13A1	Coagulation Factor XIII, A Polypeptide	2162	1.704	1.56E-10	3.04E-08
4	BDKRB2	Bradykinin Receptor B2	624	1.569	7.56E-11	1.53E-08
5	C6	Complement C6	729	1.688	3.06E-05	1.11E-03
6	C7	Complement C7	730	1.825	4.56E-16	2.03E-13
7	CR1	Complement C3b/C4b Receptor 1	1378	2.310	4.84E-27	6.20E-24
8	C4BPB	Complement Component 4 Binding Protein Beta	725	2.430	6.20E-17	3.26E-14
9	CFH	Complement Factor H	3075	1.775	9.99E-27	1.21E-23
10	CFHR3	Complement Factor H Related 3	10878	1.796	2.47E-11	5.45E-09

(B)LGG

Table 4.3.1: Differentially expressed genes of complement and coagulation pathway in PIK3CA mutated (A)UCEC and (B)LGG dataset.



(a)



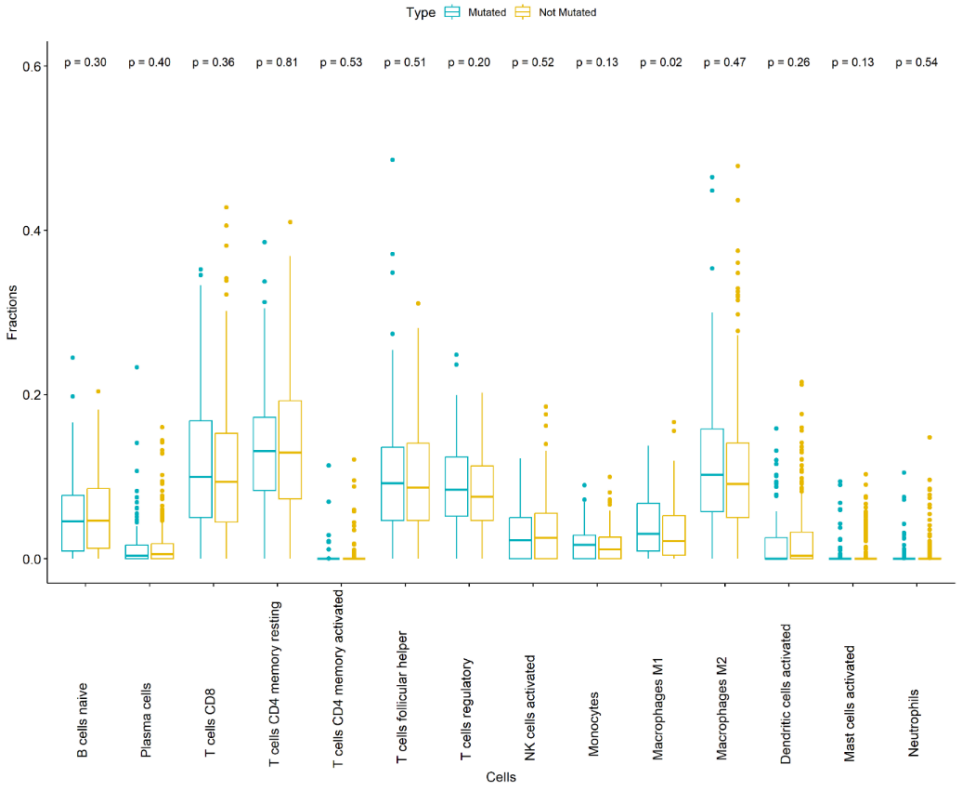
(b)

Figure 4.3.2: The complement and coagulation pathway, showing differentially expressed genes in the PIK3CA mutated dataset (a) UCEC, (b)LGG.

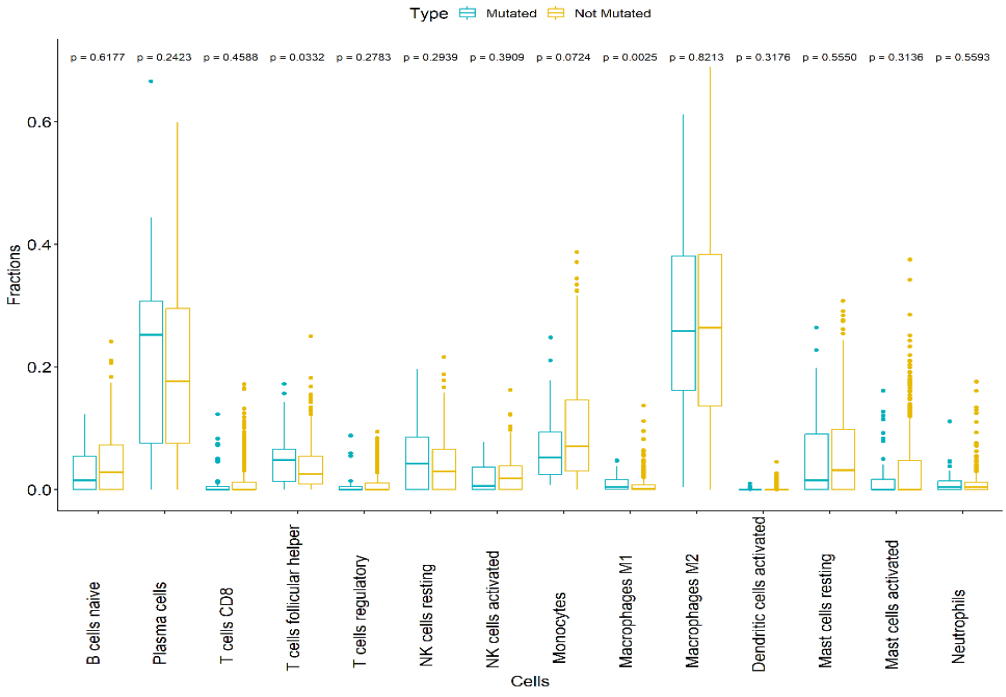
4.3.1.3. Difference in immune infiltration to tumor microenvironment in mutated versus non-mutated PIK3CA samples and their association to differentially expressed genes

The difference in recruitment of immune cells in PIK3CA mutated and non-mutated dataset was analysed. Here, student's t-test was used to determine the significance of differences between the mutated and non-mutated datasets. In UCEC (mutated=113, non-mutated=95), M1 macrophages showed a notably higher level of infiltration in datasets with mutated PIK3CA in contrast to non-mutated datasets (Figure 4.3.3(a)). In the case of mutated LGG (mutated=36, non-mutated=322), the infiltration levels of follicular helper T cells (Tfh) and M1 macrophages were found to be elevated significantly compared to non-mutated dataset (Figure 4.3.3(b)). Apart from these cells, plasma cells in LGG also had high infiltration to the mutated dataset but could not reach significance.

The correlation between immune cells with difference in infiltration levels in PIK3CA mutated versus nonmutated dataset and differentially expressed genes were analysed in both the cancers. In UCEC, Kininogen 1(KNG1) and Complement Component 3 Precursor Pseudogene (CFHR3) from the complement cascade were downregulated. Both genes KNG1 (0.20582, p-value 0.042) and CFHR3 (0.229872, p-value 0.022) positively correlate with M1 macrophage infiltration into the tumor immune microenvironment. In LGG, Tfh positively correlate with M1 macrophage infiltration (0.333439, p-value 0.022). Among the differentially expressed genes, Endothelin 2 (EDN2), Chondrolectin (CHODL), and Neurotensin Receptor 1 (NTSR1) positively correlate with both Tfh and M1 macrophages. In the PIK3CA mutated dataset of LGG, EDN2 was downregulated, whereas CHODL and NTSR1 were upregulated (Table 4.3.2).


Type
Mutated
Not Mutated
p = 0.30
p = 0.40
p = 0.36
p = 0.81
p = 0.53
p = 0.51
p = 0.20
p = 0.52
p = 0.13
p = 0.02
p = 0.47
p = 0.26
p = 0.13
p = 0.54
Fractions
0.0
0.2
0.4
0.6
B cells naive
Plasma cells
T cells CD8
T cells CD4 memory resting
T cells CD4 memory activated
T cells follicular helper
T cells regulatory
NK cells activated
Monocytes
Macrophages M1
Macrophages M2
Dendritic cells activated
Mast cells activated
Neutrophils
Cells

(a)


Type
Mutated
Not Mutated
p = 0.6177
p = 0.2423
p = 0.4588
p = 0.0332
p = 0.2783
p = 0.2939
p = 0.3909
p = 0.0724
p = 0.0025
p = 0.8213
p = 0.3176
p = 0.5550
p = 0.3136
p = 0.5593
Fractions
0.0
0.2
0.4
0.6
B cells naive
Plasma cells
T cells CD8
T cells follicular helper
T cells regulatory
NK cells resting
NK cells activated
Monocytes
Macrophages M1
Macrophages M2
Dendritic cells activated
Mast cells resting
Mast cells activated
Neutrophils
Cells

(b)

Figure 4.3.3: Difference in fractions of immune cell infiltration in mutated and non-mutated PIK3CA dataset of **(a)** UCEC (mutated=113, non-mutated=95), **(b)** LGG (mutated=36, non-mutated=322), p-value for t-test to the top of the graph X-axis represents the type of immune cells and Y-axis represent fractions of cells infiltrating the TIME.

Gene Symbol	**Gene name**	**logFC**	**p-value**	**FDR**	**Tfh**	**p-value**	**M1**	**p-value**
EDN2	Endothelin 2	-2.113	2.38E-03	2.65E-02	0.56009	4.24E-05	0.336138	0.0208
CHODL	Chondrolectin	1.594	5.18E-11	1.12E-08	0.64233	1.13E-06	0.533427	0.0001
NTSR1	Neurotensin Receptor 1	1.528	2.64E-08	3.43E-06	0.67435	2.03E-07	0.448124	0.0015

Table 4.3.2: Correlation between differentially expressed genes and immune cells with difference in infiltration levels in PIK3CA mutated versus nonmutated LGG (mutated=36, non-mutated=322) dataset i.e. follicular-helper T cells and M1 macrophages.

4.3.1.4. Over-represented pathways in TG mutated datasets

Pathway enrichment analysis of differentially expressed genes in the TG-mutated versus non-mutated dataset of BRCA (mutated=12, non-mutated=615) identified the protein digestion and absorption, adrenergic signalling in cardiomyocytes, nicotine addiction, neuroactive ligand-receptor interaction, salivary secretion, dilated cardiomyopathy, taste transduction, hypertrophic cardiomyopathy aldosterone synthesis and secretion and pancreatic secretion as the significantly enriched pathways (Table 4.3.3).

Sl.No	Pathways	Enrichment Ratio	P-value
1.	Protein digestion and absorption	3.3411	1.0218e-5
2.	Adrenergic signaling in cardiomyocytes	2.7604	2.0701e-5
3.	Dilated cardiomyopathy	3.1446	4.039e-5
4.	Hypertrophic cardiomyopathy	3.1304	7.393e-5
5.	Nicotine addiction	4.554	1.1197e-4
6.	Aldosterone synthesis and secretion	2.8919	3.0155e-4
7.	Neuroactive ligand-receptor interaction	1.8202	5.8718e-4
8.	Salivary secretion	2.8297	6.087e-4
9.	Drug metabolism	3.0927	7.419e-4
10.	Taste transduction	2.8246	9.9423e-4

Table 4.3.3: Pathways over-represented by differentially expressed genes of TG mutated BRCA dataset

4.3.1.5. Difference in immune infiltration to tumor microenvironment in mutated versus non-mutated TG samples and their association to differentially expressed genes

The difference in immune cell recruitment between TG-mutated and non-mutated datasets were investigated. Here, Student's t-test was used to determine the significance of differences between the mutated and non-mutated datasets. In BRCA (mutated=12, non-mutated=615), M1 macrophages showed a notably higher level of infiltration in datasets with mutated TG as compared to non-mutated datasets (Figure 4.3.4). The correlation analysis between immune cells with differences in infiltration levels in the TG mutant versus nonmutated datasets and differentially expressed genes showed M1 macrophages correlate positively to Indoleamine 2,3-dioxygenase 1(IDO1) and Angiotensin Converting Enzyme 2(ACE2). ACE2 was downregulated whereas IDO1 was upregulated in TG mutated BRCA dataset (Table 4.3.4).

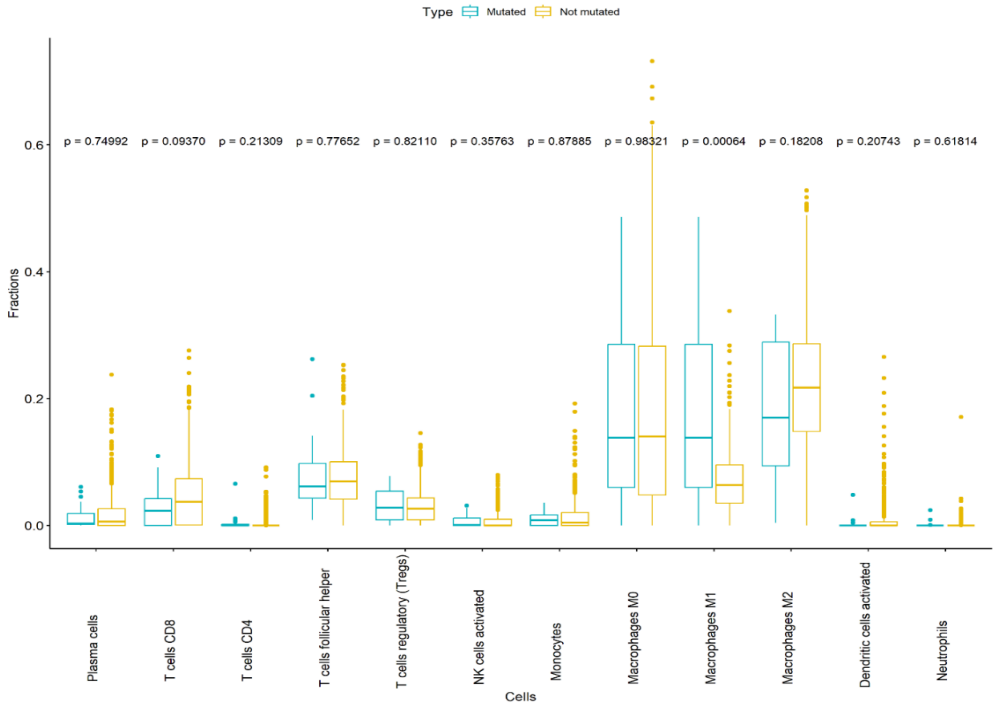


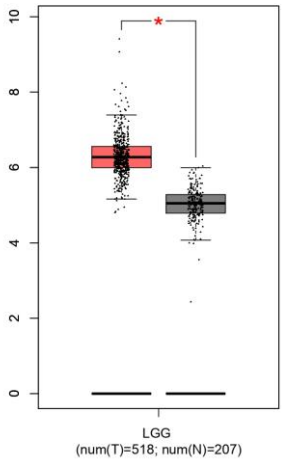
Figure 4.3.4: Difference in fractions of immune cell infiltration in mutated and non-mutated TG dataset of BRCA, p-value for t-test to the top of the graph X-axis represents the cells and Y-axis represent fractions of cells infiltrating the TIME.

Sl. No	**Gene Symbol**	**Gene name**	**logFC**	**PValue**	**M1(Rho)**	**p-value**
1	ACE2	Angiotensin Converting Enzyme 2	-3.185	0.00069	0.5014	0.0205
2	IDO1	Indoleamine 2,3-Dioxygenase 1	1.79	4.40E-06	0.6304	0.0021

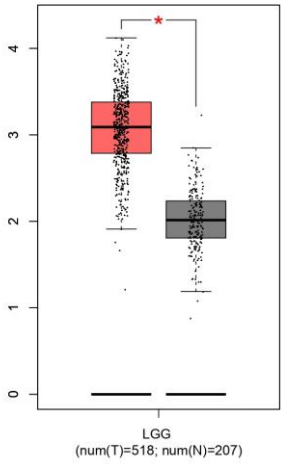
Table 4.3.4: Correlation between differentially expressed genes and immune cell with difference in infiltration levels in TG mutated versus nonmutated BRCA dataset i.e. M1 macrophages.

4.3.1.6. Differential expression of CIITA, RFX (RFXANK, RFXAP, RFX5), CREB, NFY (NFYA, NFYB, NFYC), HLA-A, HLA-B, HLADRB1, HLA-E and HLA-G in LGG

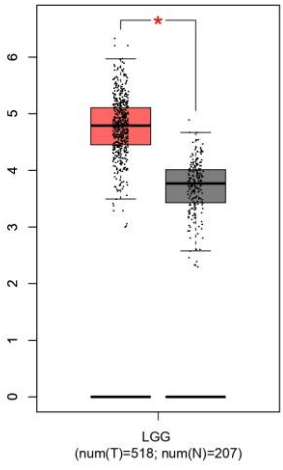
Differential expression analysis of CIITA, RFX (RFXANK, RFXAP, RFX5), CREB, NFY (NFYA, NFYB, NFYC), HLA-A, HLA-B, HLADRB1, HLA-E and HLA-G in LGG revealed significant upregulation of several genes, including RFXANK, RFXAP, RFX5(Figure 4.3.6(a(i,ii,iii))), CREB(Figure 4.3.6(c)), HLA-A(Figure 4.3.6(f)), HLA-B(Figure 4.3.6(g)), HLADRB1(Figure 4.3.6(h)) and HLA-E(Figure 4.3.6(d)). While CIITA (Figure 4.3.6(i)) and the NFY family (NFYA, NFYB, NFYC) (Figure 4.3.6(b(i,ii,iii))) were upregulated, they did not achieve statistical significance.



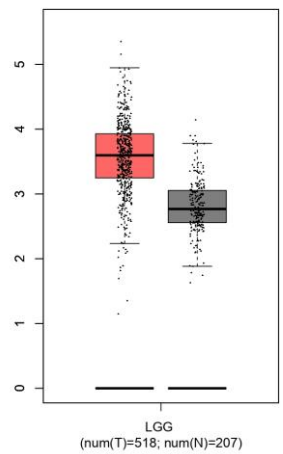
(a(i))



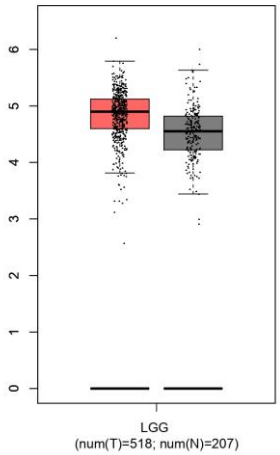
(a(ii))



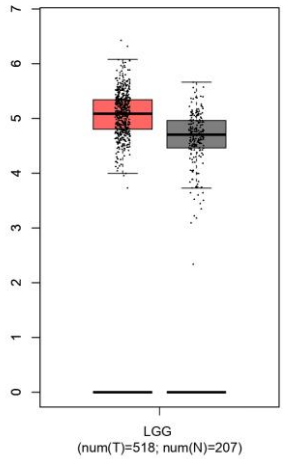
(a(iii)



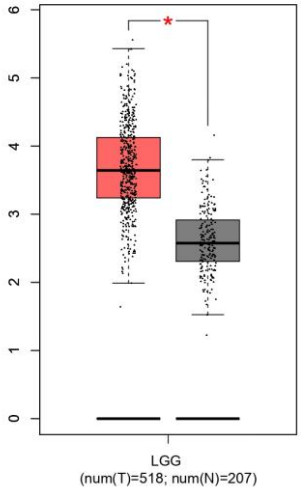
(b(i))



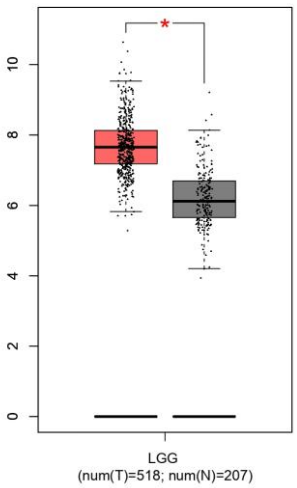
(b(ii))



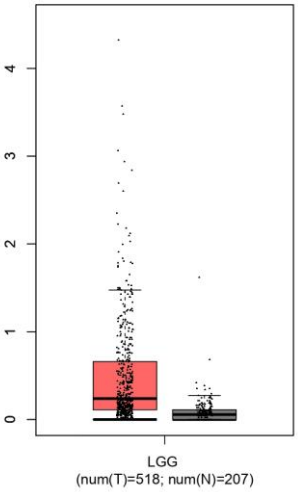
(b(iii))



(c)



(d)



(e)

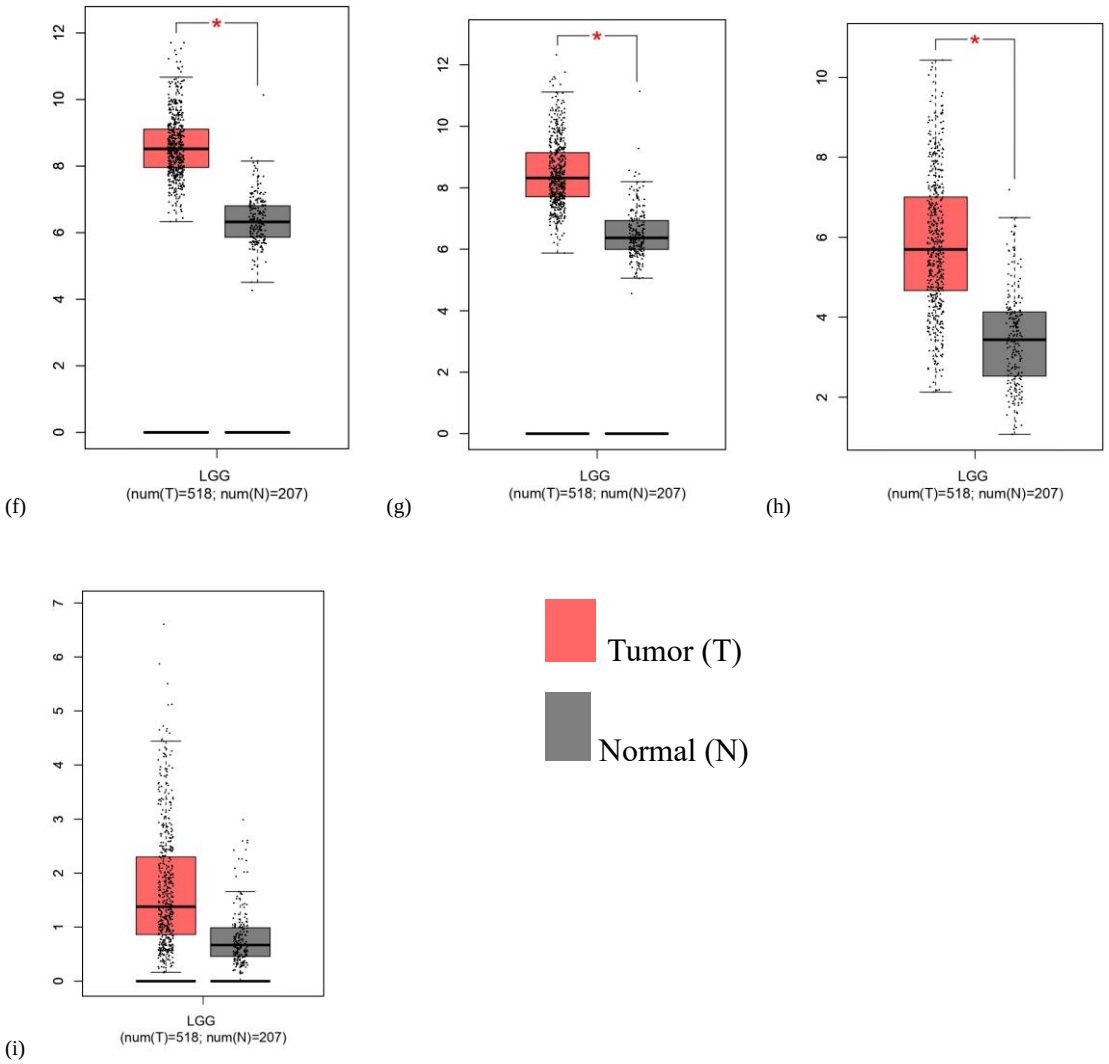
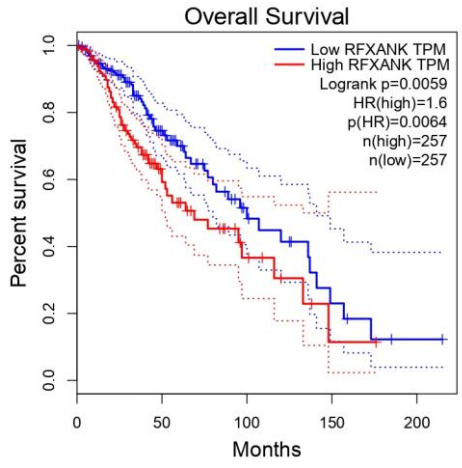


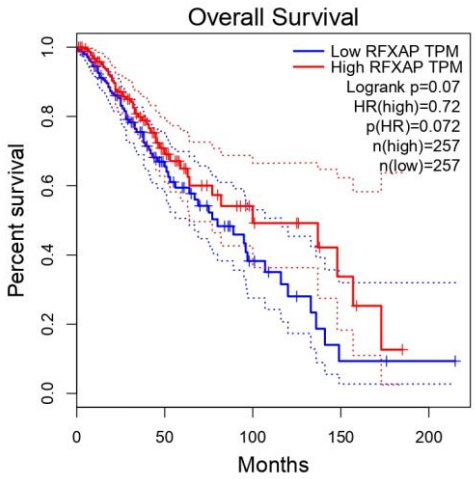
Figure 4.3.5: Differential gene expression of (a)RFX ((i)RFXANK, (ii)RFXAP, (iii)RFX5), (b) NFY((i)NFYA, (ii)NFYB, (iii)NFYC), (c)CREB, (d)HLA-E and (e)HLA-G, (f)HLA-A, (g)HLA-B, (h)HLA-DRB1(i)CIITA in tumor versus normal samples of LGG. Here, num (T) and num (N) signifies number of tumor samples and number of normal samples respectively. Y-axis represents fold change in expression levels of genes mentioned in mutated versus non-mutated samples of LGG.

4.3.1.7. Influence of RFX (RFXANK, RFXAP, RFX5), CREB, NFY (NFYA, NFYB, and NFYC), HLA-E and HLA-G expression on survival in LGG

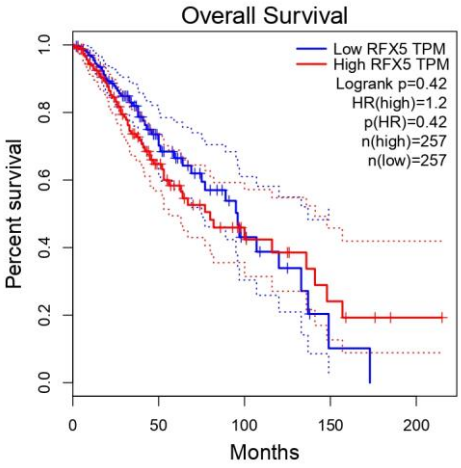
Previously, our study demonstrated that higher expression of CIITA, HLA-A, HLA-B, and HLA-DRB1 (Table 4.2.3) was correlated to poor survival in LGG patients. In alignment with the findings, we observed that higher expression levels of CREB (Figure 4.3.7(c)), HLA-E (Figure 4.3.7(d)), HLA-G(Figure 4.3.7(e)), NFYC(Figure 4.3.7(b(iii)), RFXANK(Figure 4.3.7(a(i)), and RFX5(Figure 4.3.7(a(iii)) were also linked to poor survival outcomes. However, it is important to note that RFXAP (Figure 4.3.7(a(ii)), NFYA (Figure 4.3.7(b(i)) and NFYB(Figure 4.3.7(b(ii)) could not match the above trend and did not reach statistical significance (Figure 4.3.7).



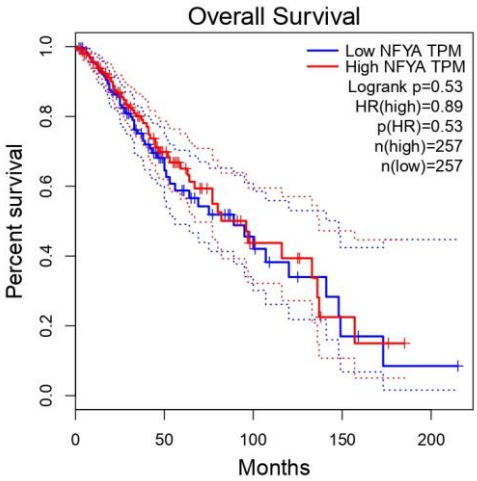
(a(i))



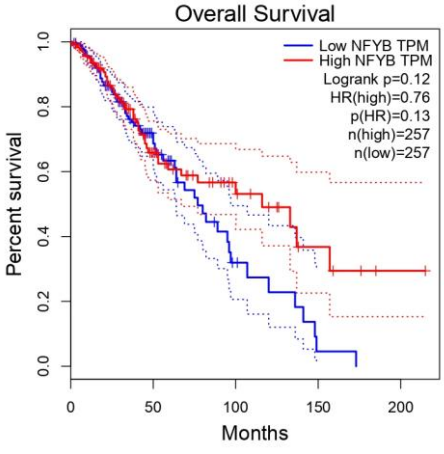
(a(ii))



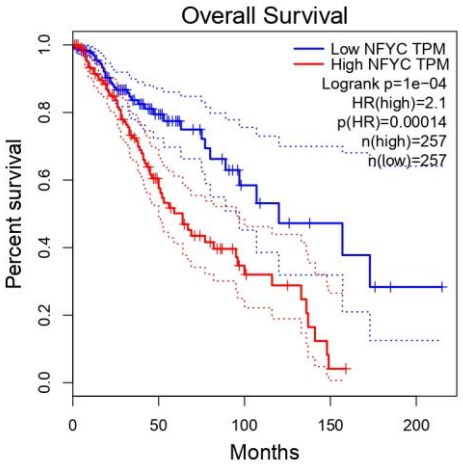
(a(iii))



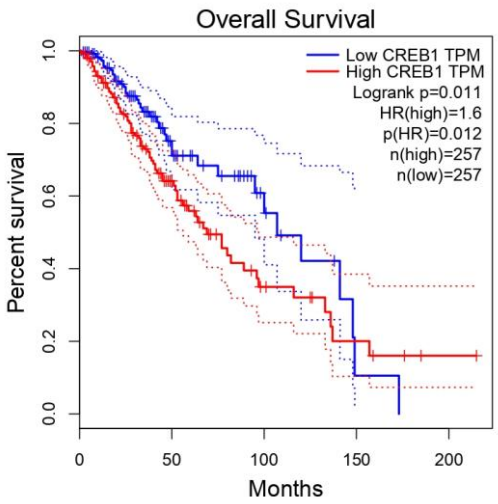
(b(i))



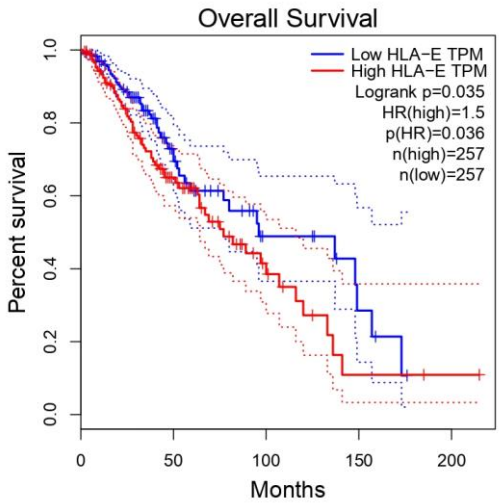
(b(ii))



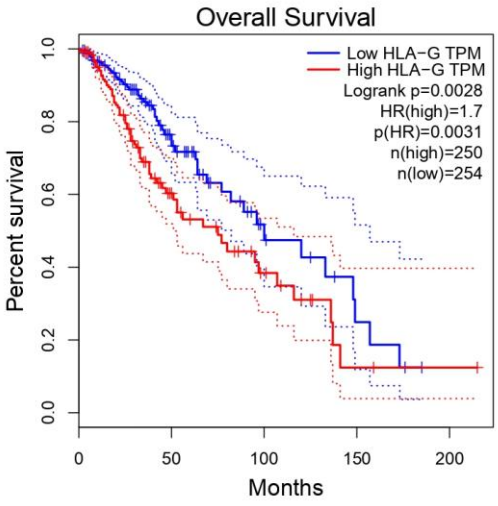
(b(iii))



(c)



(d)



(e)

Figure 4.3.6: Kaplan Meier survival curves for (a)RFX ((i)RFXANK, (ii)RFXAP, (iii)RFX5), (b) NFY((i)NFYA, (ii)NFYB, and (iii)NFYC), (c)CREB1, (d)HLA-E and (e)HLA-G. Survival curves for higher expression is depicted in red while for lower expression it is blue, the X-axis represent time (in months) whereas Y-axis represents percent survival. The dataset was divided using median.

4.3.1.8. Correlation between 4genes in our study (HLA-A, HLA-B, HLA-DRB1 and CIITA) and infiltrating immune cells

The correlation between four genes in our study (HLA-A, HLA-B, HLA-DRB1, and CIITA) and TME infiltrating immune cells was examined. In LGG, the study observed a consistent and strong positive association between the four antigen-presenting genes in our study and M2 macrophages and exhibited a negative correlation with CD4+ infiltration (Table 4.3.5).

Sl. No		(a) HLA-A		(b) HLA-B		(c) HLA-DRB1		(d) CIITA	
	CELL	Rho	P-value	Rho	P-value	Rho	P-value	Rho	P-value
1	CD8+	0.273	1.35E-09	0.247	4.32E-08	0.252	2.38E-08	0.184	5.21E-05
2	CD4+	-0.353	1.91E-15	-0.344	9.79E-15	-0.438	7.03E-24	-0.339	2.71E-14
3	Tregs	0.255	1.65E-09	0.239	1.26E-07	0.264	4.56E-09	0.183	5.57E-05
4	B cell	0.16	3.74E-04	-0.334	6.79E-14	-0.457	4.25E-26	-0.365	1.8E-16
5	M0	0.44	0.33	0.059	1.98E-01	-0.032	4.81E-01	-0.057	2.11E-01
6	M1	0.371	4.52E-17	0.396	2.10E-19	0.328	1.78E-13	0.333	7.41E-14
7	M2	0.67	1.06E-64	0.692	2.72E-69	0.781	1.95E-99	0.751	5.45E-88
8	DC	0.133	3.7E-03	0.13	4.57E-03	0.121	8.26E-03	0.122	7.54E-03
9	NK	0.2	1.06E-05	0.229	4.33E-07	0.255	1.54E-08	0.285	2.13E-10

Table 4.3.5: Association between expression of the four antigen presenting genes in our study (a) HLA-A, (b) HLA-B, (c) HLA-DRB1, (d) CIITA and immune cell infiltration to the TME in LGG.

4.3.1.9. Correlation between 4genes (HLA-A, HLA-B, HLA-DRB1 and CIITA) and HLA-E

The correlation between four genes (HLA-A, HLA-B, HLA-DRB1, and CIITA) in our study and HLA-E was examined. Here, HLA-E showed a strong positive correlation to HLA-A, HLA-B, HLA-DRB1 and CIITA in LGG dataset (Table 4.3.6).

		HLA-A	HLA-B	HLA-DRB1	CIITA
HLA-E	Rho	0.698	0.795	0.756	0.692
	p-value	7.95e-77	4.90e-114	7.75e-97	6.21e-75

Table 4.3.6: Correlation between 4genes in our study (HLA-A, HLA-B, HLA-DRB1 and CIITA) and HLA-E

4.3.1.10. Correlation among CIITA, RFXANK, RFXAP, RFX5, CREB, HLA-A, HLA-B, HLA-DRB1 and HLA-E

The correlation analysis among the genes CIITA, RFXANK, RFXAP, RFX5, CREB, HLA-A, HLA-B, HLA-DRB1, and HLA-E showed that HLA-A, HLA-B, HLA-DRB1, HLA-E, CIITA and RFXANK correlated positively to each other. RFXANK showed negative correlation to RFXAP and RFX5. RFXAP correlates negatively to HLA-A, HLA-B, HLA-DRB1, RFXAP and CIITA. CREB showed negative correlation to HLA-A, HLA-B, HLA-DRB1 and CIITA and positive corelation to RFXAP and RFX5 (Table 4.3.7).

	CIITA	RFXANK	RFXAP	RFX5	CREB	HLA-A	HLA-B	HLA-DRB1	HLA-E
CIITA	1	0.153	0.282	0.165	0.08	0.613	0.691	0.846	0.69
RFXANK	0.153	1	-0.214	-0.261	0.08	0.387	0.34	0.228	0.227
RFXAP	0.282	-0.214	1	0.347	0.461	-0.314	-0.333	-0.387	-0.3
RFX5	0.165	-0.261	0.347	1	0.594	0.01	-0.008	0.04	-0.09
CREB	0.08	0.08	0.461	0.594	1	-0.129	-0.148	-0.1	-0.15
HLA-A	0.613	0.387	-0.314	0.01	-0.129	1	0.9	0.68	0.69
HLA-B	0.691	0.34	-0.333	-0.008	-0.148	0.9	1	0.75	0.79
HLA-DRB1	0.846	0.228	-0.387	0.04	-0.1	0.68	0.75	1	0.75
HLA-E	0.69	0.227	-0.3	-0.09	-0.15	0.69	0.79	0.75	1

Table 4.3.7: Correlation among CIITA, RFXANK, RFXAP, RFX5, CREB, HLA-A, HLA-B, HLA-DRB1 and HLA-E.

4.3.2. Discussion:

In UCEC, our data suggest protection conferred by the PIK3CA mutation may operate through complement and coagulation pathways (Figure 4.3.1) and increased infiltration of M1 macrophage cells (Figure 4.3.3(a)). M1 macrophages act as key allies in the fight against cancer. They stimulate the immune system by presenting antigens and activating Th1 cells, while also secreting pro-inflammatory factors that promote inflammation and directly kill tumor cells. The complement system possesses the ability to directly lyse tumor cells, exerts pro-inflammatory effects, and can induce opsonization, whereby complement fragments such as C3b coat tumor cells. This coating facilitates the recognition and phagocytosis of tumor cells by immune cells. In cancer, receptors for anaphylatoxins, such as C5aR, trigger prosurvival and antiapoptotic responses. For instance, the binding of C5a to C5aR enhances cell proliferation in both colon cancer cell lines and endothelial cells. Similarly, the activation of C3aR is crucial in guiding collective cell migration and facilitating epithelial-mesenchymal transition, both of which are key mechanisms in metastasis. At sublytic densities, MAC (membrane attack complex) accumulation on the cell membrane promotes cell proliferation and differentiation, inhibits apoptosis, and protects cells against complement-mediated lysis. Cancer cells exploit this duality of the complement system benefiting from the initial stages but actively inhibiting full MAC assembly(140). In UCEC, the complement regulatory gene Complement Factor H Related 5 was downregulated. Similarly, another component of the complement system, Complement Component 3 Precursor Pseudogene, was also downregulated (Table 4.3.1). The gene is predicted to enable endopeptidase inhibitor activity and negatively regulate endopeptidase activity. It indicates existence of a robust and functional complement system that can facilitate the destruction of cancer cells through MAC formation. Kininogen 1, a crucial component of the kallikrein-kinin system, coagulation, and inflammation pathways, was downregulated. Kinins can promote cell proliferation, migration, angiogenesis, and increase vascular permeability. Kininogen 1(0.205, p-value= 0.042) and Complement Component 3 Precursor

Pseudogene (0.229, p-value= 0.022) positively correlated with M1 macrophage infiltration. This suggests that PIK3CA mutations might influence the generation of an anti-tumor microenvironment in the TIME by altering the complement and coagulation pathways.

In LGG, the complement and coagulation cascade pathway were an enriched pathway in the PIK3CA mutated dataset. Complement regulatory proteins, including Complement Component 4 Binding Protein Beta, Complement Factor H, and Complement Factor H Related 3, were upregulated. This clearly indicates that cancer cells are benefitting from the early stages of complement activation and the production of anaphylatoxins while actively inhibiting the generation of MAC at higher concentrations. Infiltration of Tfh cells and M1 macrophages were significantly elevated in the PIK3CA mutated dataset of LGG compared to its non-mutated counterpart (Figure 4.3.3(b)). Both the cell types are anti-tumorigenic and pro-inflammatory in nature. Although not significant, higher infiltration of plasma cells into TME of PIK3CA mutated LGG was noted. Tfh cells do not correlate with plasma cell infiltration, indicating a disruption in their proper functioning. In LGG, plasma cells correlate with CCL1 (0.298223, p-value = 0.0417), an immunomodulatory chemokine involved in attracting Tregs, facilitating angiogenesis, and promoting resistance to therapeutic drugs. Additionally, Tfh cells and M1 macrophages correlate with each other and with three other genes: Endothelin 2 (a monocyte chemoattractant protein), Chondrolectin, and Neurotensin Receptor 1 (Table 4.3.2). Endothelin 2 has a complex role, sometimes acting as a pro-tumor and other times as an anti-tumor agent(141). Chondrolectin and Neurotensin Receptor 1 are pro-tumor genes. Chondrolectin promotes growth and invasive activity(142), while Neurotensin Receptor 1 supports progression in highly malignant pancreatic cancer cells and plays a key role in metastasis. These genes have the potential to bring changes in TME of cancers related to the digestive system(143, 144). Relation of Tfh and M1 macrophages to pro-tumorigenic genes validate the immunosuppressive nature of TIME in LGG.

Overrepresented pathways by differentially expressed genes in the dataset with TG mutations highlighted several pathways related to cellular homeostasis. However, from these results, we cannot conclusively establish a relationship between TG mutations and immune responses. Notably, M1 macrophage infiltration was significantly higher in TG-mutated samples compared to non-mutated ones (Figure 4.3.4). In TG-mutated samples, M1 macrophages were associated with Angiotensin Converting Enzyme 2 (ACE2) and Indoleamine 2,3-Dioxygenase 1 (IDO1) (Table 4.3.4). ACE2 was downregulated in TG-mutated samples and is known to suppress angiogenesis in BRCA(145). Conversely, IDO1 was upregulated, which can convert an immunogenic TIME into a tolerogenic TIME(146). IDO1 upregulation is associated with compromised and immunosuppressive macrophages in breast cancer(147) and it is highly expressed on macrophages in last stages oral carcinoma(148). This suggests that IDO1 influenced M1 macrophages might act as immunosuppressive component in TME of TG mutated BRCA. Additional studies may help us understand the mechanism behind the influence of TG mutations on survival outcomes and immune infiltration in BRCA.

The entries of the antigen processing and presentation pathway, including CIITA, RFXANK, HLA-A, HLA-B, HLA-DRB1, and HLA-E, were upregulated (Figure 4.3.6) and showed positive correlation with each other (Table 4.3.7). Our study previously demonstrated that higher expression levels of HLA-A, HLA-B, HLA-DRB1 and CIITA were associated with poor survival(84). Additionally, we found that increased expression of RFX, CREB, and HLA-E also correlated with poorer survival outcomes (Figure 4.3.7). Strong positive association of 4genes (HLA-A, HLA-B, HLA-DRB1 and CIITA) to HLA-E suggests they might function together to induce immune suppression (Table 4.3.6). HLA-E corresponds to immunological suppression as it can inhibit the proper functioning of a subset of cytotoxic T cells and NK cells via KLRC1 (Killer Cell Lectin Like Receptor C1)/NKG2A(Natural Killer Cell Receptor 2A) receptors (149). CIITA, HLA-A, HLA-B, and HLA-DRB1 were negatively associated with CD4+ T cell infiltration and had a strong positive association to M2 macrophages (Table 4.3.5). CD4+ T cells are now recognized

as anti-tumor effector cells as they can contribute to several anti-tumor immune responses like inducing senescence in cancer cells, stimulating innate immune cells against cancer(150) and they can have cytolytic effector function against tumor cells expressing MHC class II (151). M2 macrophages are involved in tissue repair, wound healing, and immune regulation. They have the potential to enhance cancer growth by supporting angiogenesis and suppressing anti-tumor immune responses by the release of several proliferation-inducing and immunosuppressive cytokines and chemokines (152). Our finding suggests that the higher expression of 4 frequently mutated genes in antigen processing and presentation category might influence generation of an immune suppressive TME in LGG with upregulation of HLA-E and promoting infiltration of M2 macrophages while inhibiting infiltration of CD4+ T cells. Understanding the complex associations might help us appreciate the complex interplay between antigen presentation and immune evasion mechanisms in LGG. This highlights the importance of alterations in the expression levels of these genes for cancer survival and immune escape. This knowledge could be pivotal in developing targeted therapies that modulate the immune microenvironment to improve patient outcomes.