Supplementary Information of

"Driver gene mutations based clustering of tumors: methods and applications"

Wensheng Zhang[1], Erik K. Flemington[2], Kun Zhang[1]

[1] *Department of Computer Science, Bioinformatics facility of Xavier RCMI Cancer Research Center, Xavier University of Louisiana, New Orleans, LA 70125, USA*

[2]*Tulane School of Medicine, Tulane Cancer Center, Tulane University, 1700 Tulane Ave, New Orleans, LA 70112*

Supplementary Files	Titles
Supplementary Text S1, page 2.	Additional description of somatic mutation datasets.
Supplementary Text S2, pages 3-4.	Legends of supplementary figure sets.
Supplementary Text S3, page 4.	Member genes in the selected prognostic signature for blade cancer.
Supplementary Text S4, page 4.	Member genes in the selected prognostic signature for liver cancer.
Supplementary Table S1, page 5	Summary of tumor samples with both somatic mutation data and clinical data.
Supplementary Table S2, page 6	File names and archives of the used clinical and somatic mutation data in the TCGA database.
Supplementary Figure Set 1, pages 7-22	Results and evaluation of 2GeneModel.
Supplementary Figure Set 2, pages 23-38	Results and evaluation of xGeneModel. Two gene pairs (2-4 genes) are considered
Supplementary Figure Set 3, pages 39-54	Results and evaluation of ccpwModel.
Supplementary Figure Set 4, pages 55-70	Results and evaluation of nbsModel.
Supplementary Figure Set 5, page 71	Prediction strength and robustness of the prognostic signature identified from the result of nbsModel for bladder cancer (BLCA)
Supplementary Figure Set 6, page 72	Prediction strength and robustness of the prognostic signature identified from the result of nbsModel for liver cancer (LIHC)

Supplementary Text S1

We downloaded the clinical and somatic data from the TCGA database (http://cancergenome.nih.gov/) on April 24, 2015. Those data, contributed by different institutes, are generated using various sequencing platforms, somatic mutation calling algorithms and computational tools. Except for ovarian carcinomas (OV), we choose one representative dataset for each cancer type according to the following criteria. First, the selected dataset contains the largest number of tumor samples (or patients). Second, if two or more datasets are of the same size, we choose the one in which the mutations are measured by the IlluminaGA DNASeq platform and are called by the latest automated system. Lastly, if the decision cannot be reached by the previous two steps, we select the dataset provided by the UCSC Genome Browser. For OV, we employ the datasets from Massachusetts Institute of Technology and Washington University in St. Louis. The basic information of the used somatic and clinical datasets is summarized in Supplementary Table S2.

Supplementary Text S2

Supplementary Figure Set 1: results of 2GeneModel. This set contains 16 figures, each of which is for one cancer type and consists of four blocks, i.e. top-left, top-middle, top-right and bottom. The tumor clusters (groups) in all the plots of a figure are consistently represented by red, green, blue and purple. **Top-left**: The genotype-based stratification (clustering) of tumors. **Top-middle**: The cluster-specific Kaplan-Meier survival curves. The p-value is calculated for the comparison between one cluster (or the aggregate of two clusters) and the aggregate of the other clusters. For example, in the plot for UCEC, the p-value is calculated by comparing Cluster-4 (C4) with the aggregate of other three clusters. **Top-right**: The association between tumor clusters and patient races. AN, BL and WH indicate Asian, black and white Americans, respectively. Beside each race ID is the corresponding number of tumor samples. **Bottom**: The mutation characteristics of individual clusters. The bar length denotes the proportion of tumors (patients) with at least one mutation in the corresponding gene.

Supplementary Figure Set 2: results of xGeneModel. This set contains 16 figures, each of which is for one cancer type and consists of four blocks, i.e. top-left, top-middle, top-right and bottom. The tumor clusters (groups) in all the plots of a figure are consistently represented by red, green, blue and purple. **Top-left**: The dendrogram from the mutation-based clustering of tumors. **Top-middle**: The cluster-specific Kaplan-Meier survival curves. The p-value is calculated for the comparison between one cluster (or the aggregate of two clusters) and the aggregate of the other clusters. For example, in the plot for BLCA, the p-value is calculated by comparing the aggregate of Cluster-1 and Cluster-2 with the aggregate of other two clusters. **Top-right**: The association between tumor clusters and patient races. AN, BL and WH indicate Asian, black and white Americans, respectively. Beside each race ID is the corresponding number of tumor samples. **Bottom**: The mutation characteristics of individual clusters. The bar length denotes the proportion of tumors (patients) with at least one mutation in the corresponding gene.

Supplementary Figure Set 3: results of ccpwModel. This set contains 16 figures, each of which is for one cancer type and consists of four blocks, i.e. top-left, top-middle, top-right and bottom. The tumor clusters (groups) in all the plots of a figure are consistently represented by red, green, blue and purple. **Top left**: The dendrogram from the mutation-based clustering of tumors. **Top middle**: The cluster-specific Kaplan-Meier survival curves. The p-value is calculated for the comparison between one cluster (or the aggregate of two clusters) and the aggregate of the other clusters. For example, in the plot for LIHC, the p-value is calculated by comparing the aggregate of Cluster-2 and Cluster-4 with the aggregate of other two clusters. **Top right**: the association between tumor clusters and patient races. AN, BL and WH indicate Asian, black and white Americans, respectively. Beside each race ID is the corresponding number of tumor samples. Bottom: The mutation characteristics of individual clusters. The bar length denotes the proportion of tumors (patients) with at least one mutation in the member genes of the corresponding cancer pathway. Among the abbreviated terms, “Trans.”, “Regu.”, “Chrom.”, “Mod.”, “Apo.”, “Dam.” and “Con.” represent “Transcription”, “Regulation”, “Chromatin”, “Modification”, “Apoptosis”, “Damage” and “Control”, respectively.

Supplementary Figure Set 4: results of nbsModel. This set contains 16 figures, each of which is for one cancer type and consists of four blocks, i.e. top-left, top-middle, top-right and bottom. The tumor clusters (groups) in all the plots of a figure are consistently represented by red, green, blue and purple. **Top-left**: The dendrogram from the mutation-based clustering of tumors. **Top-middle**: The cluster-specific Kaplan-Meier survival curves. The p-value is calculated for the comparison between one cluster (or the aggregate of two clusters) and the aggregate of the other clusters. For example, in the plot for BLCA, the p-value is

calculated by comparing the aggregate of Cluster-2 (C2) and Cluster-3 (C3) with the aggregate of other two clusters. **Top-right**: The association between tumor clusters and patient races. AN, BL and WH indicate Asian, black and white Americans, respectively. Beside each race ID is the corresponding number of tumor samples. **Bottom**: The mutation characteristics of individual clusters. The bar length denotes the proportion of tumors (patients) with at least one mutation in the corresponding gene.

Supplementary Text S3

ABCA5, ACADSB, ALDH1L1, ANKRA2, ANP32B, ANP32E, ARPC5L, ASPM, ATAD3A, AVPI1, BTG2, BUB1, C14orf142, C15orf23, C17orf91, C1orf112, C1orf135, C21orf15, C3orf26, C9orf40, CBR4, CBX3, CCNB2, CCNDBP1, CDC20, CDC25B, CDC45, CDCA3, CDCA5, CDCA8, CDK5RAP2, CDKAL1, CDKN2A, CDT1, CEACAM1, CENPA, CENPF, CEP78, CHADL, CHCHD3, CHEK1, CIZ1, CKS1B, CKS2, COBLL1, COBRA1, CPSF3, CTSL2, CXCR2, CYP3A5, DBF4, DEPDC1, DGKA, DHDPSL, DLGAP5, DNAJC9, DNM1, DNMT1, DSN1, E2F1, E2F3, EDA2R, EXOSC2, EZH2, FAH, FAM13A, FAM160B1, FAM184A, FAM64A, FAM72B, FAM72D, FANCC, FBN2, FBXO45, FCER1A, FEN1, FMNL2, FOSB, FOXM1, GARS, GGCT, GPR172B, GPR19, GPRC5C, GTPBP4, H2AFV, HDGF, HHAT, HIST1H2BJ, HIST1H4H, HIST2H4A, HMGN2, IFRD1, IFT140, INPP5D, ITM2B, KIAA0895, KIF15, KIF2C, KIF4A, KIFC1, KLHL3, KPNA2, LOC284233, LPAR6, MAD2L1, MAD2L2, MCM2, MCM7, MCTP2, MDM2, MELK, MRPL9, MSH2, MTHFD2, MYBL2, MYB, NASP, NCRNA00086, NDC80, NEK2, NETO2, NUDC, NUDT4, NUF2, NUP155, ODF2, OIP5, ORAI3, ORC6L, OXCT1, PAQR4, PARP2, PDGFD, PHF19, PHGDH, PIGB, PIK3IP1, PLK1, POLE3, POLH, PPIL5, PPM1G, PRPF38A, PSMC3IP, QSOX2, RAD18, RAD54L, RANBP1, RBM17, RBPMS, REN, RFC2, RFC4, RHBDF1, RHEBL1, RNASEH2A, RPN1, RPS27L, RTN3, RUVBL1, SAMD1, SEMA4B, SEPT3, SESN1, SKA1, SKA3, SLC16A10, SLC16A5, SLC24A1, SLC45A3, SLC48A1, SLMO1, SMAD3, SMC2, SPAG5, SPATA18, SPATA20, SPC24, SSRP1, SSU72, STMN1, STRBP, SUV39H2, TARDBP, TBCK, TMBIM4, TMEFF1, TMEM203, TNFRSF10B, TNFRSF10C, TP53INP1, TPX2, TSPAN14, TSPAN31, TTF1, TTK, TUBA1B, TUBB2A, TUBB2B, TUBB4, TUBG2, UBAC1, UBE2C, UCK2, UNC5B, USP1, VPS36, VRK1, WDR34, WDR62, WDR63, WHAMM, YARS, YKT6, ZMYND19, ZNF561

Supplementary Text S4

ADAMTS13, C19orf40, C5orf30, C5orf4, C6orf138, CBX2, CCDC123, CCT5, CCT6A, CD3EAP, CDC25B, CENPO, CETP, CLEC3B, CSTF2, CYB5D2, CYGB, CYYR1, DDB2, DNAJC6, DNASE1L3, DPH2, EMCN, ENO1, ERG, EXOC3L2, EZH2, FAF1, FAM164A, FLT4, FXYD6, G6PD, GRB2, GRHPR, GTPBP4, IFIT1, IFIT5, KBTBD11, KIAA1524, KLRF1, LRRC59, LRRC70, MADCAM1, MEIS3P1, MYCT1, NEIL3, NUP62, ORC1L, PDE2A, PHLDA3, PLCB1, PNMAL2, PRCD, PRDX1, PRPF19, PRR18, PTP4A2, RAMP3, RNF24, ROBO4, RTN3, S1PR1, SAE1, SAPS1, SGOL2, SPATS2, SRL, STIP1, TAL1, TBXA2R, TIE1, TINAGL1, TMEM204, TRAIP, TRIP13, TUBA1C, ZNF470, ZSCAN20

Supplementary Table S1

Cancer	Total samples	White	Black	Asian
BLCA	382 (233)	300 (183)	22 (13)	42 (26)
GBM	594 (285)	505 (256)	50 (17)	13 (5)
HNSC	522 (504)	447 (439)	45 (36)	11 (11)
KIRC	533 (417)	466 (390)	51 (14)	8 (7)
LUAD	521 (488)	391 (385)	52 (29)	8 (8)
LUSC	496 (178)	341 (111)	31 (9)	9 (5)
BRCA	1080 (967)	747 (698)	172 (116)	61 (57)
OV	588 (371)	498 (324)	34 (17)	20 (12)
UCEC	538 (248)	372 (193)	104 (25)	20 (13)
COAD	455 (216)	214 (177)	54 (19)	11 (7)
THCA	506 (402)	329 (263)	27 (18)	52 (39)
CESC	305 (198)	210 (142)	30 (16)	19 (19)
ESCA	174 (171)	110 (109)	2 (2)	41 (41)
KIRP	272 (168)	189 (108)	60 (43)	5 (2)
LIHC	363 (197)	175 (120)	17 (14)	159 (54)
STAD	453 (288)	288 (167)	12 (4)	89 (76)

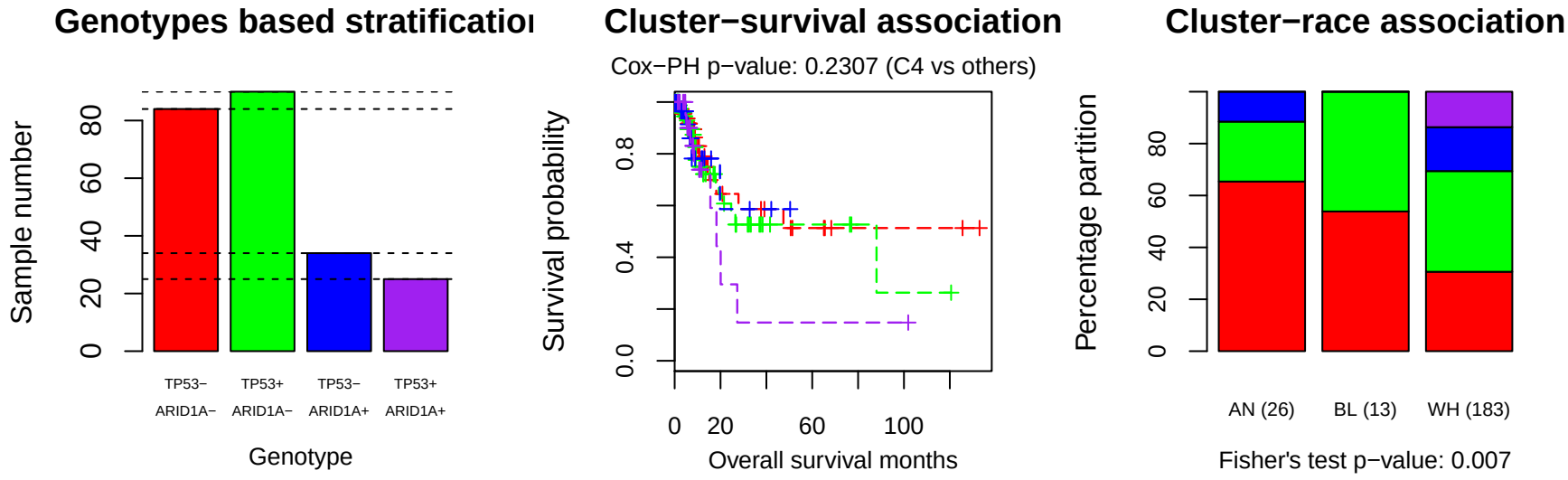
Note: Outside the brackets are the numbers of samples with clinical information only. Inside the brackets are the numbers of samples with both clinical and genomic information. Some samples do not belong to any racial group of White, Black or Asian.

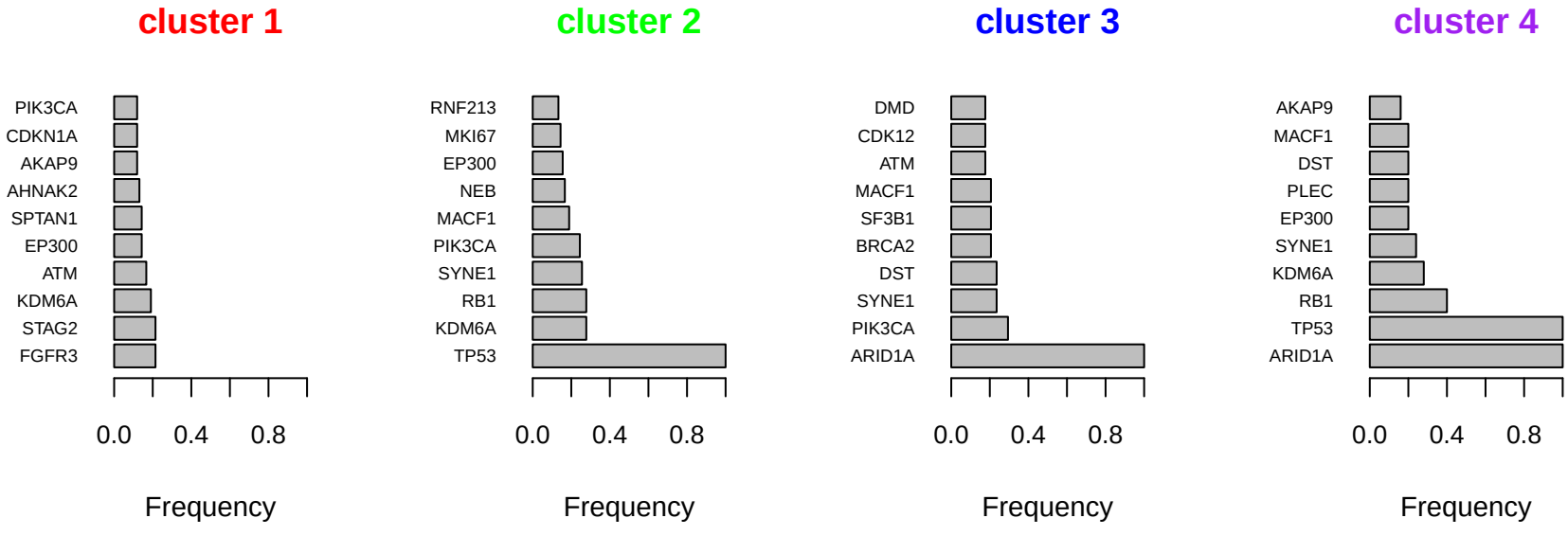
Supplementary Table S2

Column 1: Cancer types
Column 2: Archive names of somatic mutation data
Column 3: File names of somatic mutation data
Column 4: Archive names of clinical data
Column 5: File names of clinical data

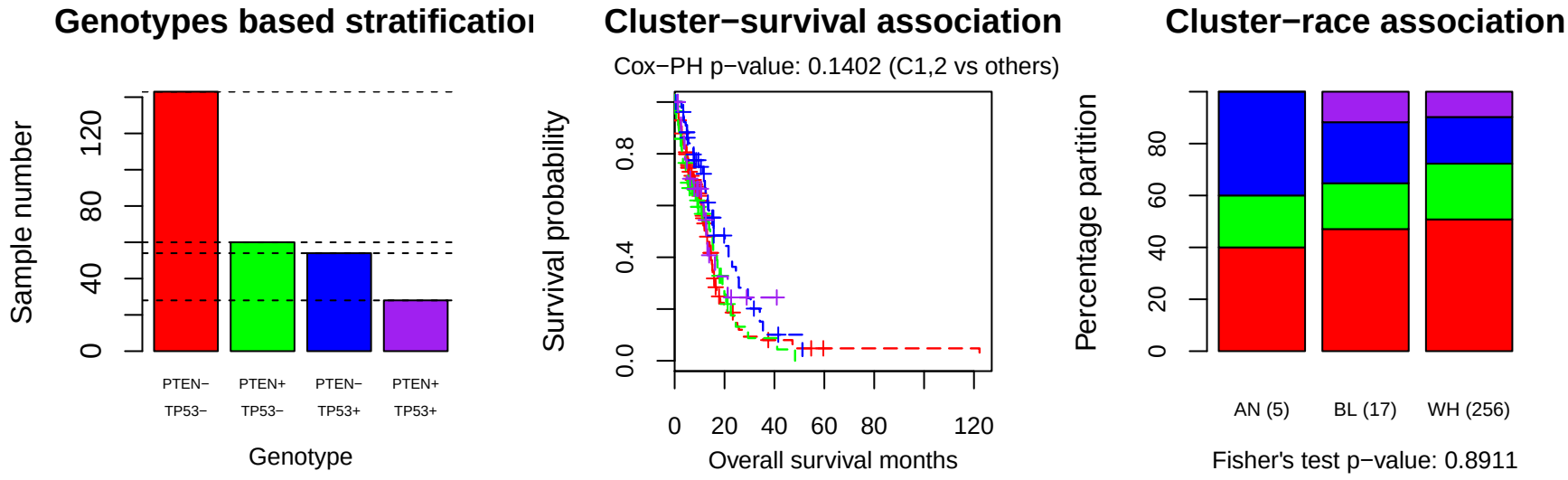
Bladder urothelial carcinoma [BLCA]	broad.mit.edu_BLCA.IlluminaGA_DNASeq_automated.Level_2.1.4.0	PR_TCGA_BLCA_PAIR_Capture_All_Pairs_QCPASS_v5.aggregated.capture.tcga.uuid.automated.somatic.maf	nationwidechildrens.org_BLCA.bio.Level_2.0.27.0	nationwidechildrens.org_clinical_patient_blca.txt
Glioblastoma multiforme [GBM]	broad.mit.edu_GBM.IlluminaGA_DNASeq.Level_2.100.1.0	step4_gbm_liftover.aggregated.capture.tcga.uuid.maf2.4.migrated.somatic.maf	nationwidechildrens.org_GBM.bio.Level_2.0.25.0	nationwidechildrens.org_clinical_patient_gbm.txt
Head and neck squamous cell carcinoma [HNSC]	broad.mit.edu_HNSC.IlluminaGA_DNASeq_automated.Level_2.1.4.0	PR_TCGA_HNSC_PAIR_Capture_All_Pairs_QCPASS_v4.aggregated.capture.tcga.uuid.automated.somatic.maf	nationwidechildrens.org_HNSC.bio.Level_2.0.21.0	nationwidechildrens.org_clinical_patient_hnsc.txt
kidney renal clear cell carcinoma (KIRC)	hgsc.bcm.edu_KIRC.Mixed_DNASeq.Level_2.1.2.0	hgsc.bcm.edu_KIRC.Mixed_DNASeq.1.somatic.maf	nationwidechildrens.org_KIRC.bio.Level_2.0.21.0	nationwidechildrens.org_clinical_patient_kirc.txt
Lung adenocarcinoma [LUAD]	broad.mit.edu_LUAD.IlluminaGA_DNASeq_automated.Level_2.1.5.0	PR_TCGA_LUAD_PAIR_Capture_All_Pairs_QCPASS_v4.aggregated.capture.tcga.uuid.automated.somatic.maf	nationwidechildrens.org_LUAD.bio.Level_2.0.24.0	nationwidechildrens.org_clinical_patient_luad.txt
Lung squamous cell carcinoma [LUSC]	broad.mit.edu_LUSC.IlluminaGA_DNASeq.Level_2.100.1.0	step4_LUSC_Paper_v8.aggregated.tcga.maf2.4.migrated.somatic.maf	nationwidechildrens.org_LUSC.bio.Level_2.0.21.0	nationwidechildrens.org_clinical_patient_lusc.txt
Breast invasive carcinoma [BRCA]	genome.wustl.edu_BRCA.IlluminaGA_DNASeq_curated.Level_2.1.1.0	genome.wustl.edu_BRCA.IlluminaGA_DNASeq.Level_2.1.1.0.curated.somatic.maf	nationwidechildrens.org_BRCA.bio.Level_2.0.24.0	nationwidechildrens.org_clinical_patient_brca.txt
Ovarian serous cystadenocarcinoma [OV]	broad.mit.edu_OV.IlluminaGA_DNASeq.Level_2.100.1.0	step4_ov_liftover.aggregated.capture.tcga.uuid.maf2.4.migrated.somatic.maf	nationwidechildrens.org_OV.bio.Level_2.0.23.0	nationwidechildrens.org_clinical_patient_ov.txt
Ovarian serous cystadenocarcinoma [OV]	genome.wustl.edu_OV.IlluminaGA_DNASeq.Level_2.1.3.0	genome.wustl.edu_OV.IlluminaGA_DNASeq.1.3.somatic.maf	nationwidechildrens.org_OV.bio.Level_2.0.23.0	nationwidechildrens.org_clinical_patient_ov.txt
Ovarian serous cystadenocarcinoma [OV]	genome.wustl.edu_OV.IlluminaGA_DNASeq.Level_2.2.1.0	genome.wustl.edu_OV.IlluminaGA_DNASeq.Level_2.2.0.0.somatic.maf	nationwidechildrens.org_OV.bio.Level_2.0.23.0	nationwidechildrens.org_clinical_patient_ov.txt
Uterine corpus endometrial carcinoma [UCEC]	genome.wustl.edu_UCEC.IlluminaGA_DNASeq.Level_2.1.7.0	genome.wustl.edu_UCEC.IlluminaGA_DNASeq.Level_2.1.7.somatic.maf	nationwidechildrens.org_UCEC.bio.Level_2.0.22.0	nationwidechildrens.org_clinical_patient_ucec.txt
Cervical squamous cell carcinoma and endocervical adenocarcinoma [CESC]	genome.wustl.edu_UCEC.IlluminaGA_DNASeq.Level_2.1.7.0	genome.wustl.edu_UCEC.IlluminaGA_DNASeq.Level_2.1.7.somatic.maf	nationwidechildrens.org_UCEC.bio.Level_2.0.22.0	nationwidechildrens.org_clinical_patient_ucec.txt
esophageal carcinoma (ESCA)	ucsc.edu_ESCA.IlluminaGA_DNASeq_automated.Level_2.1.0.0	ucsc.edu_ESCA.IlluminaGA_DNASeq_automated.Level_2.1.0.0.somatic.maf	nationwidechildrens.org_ESCA.bio.Level_2.0.34.0	nationwidechildrens.org_clinical_patient_esca.txt
kidney renal papillary cell carcinoma (KIRP)	ucsc.edu_KIRP.IlluminaGA_DNASeq_automated.Level_2.1.2.0	ucsc.edu_KIRP.IlluminaGA_DNASeq_automated.Level_2.1.2.0.somatic.maf	nationwidechildrens.org_KIRP.bio.Level_2.0.34.0	nationwidechildrens.org_clinical_patient_kirp.txt
Colon adenocarcinoma [COAD]	hgsc.bcm.edu_COAD.IlluminaGA_DNASeq.Level_2.1.5.0	hgsc.bcm.edu_COAD.IlluminaGA_DNASeq.1.somatic.maf	nationwidechildrens.org_COAD.bio.Level_2.0.23.0	nationwidechildrens.org_clinical_patient_coad.txt
Esophageal carcinoma [ESCA]	ucsc.edu_ESCA.IlluminaGA_DNASeq_automated.Level_2.1.0.0	ucsc.edu_ESCA.IlluminaGA_DNASeq_automated.Level_2.1.0.0.somatic.maf	nationwidechildrens.org_ESCA.bio.Level_2.0.34.0	nationwidechildrens.org_clinical_patient_esca.txt
Liver hepatocellular carcinoma [LIHC]	ucsc.edu_LIHC.IlluminaGA_DNASeq_automated.Level_2.1.1.0	ucsc.edu_LIHC.IlluminaGA_DNASeq_automated.Level_2.1.1.0.somatic.maf	nationwidechildrens.org_LIHC.bio.Level_2.0.38.0	nationwidechildrens.org_clinical_patient_lihc.txt
Stomach adenocarcinoma [STAD]	hgsc.bcm.edu_STAD.IlluminaGA_DNASeq_automated.Level_2.1.0.0	hgsc.bcm.edu_STAD.IlluminaGA_DNASeq.1.somatic.maf	nationwidechildrens.org_STAD.bio.Level_2.0.31.0	nationwidechildrens.org_clinical_patient_stad.txt

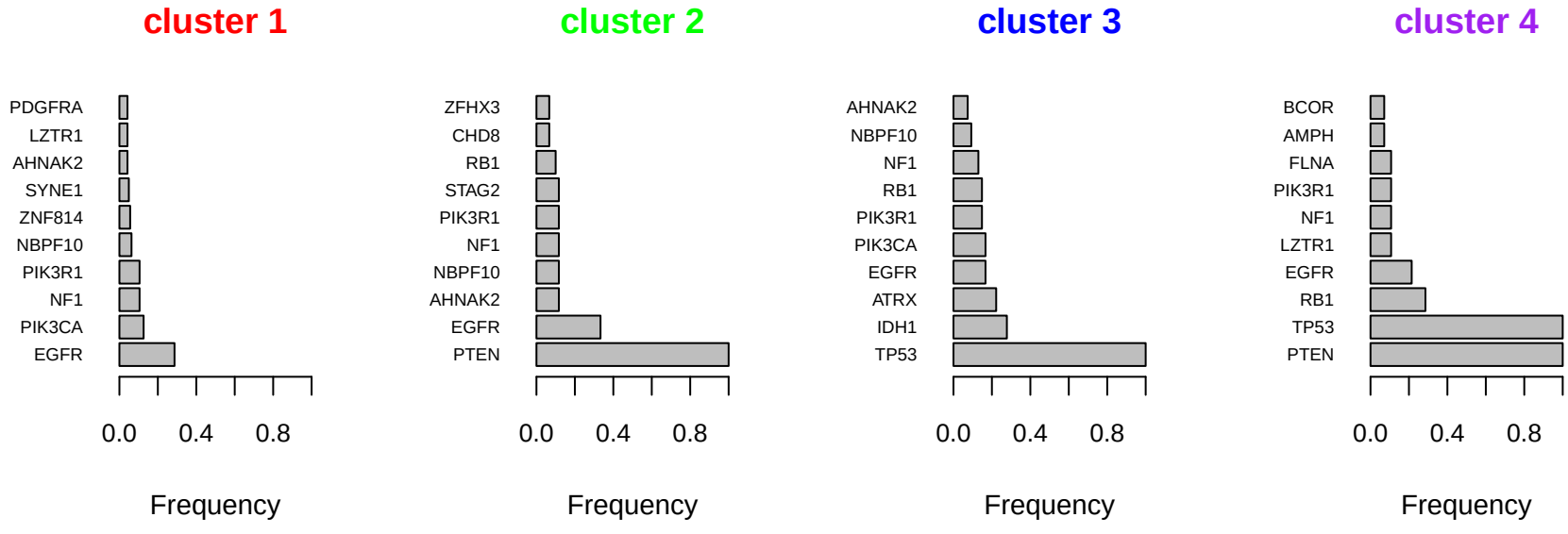
BLCA − results of 2GeneModel



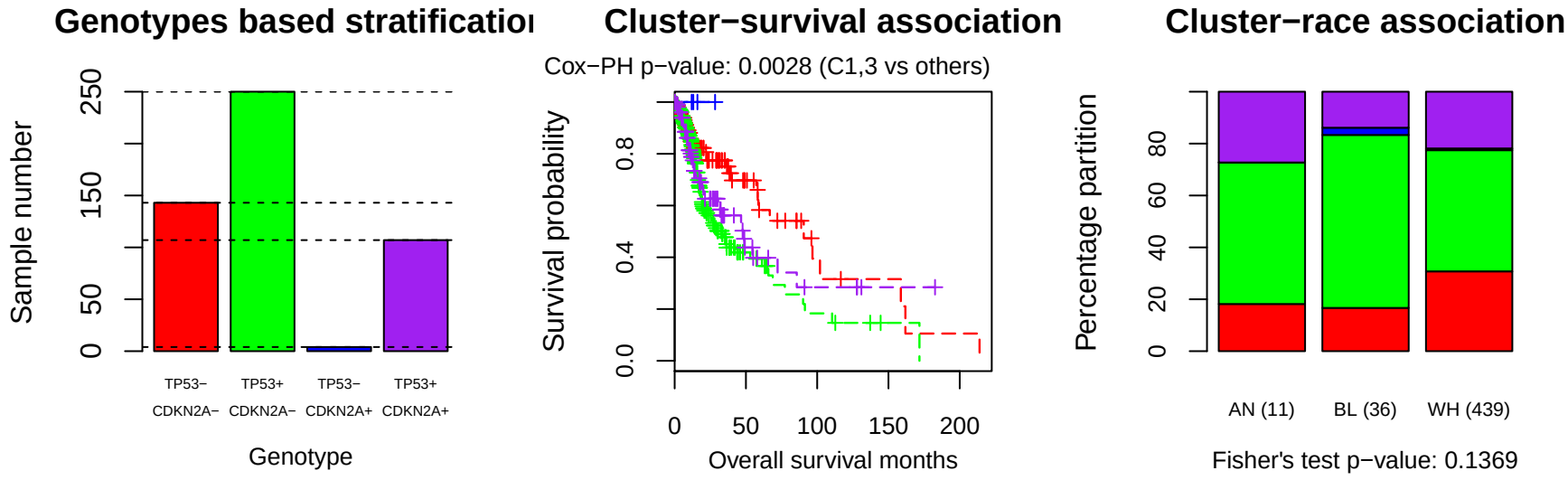


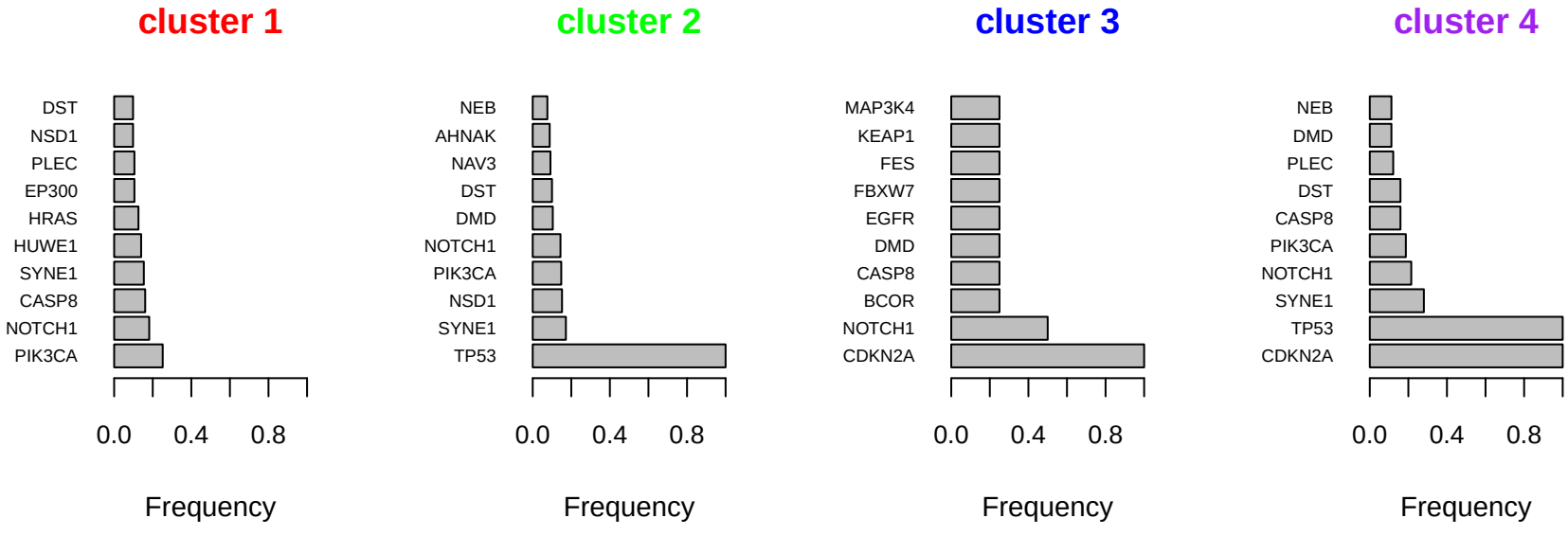
GBM − results of 2GeneModel





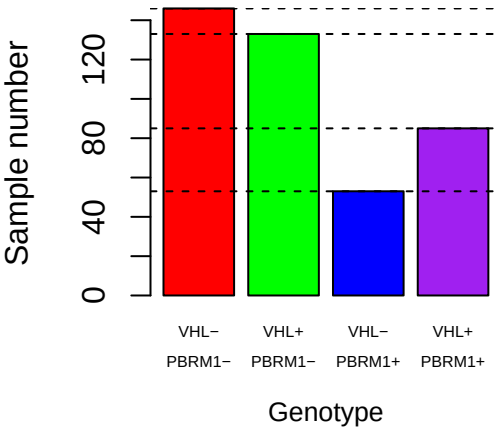
HNSC − results of 2GeneModel





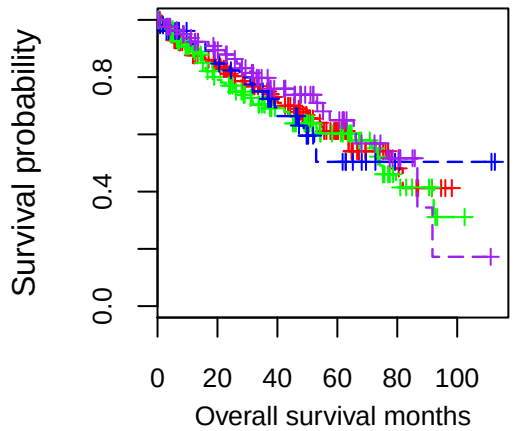
KIRC − results of 2GeneModel

Genotypes based stratificatio

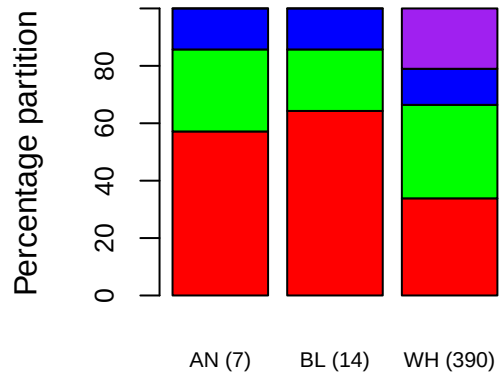


Cluster−survival association

Cox−PH p−value: 0.3139 (C4 vs others)

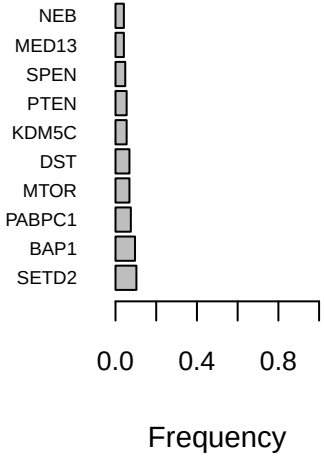


Cluster−race association

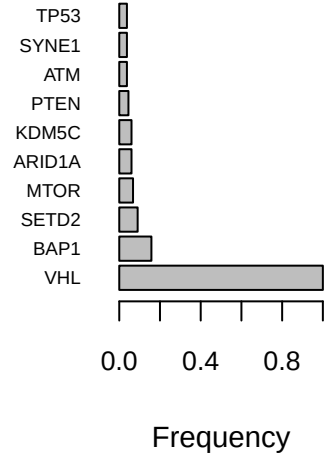


Fisher's test p−value: 0.1154

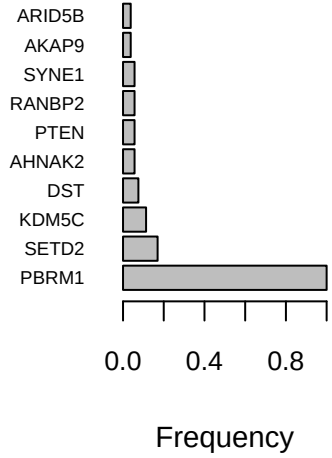
cluster 1



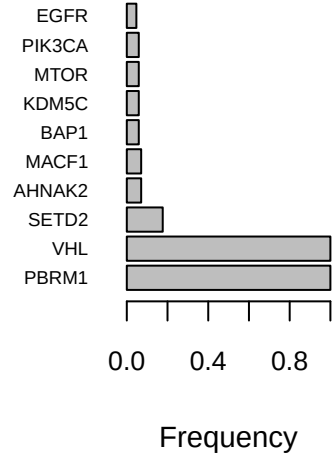
cluster 2



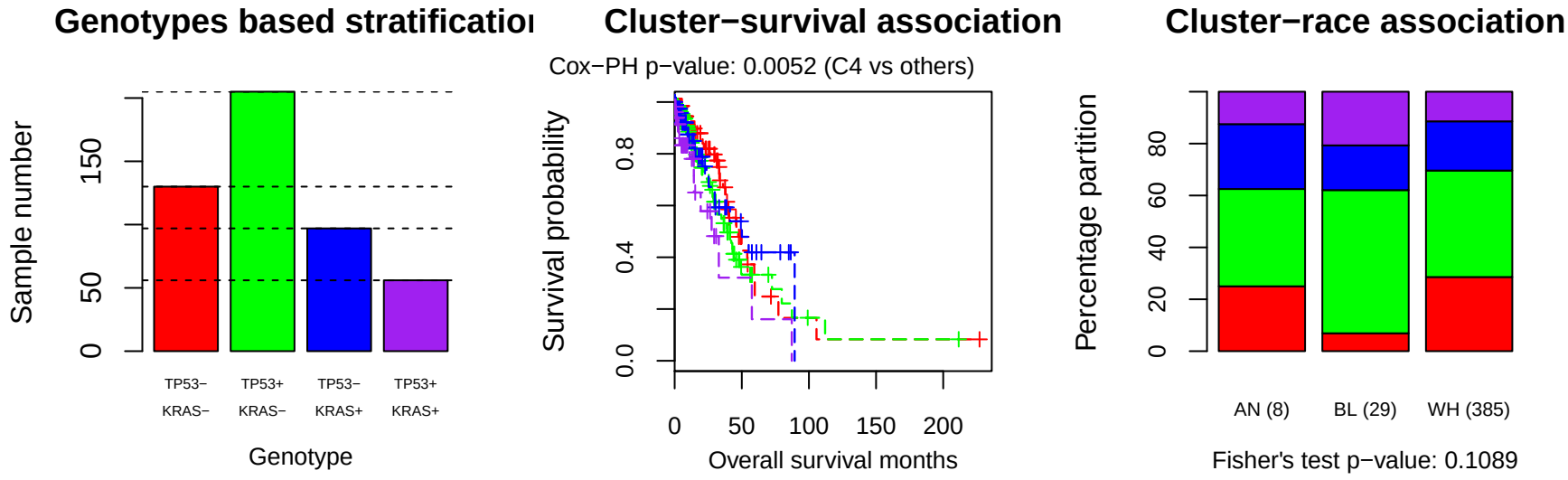
cluster 3

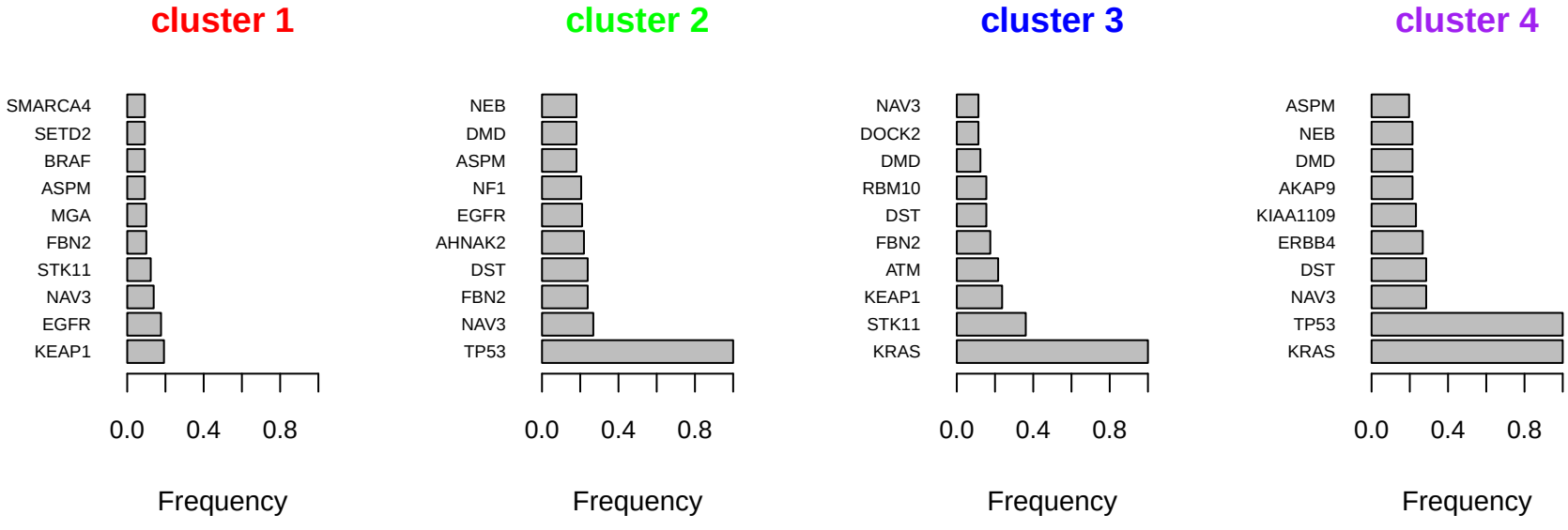


cluster 4

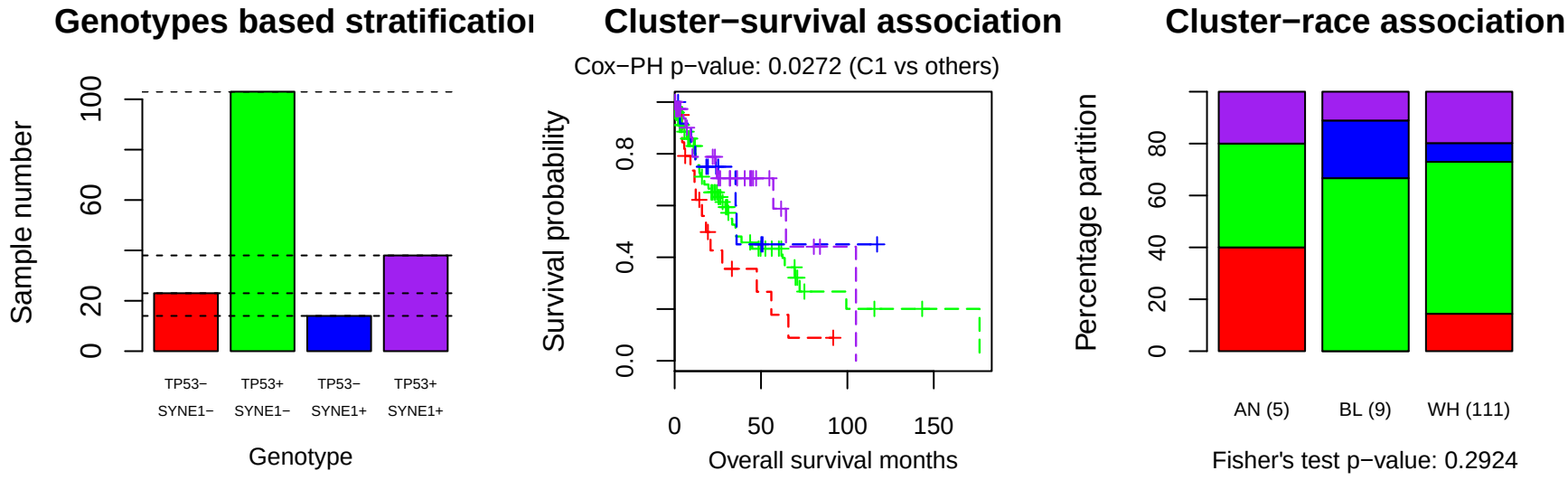


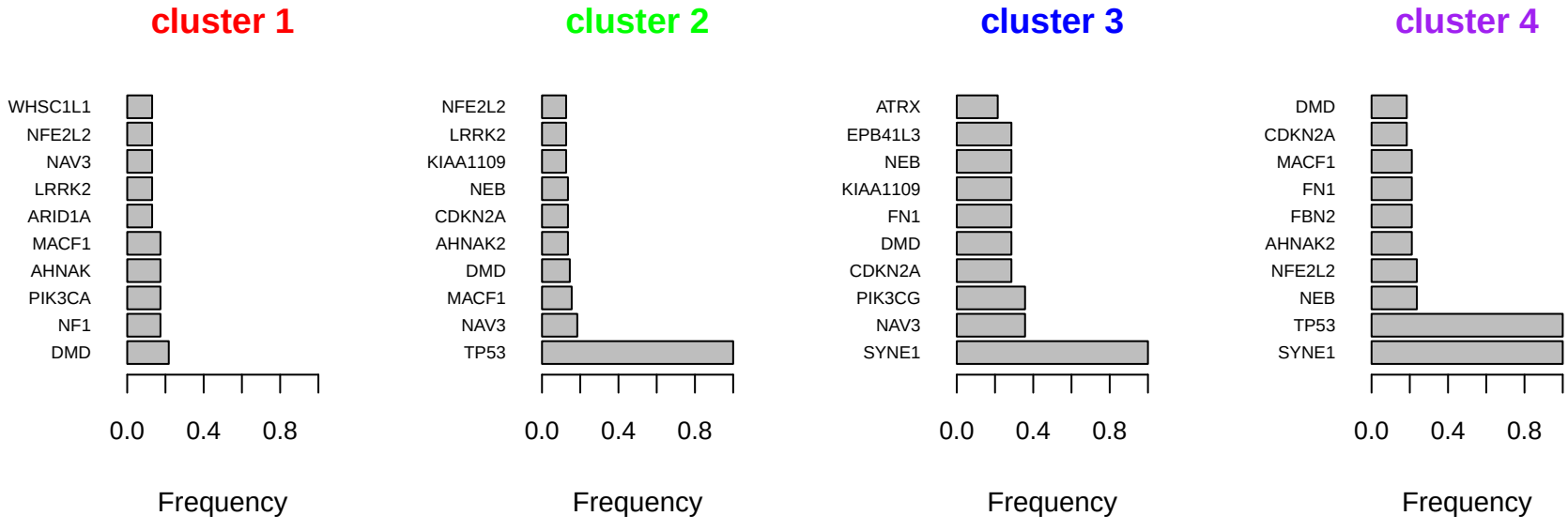
LUAD − results of 2GeneModel





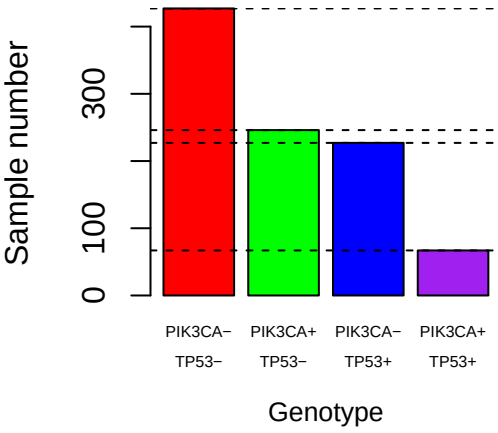
LUSC − results of 2GeneModel





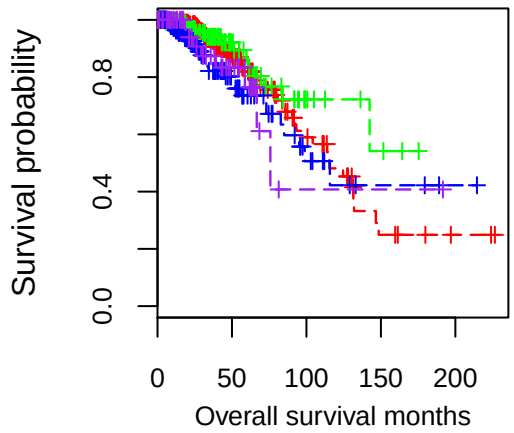
BRCA − results of 2GeneModel

Genotypes based stratificatio

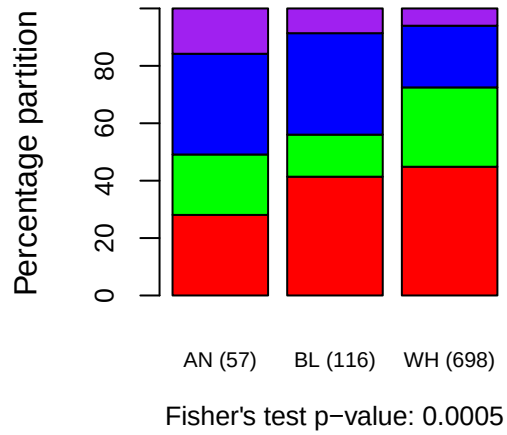


Cluster−survival association

Cox−PH p−value: 0.0054 (C1,2 vs others)

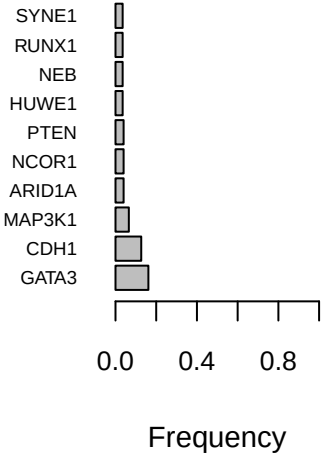


Cluster−race association

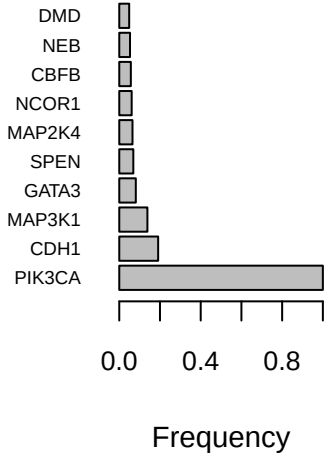


Fisher's test p−value: 0.0005

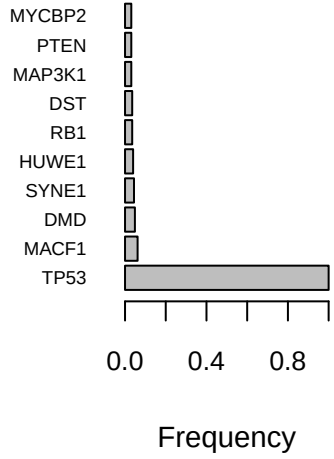
cluster 1



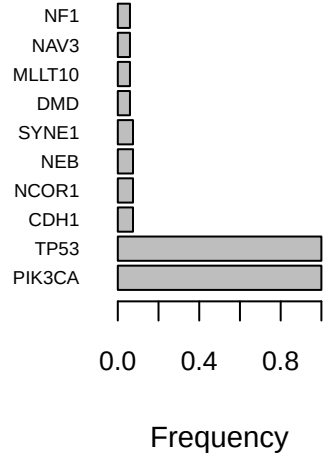
cluster 2



cluster 3

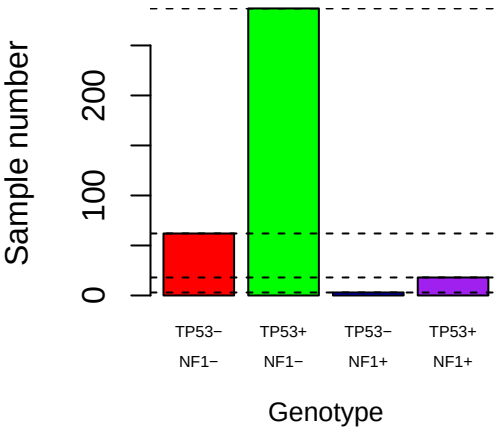


cluster 4



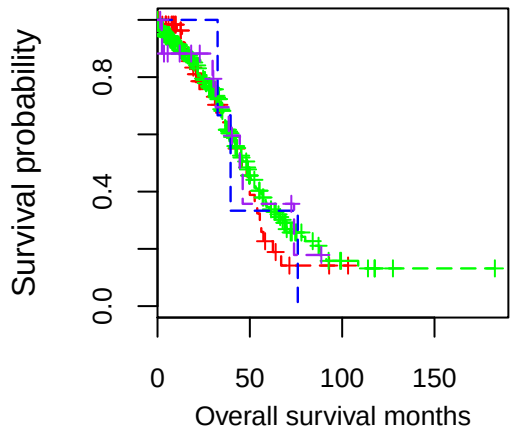
OV − results of 2GeneModel

Genotypes based stratificatio

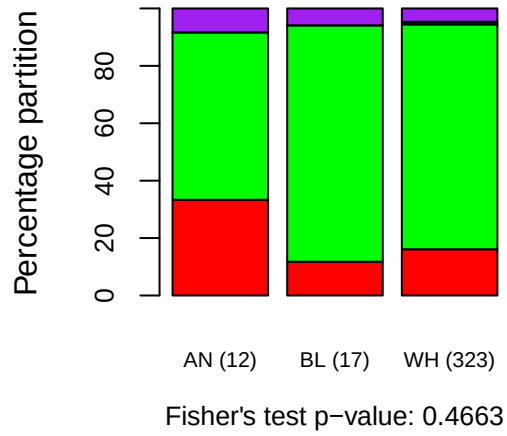


Cluster−survival association

Cox−PH p−value: 0.2186 (C3 vs others)

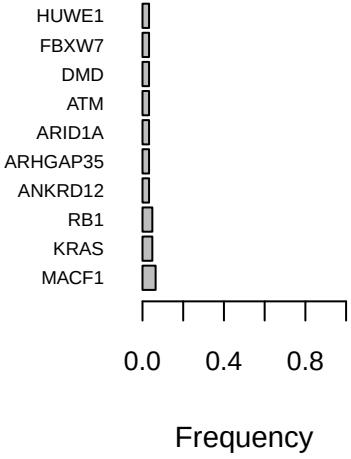


Cluster−race association

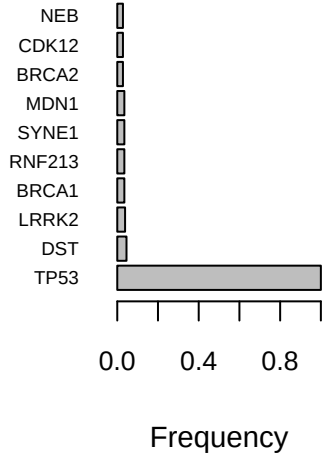


Fisher's test p−value: 0.4663

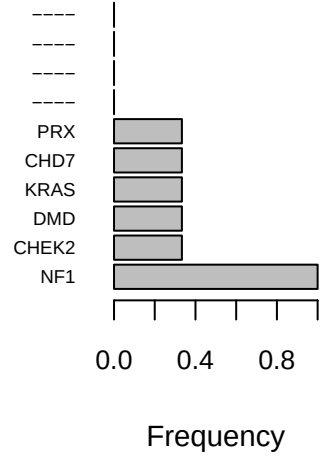
cluster 1



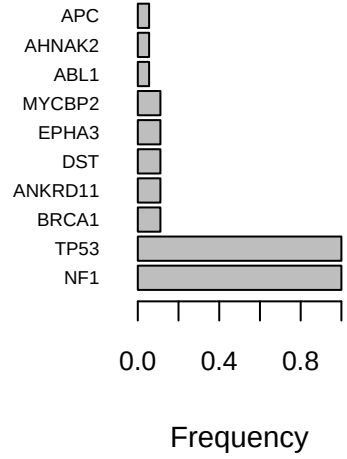
cluster 2



cluster 3

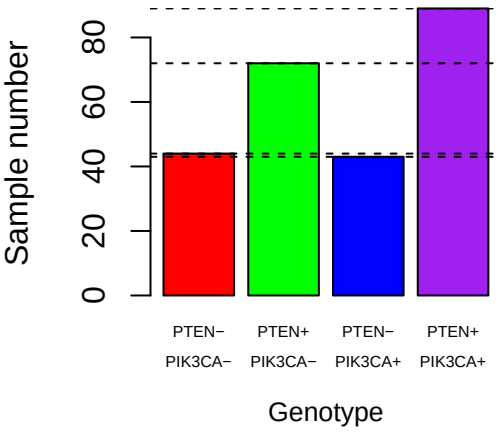


cluster 4



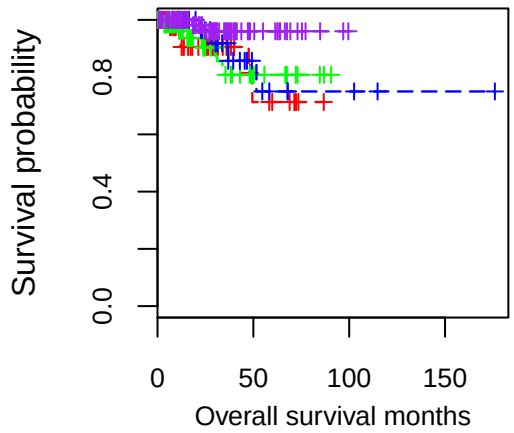
UCEC − results of 2GeneModel

Genotypes based stratificatio

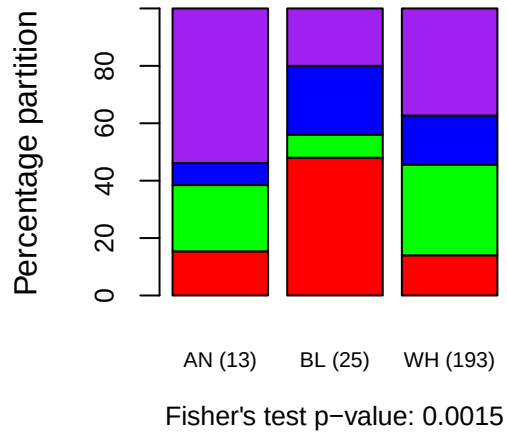


Cluster−survival association

Cox−PH p−value: 0.0401 (C4 vs others)

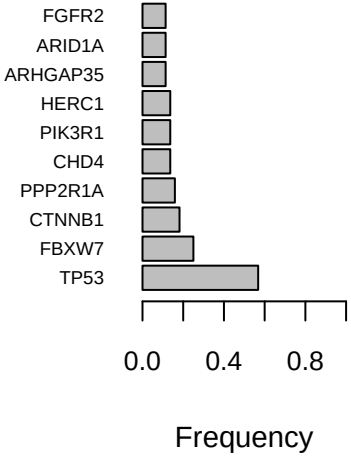


Cluster−race association

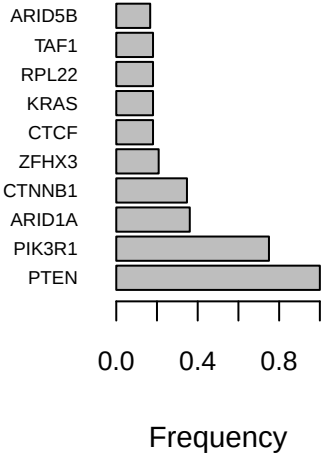


Fisher's test p−value: 0.0015

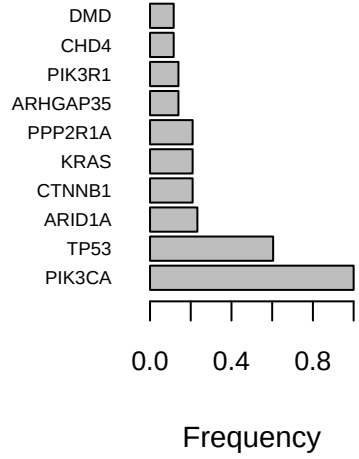
cluster 1



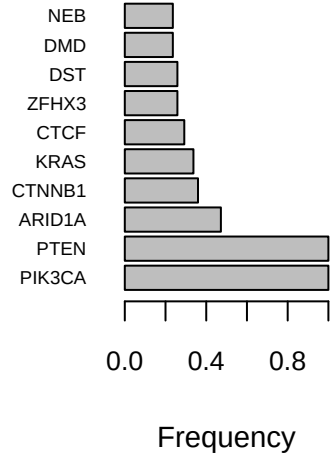
cluster 2



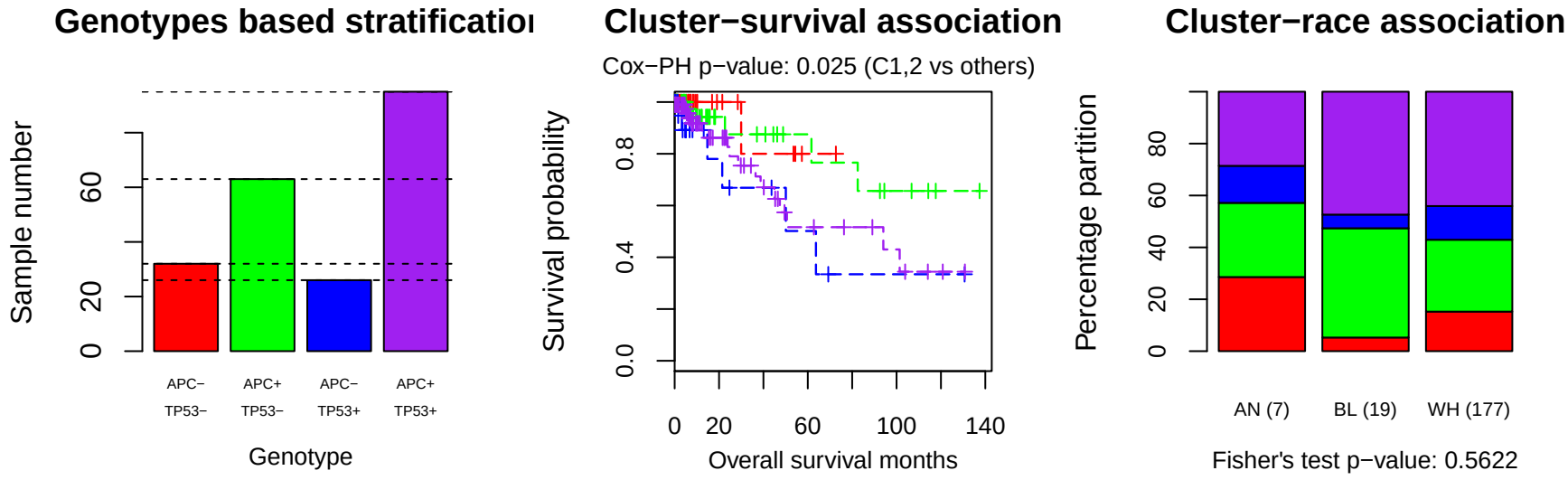
cluster 3

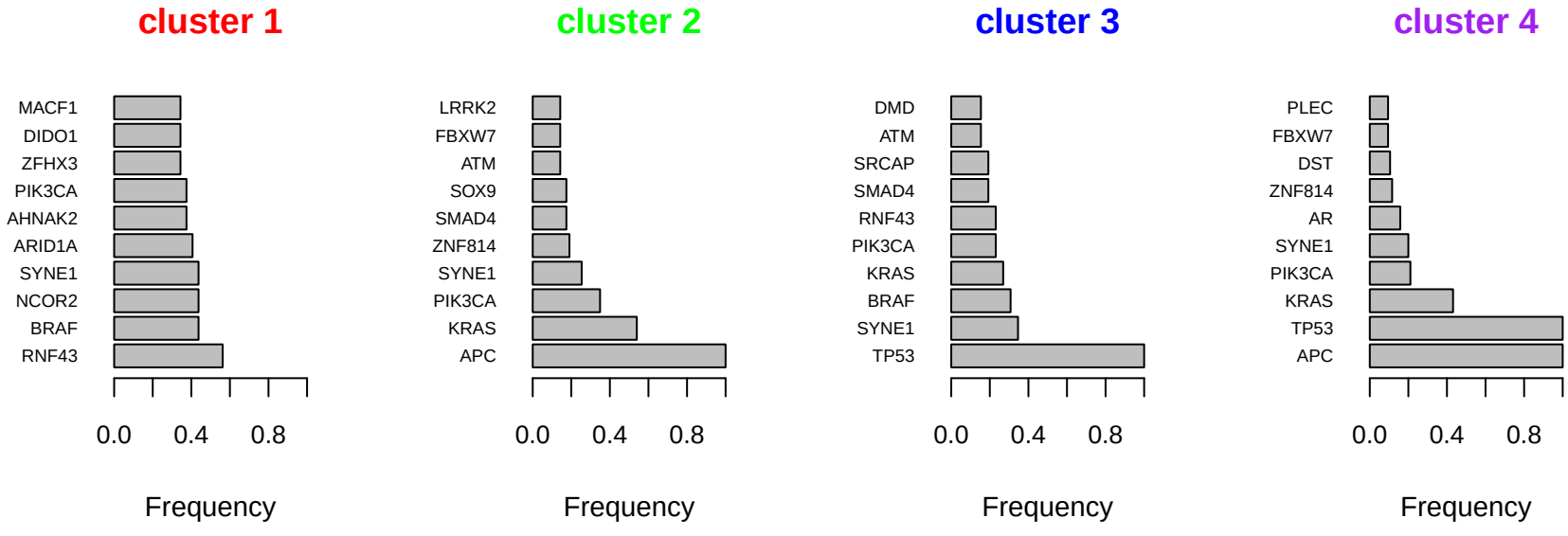


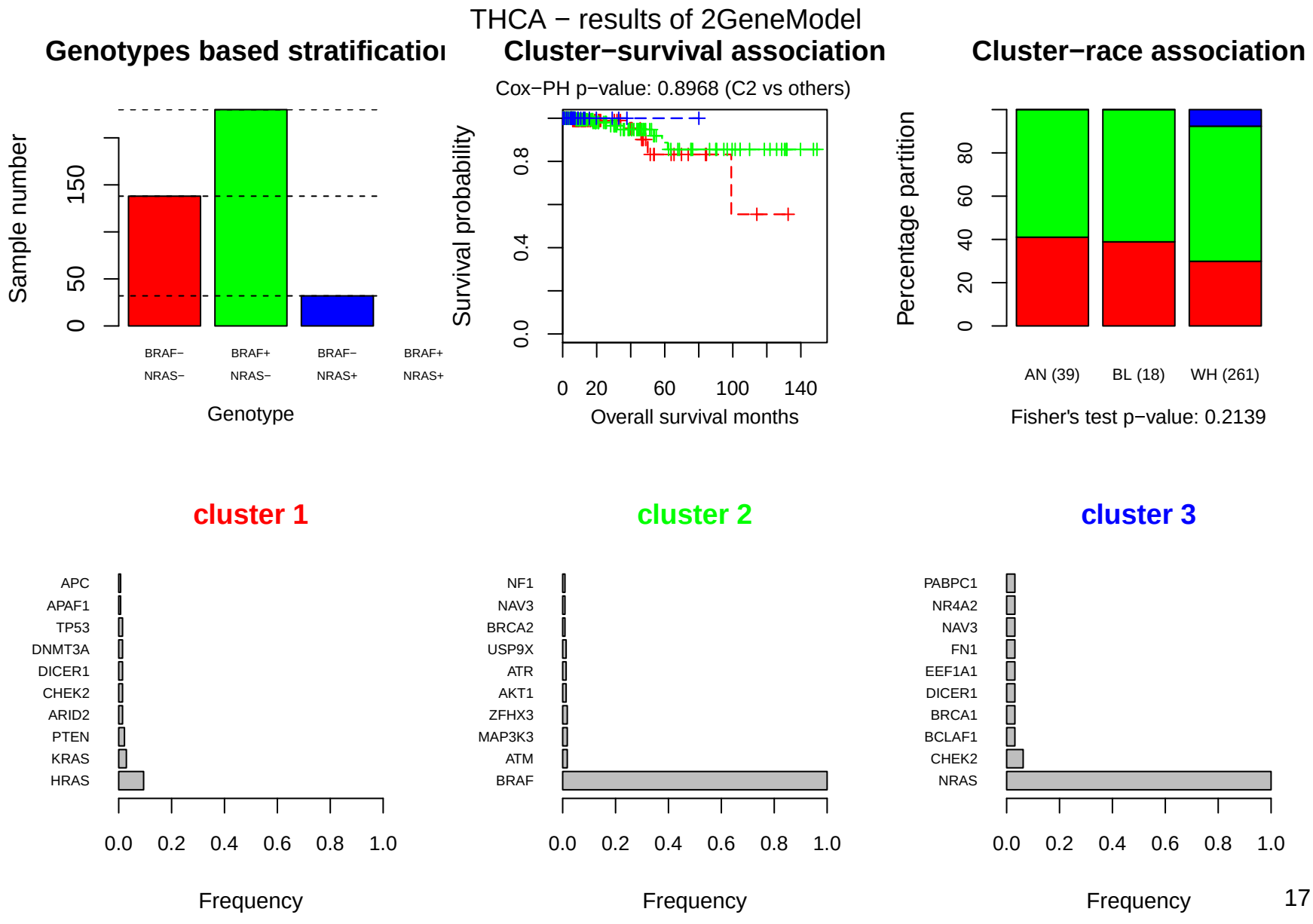
cluster 4



COAD − results of 2GeneModel

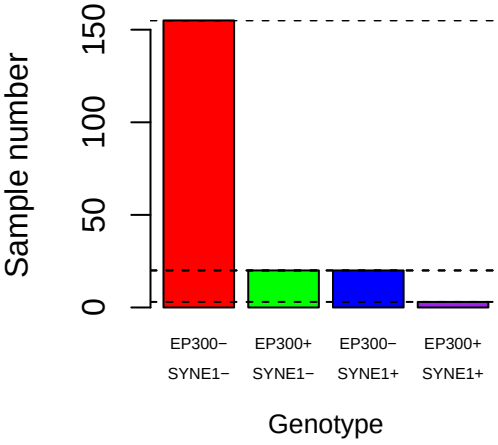





THCA − results of 2GeneModel
Genotypes based stratificatio
Cluster−survival association
Cluster−race association
Cox−PH p−value: 0.8968 (C2 vs others)
Sample number
0
50
150
BRAF−
NRAS−
BRAF+
NRAS−
BRAF−
NRAS+
BRAF+
NRAS+
Genotype
Survival probability
0.0
0.4
0.8
0
20
60
100
140
Overall survival months
Percentage partition
0
20
40
60
80
AN (39)
BL (18)
WH (261)
Fisher's test p−value: 0.2139
cluster 1
cluster 2
cluster 3
APC
APAF1
TP53
DNMT3A
DICER1
CHEK2
ARID2
PTEN
KRAS
HRAS
NF1
NAV3
BRCA2
USP9X
ATR
AKT1
ZFHX3
MAP3K3
ATM
BRAF
PABPC1
NR4A2
NAV3
FN1
EEF1A1
DICER1
BRCA1
BCLAF1
CHEK2
NRAS
0.0
0.2
0.4
0.6
0.8
1.0
Frequency

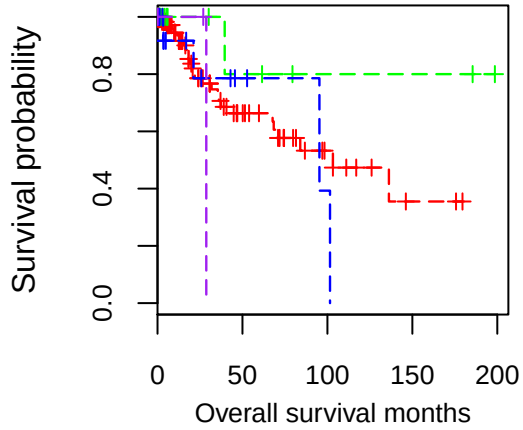
CESC − results of 2GeneModel

Genotypes based stratificatio

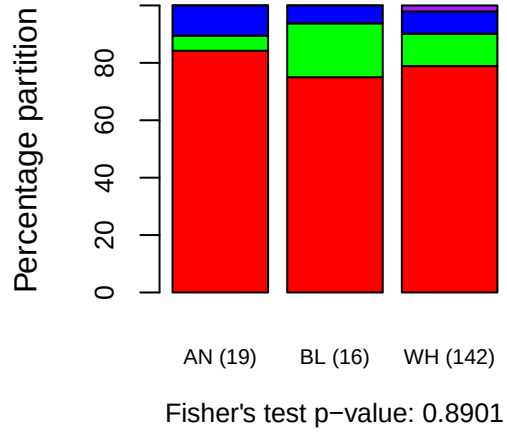


Cluster−survival association

Cox−PH p−value: 0.1421 (C2 vs others)

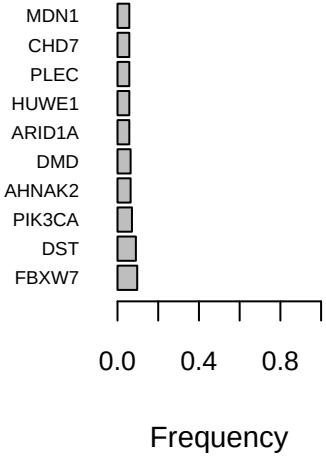


Cluster−race association

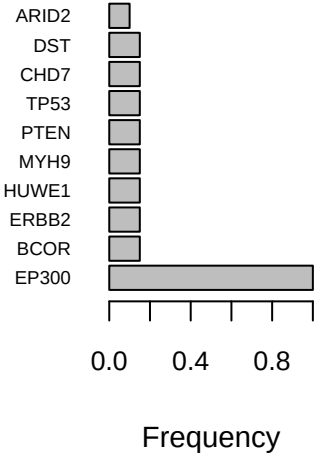


Fisher's test p−value: 0.8901

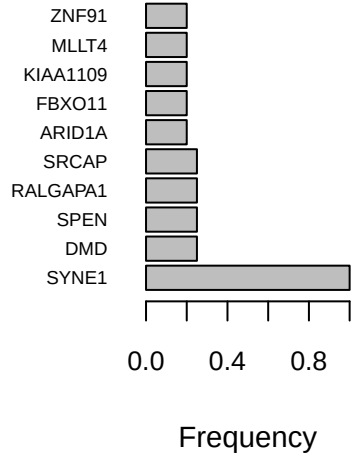
cluster 1



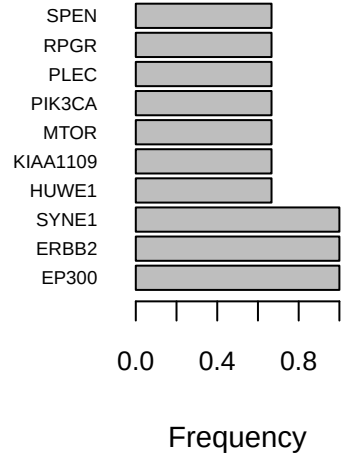
cluster 2

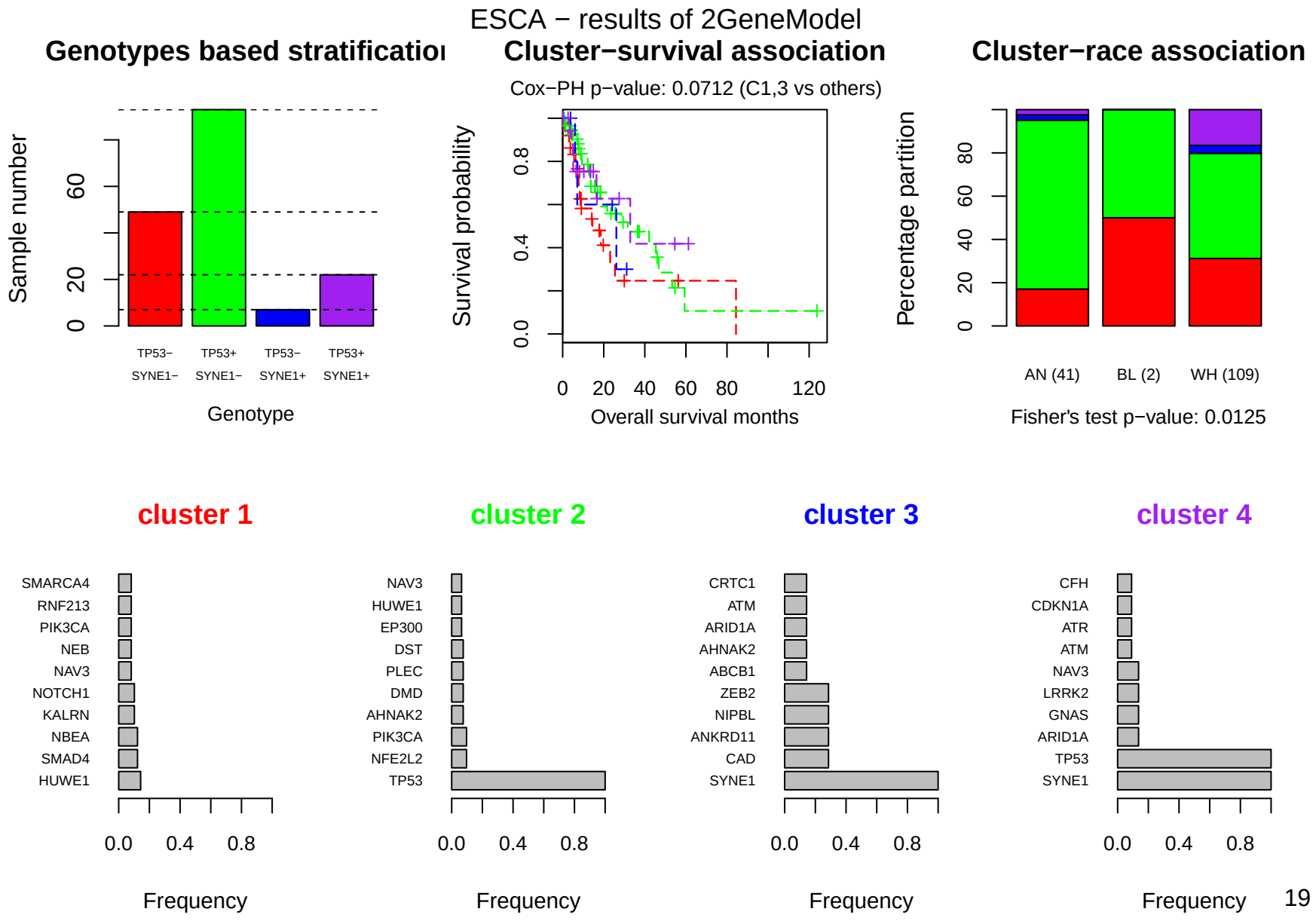


cluster 3



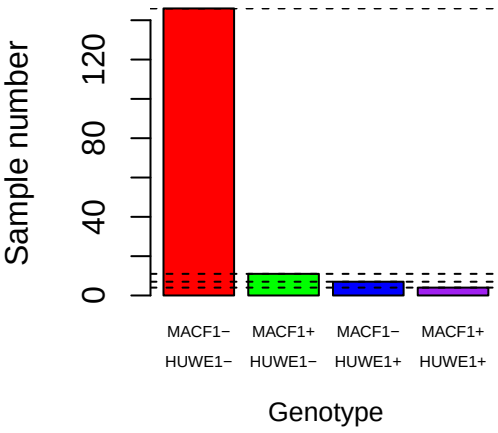
cluster 4




ESCA − results of 2GeneModel
Genotypes based stratificatio
Cluster−survival association
Cluster−race association
Sample number
0
20
60
TP53−
SYNE1−
TP53+
SYNE1−
TP53−
SYNE1+
TP53+
SYNE1+
Genotype
Cox−PH p−value: 0.0712 (C1,3 vs others)
Survival probability
0.0
0.4
0.8
0
20
40
60
80
120
Overall survival months
Percentage partition
0
20
40
60
80
AN (41)
BL (2)
WH (109)
Fisher's test p−value: 0.0125
cluster 1
SMARCA4
RNF213
PIK3CA
NEB
NAV3
NOTCH1
KALRN
NBEA
SMAD4
HUWE1
0.0
0.4
0.8
Frequency
cluster 2
NAV3
HUWE1
EP300
DST
PLEC
DMD
AHNAK2
PIK3CA
NFE2L2
TP53
0.0
0.4
0.8
Frequency
cluster 3
CRTC1
ATM
ARID1A
AHNAK2
ABCB1
ZEB2
NIPBL
ANKRD11
CAD
SYNE1
0.0
0.4
0.8
Frequency
cluster 4
CFH
CDKN1A
ATR
ATM
NAV3
LRRK2
GNAS
ARID1A
TP53
SYNE1
0.0
0.4
0.8
Frequency

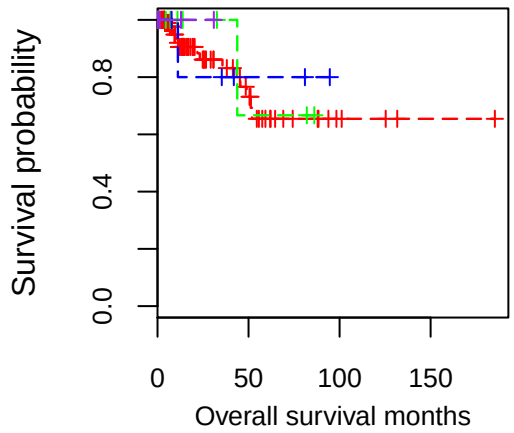
KIRP − results of 2GeneModel

Genotypes based stratificatio

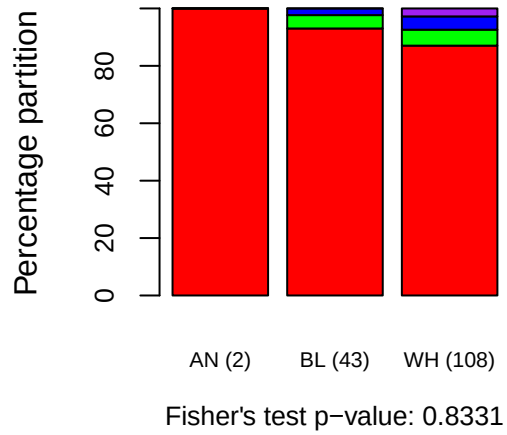


Cluster−survival association

Cox−PH p−value: 0.5227 (C1,2 vs others)

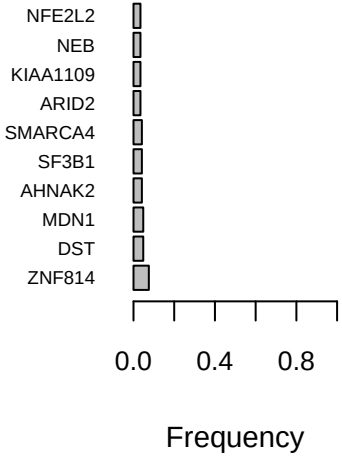


Cluster−race association

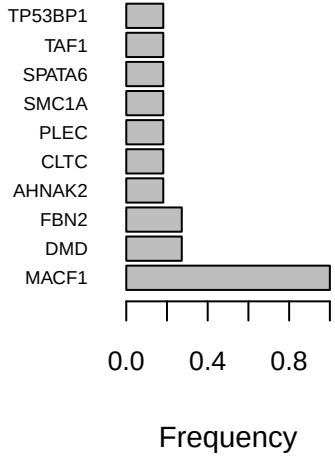


Fisher's test p−value: 0.8331

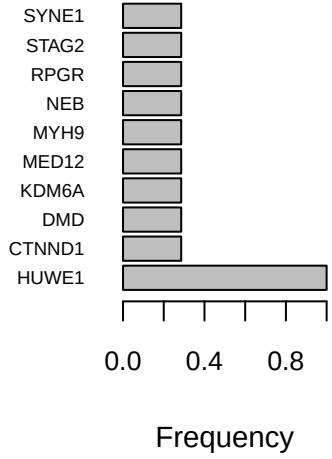
cluster 1



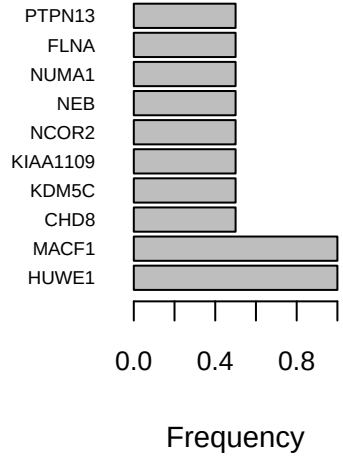
cluster 2



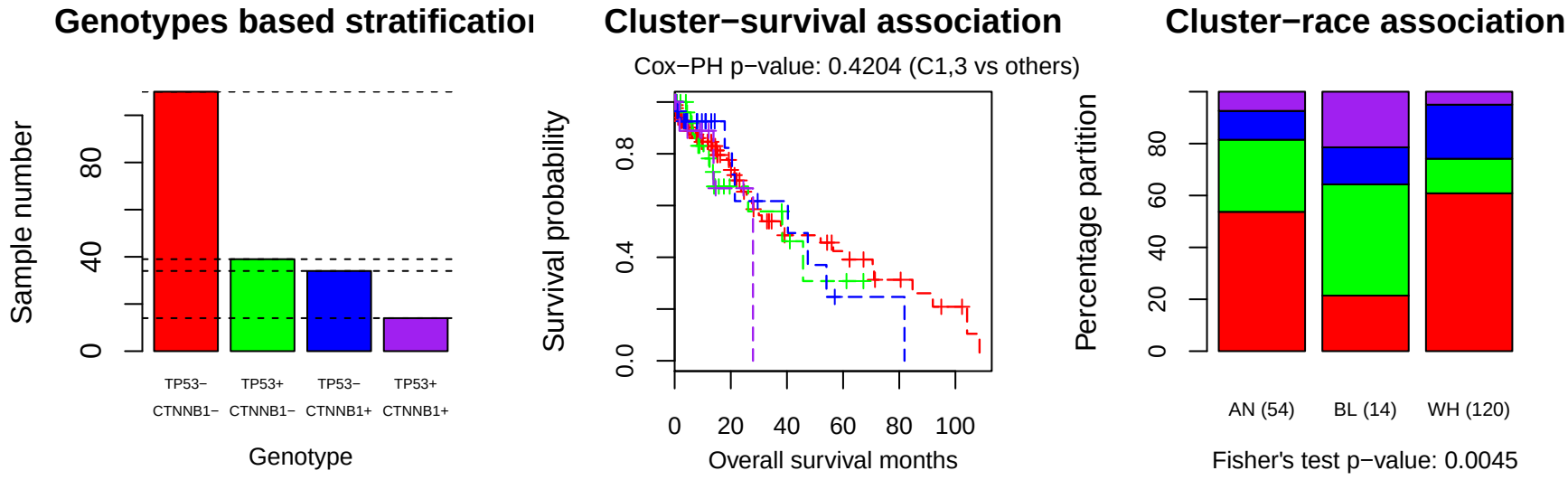
cluster 3

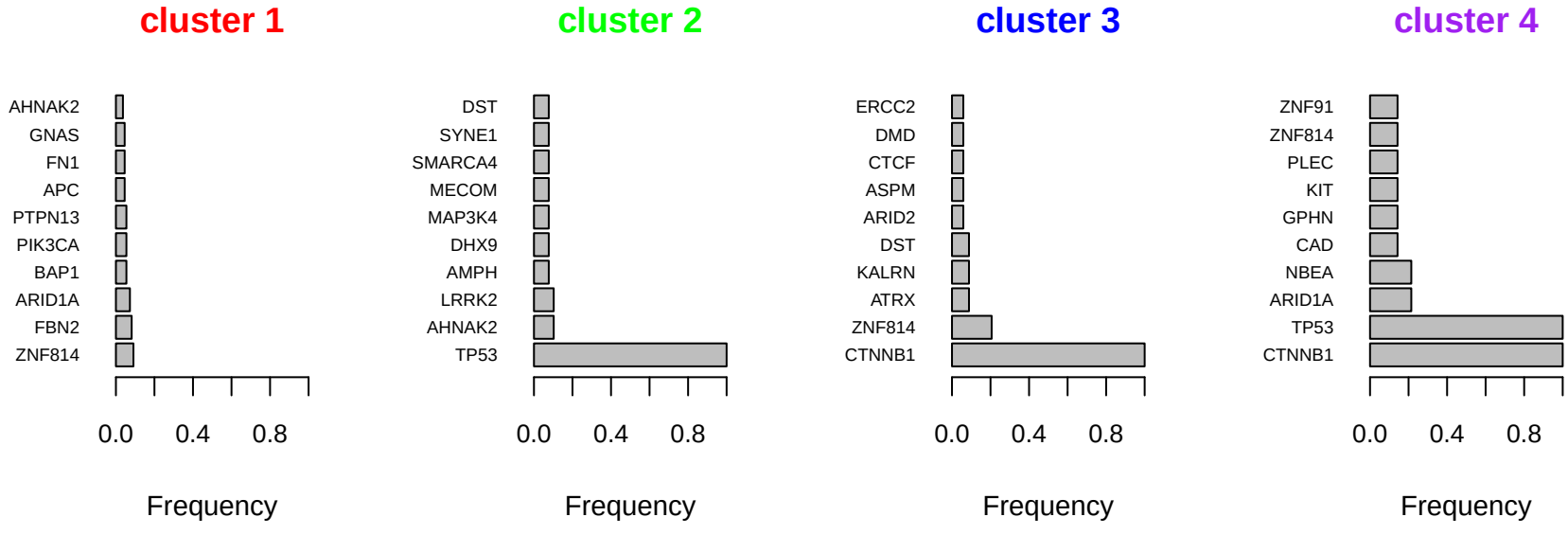


cluster 4



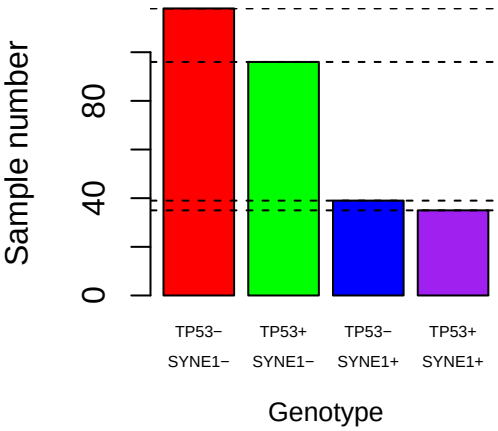
LIHC − results of 2GeneModel



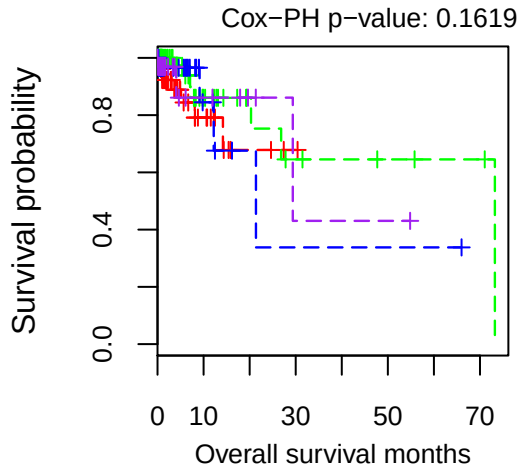


STAD − results of 2GeneModel

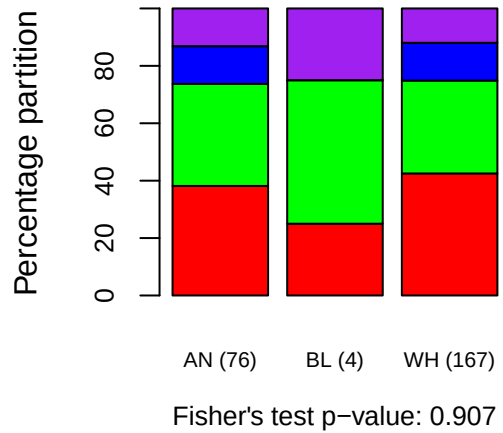
Genotypes based stratificatio



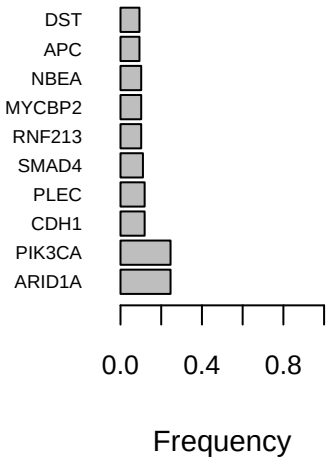
Cluster−survival association



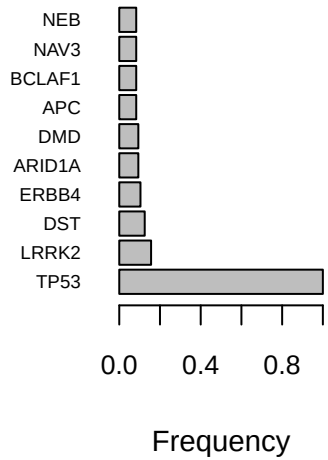
Cluster−race association



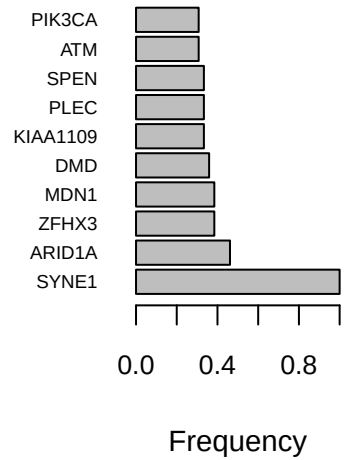
cluster 1



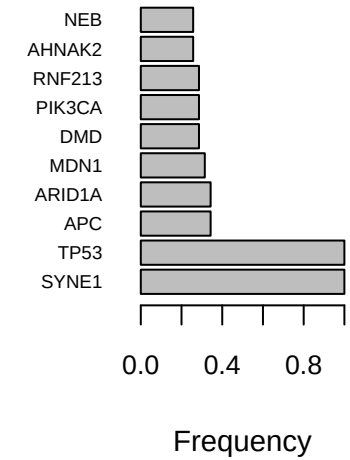
cluster 2



cluster 3

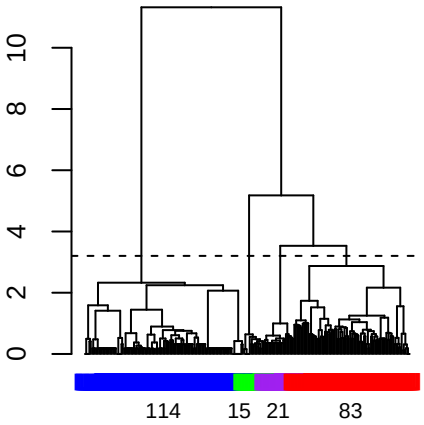


cluster 4



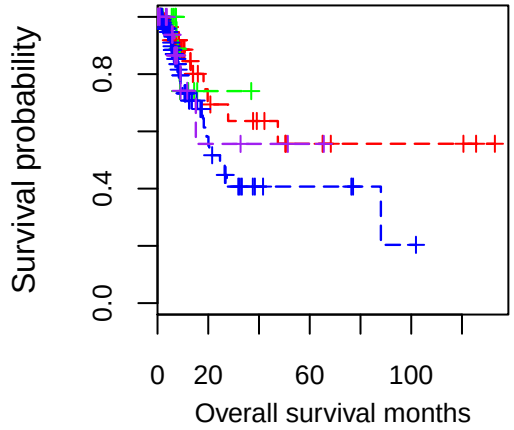
BLCA − results of xGeneModel

Mutation−based clustering

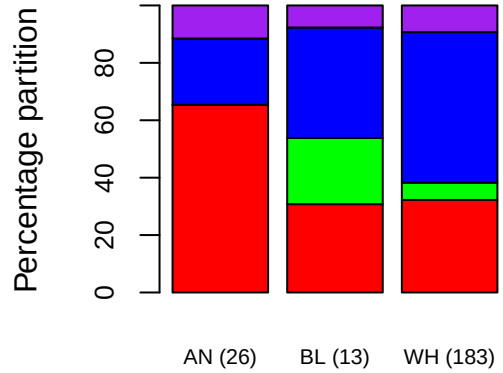


Cluster−survival association

Cox−PH p−value: 0.0617 (C1,2 vs others)

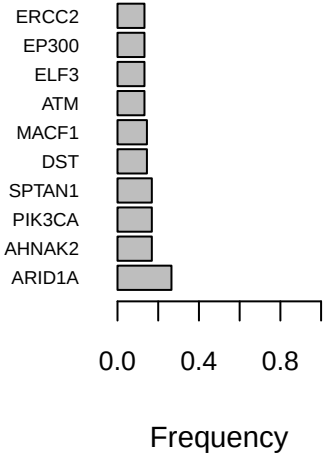


Cluster−race association

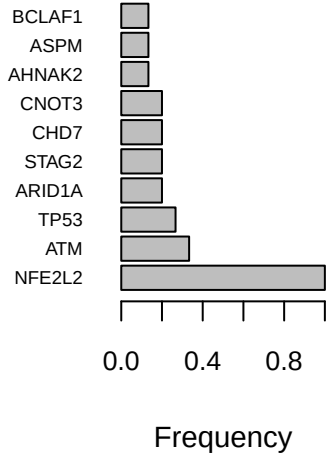


Fisher's test p−value: 0.004

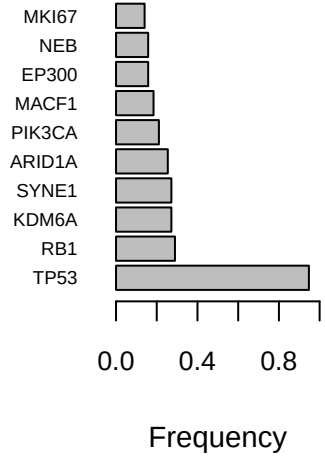
cluster 1



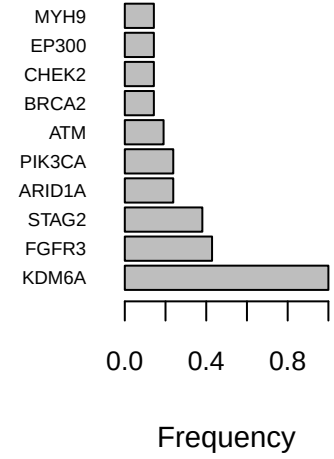
cluster 2



cluster 3

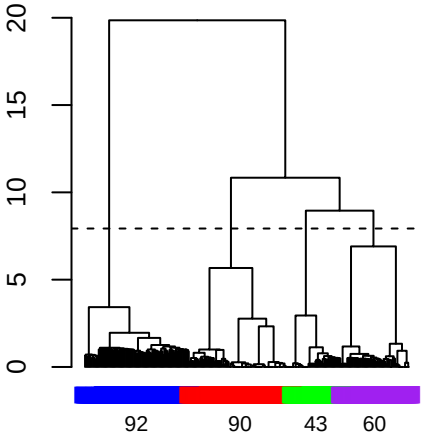


cluster 4



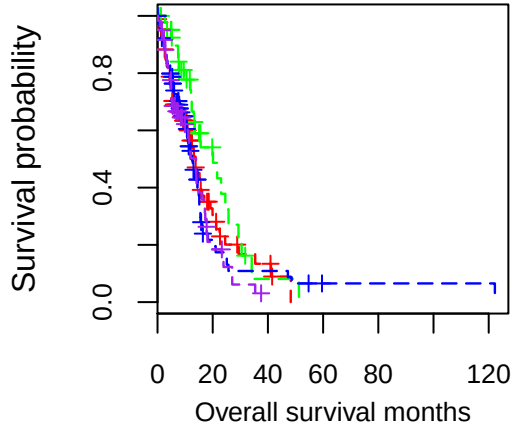
GBM − results of xGeneModel

Mutation−based clustering

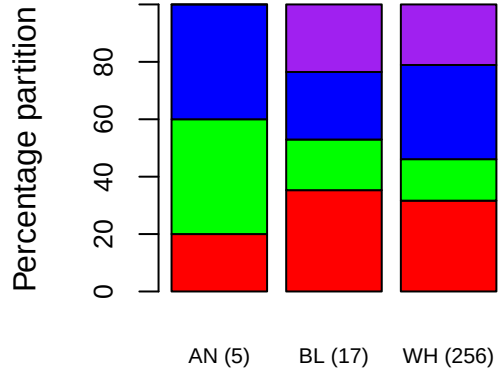


Cluster−survival association

Cox−PH p−value: 0.1977 (C1,2 vs others)

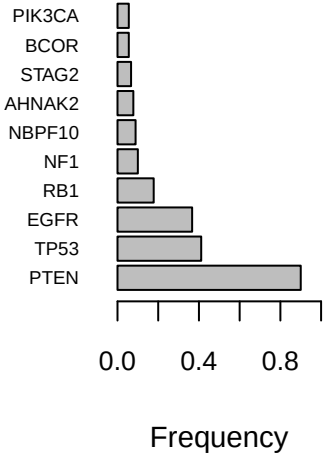


Cluster−race association

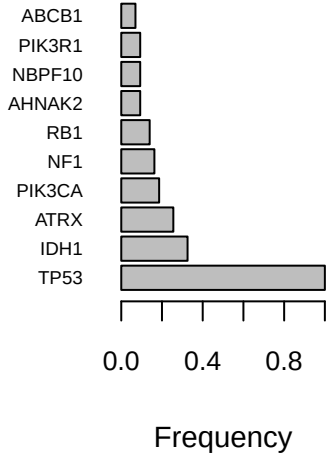


Fisher's test p−value: 0.6892

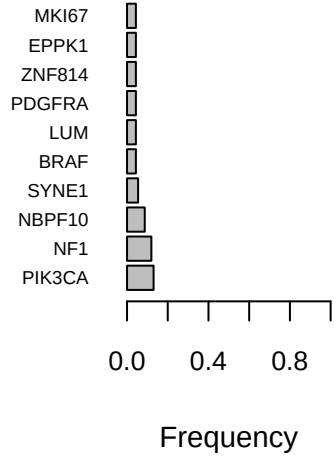
cluster 1



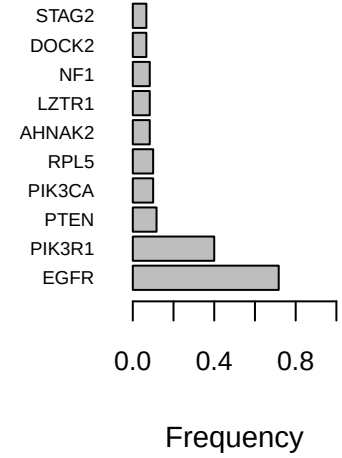
cluster 2



cluster 3

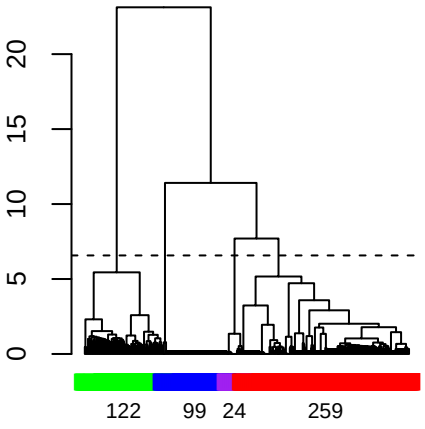


cluster 4



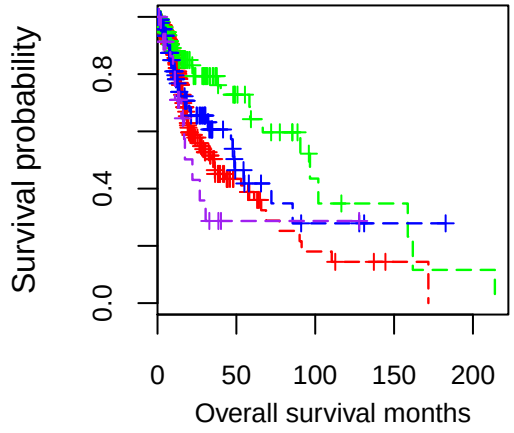
HNSC − results of xGeneModel

Mutation−based clustering

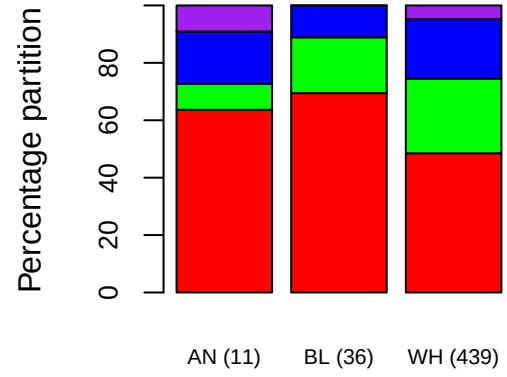


Cluster−survival association

Cox−PH p−value: 0.0017 (C2 vs others)

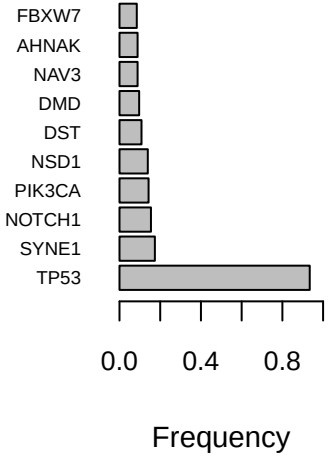


Cluster−race association

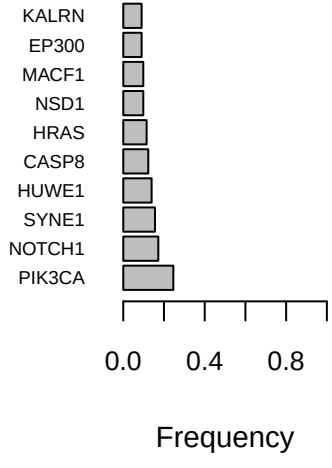


Fisher's test p−value: 0.1819

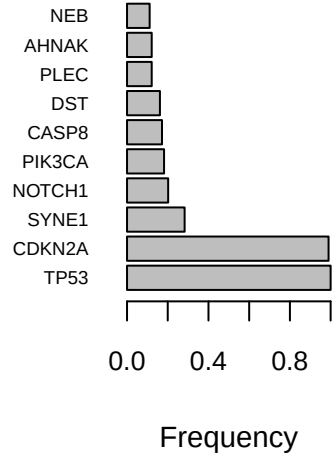
cluster 1



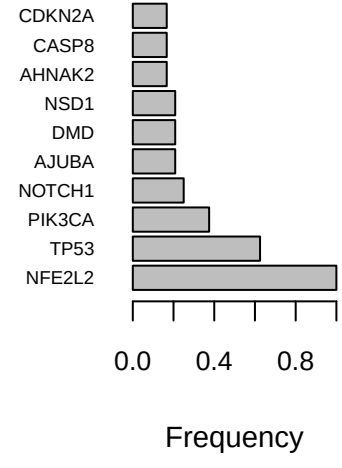
cluster 2



cluster 3

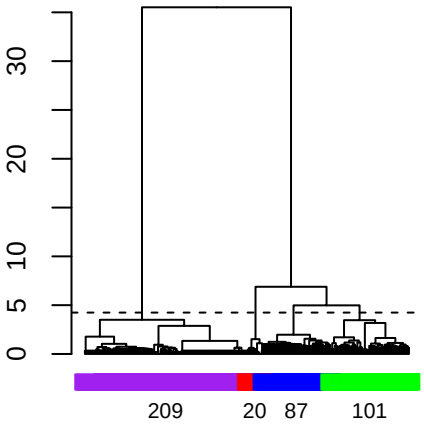


cluster 4



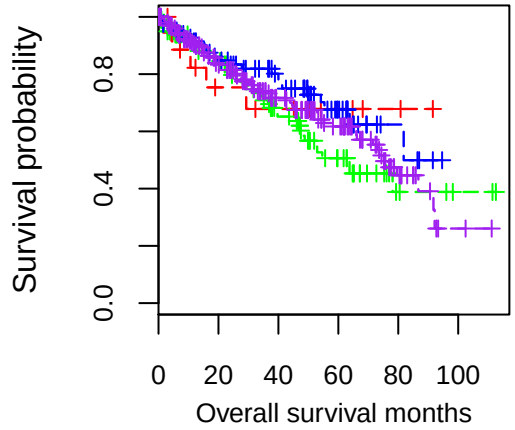
KIRC − results of xGeneModel

Mutation−based clustering

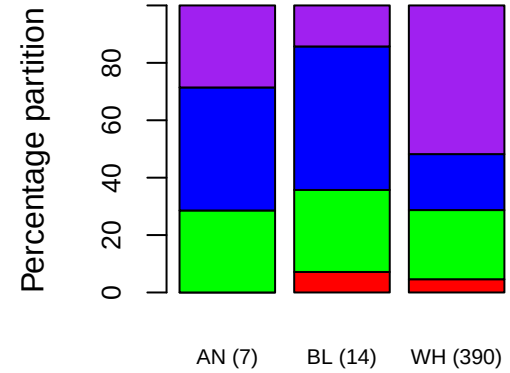


Cluster−survival association

Cox−PH p−value: 0.2785 (C2 vs others)

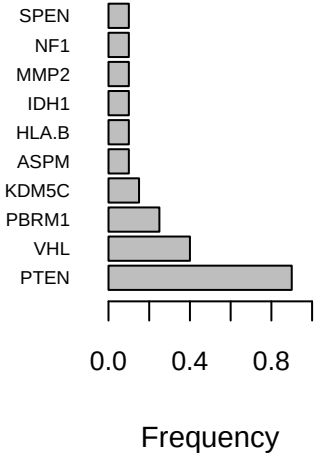


Cluster−race association

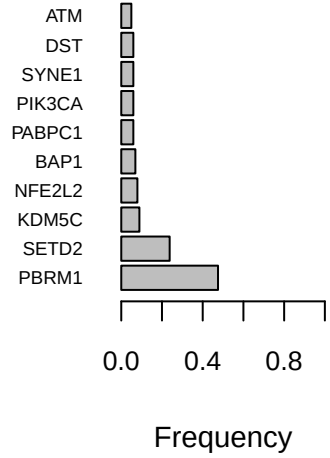


Fisher's test p−value: 0.022

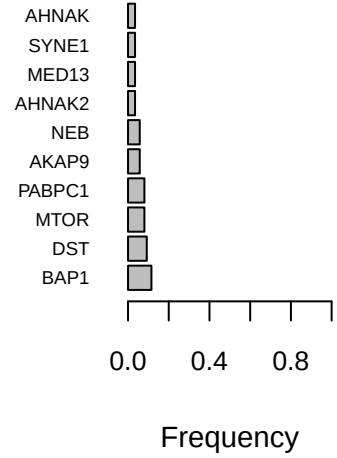
cluster 1



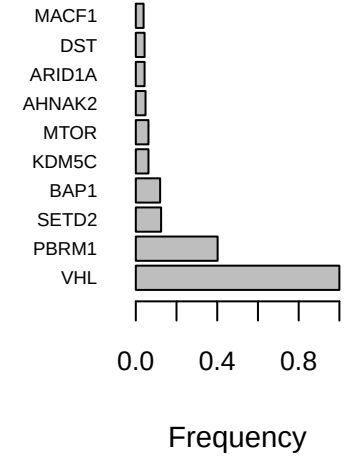
cluster 2



cluster 3

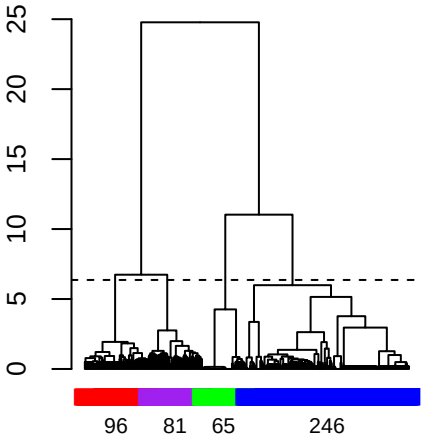


cluster 4



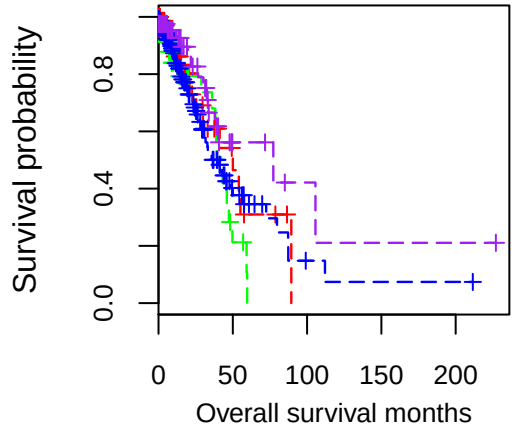
LUAD − results of xGeneModel

Mutation−based clustering

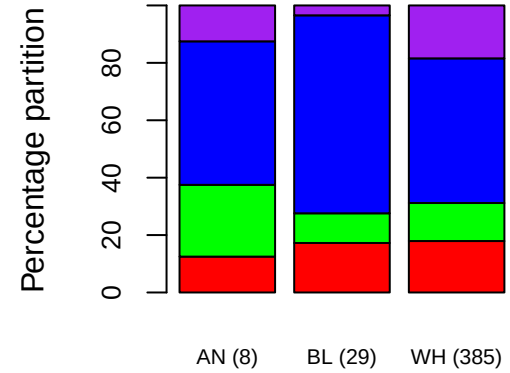


Cluster−survival association

Cox−PH p−value: 0.0148 (C4 vs others)

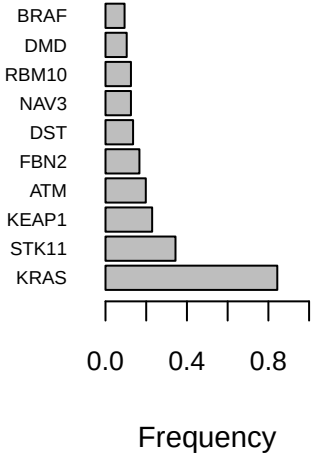


Cluster−race association

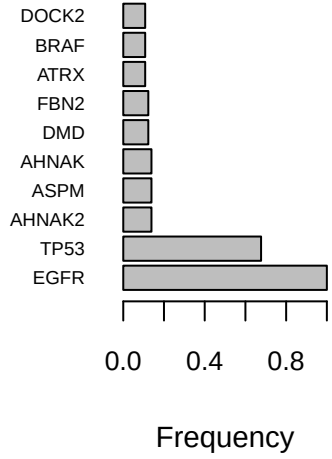


Fisher's test p−value: 0.2659

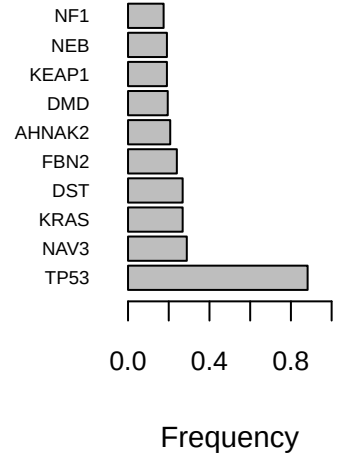
cluster 1



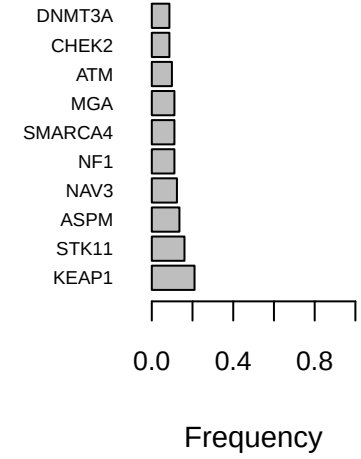
cluster 2



cluster 3

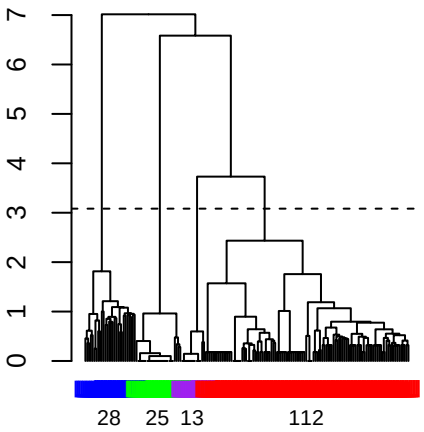


cluster 4



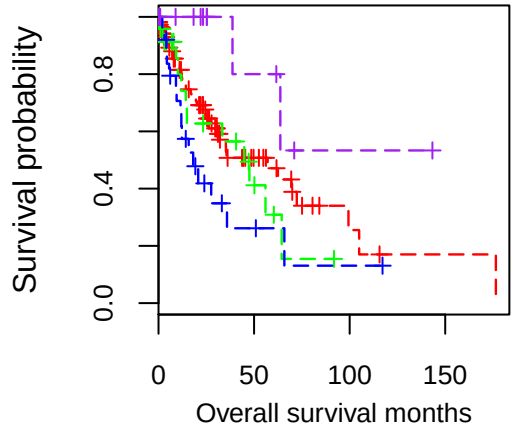
LUSC − results of xGeneModel

Mutation−based clustering

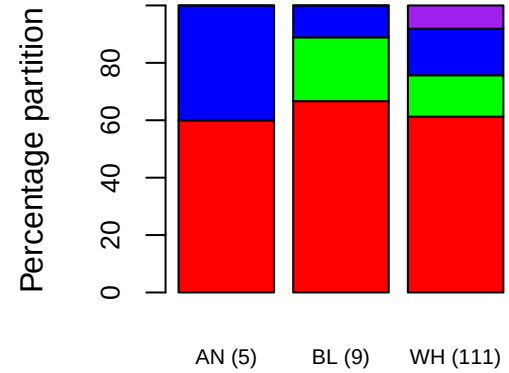


Cluster−survival association

Cox−PH p−value: 0.0115 (C3 vs others)

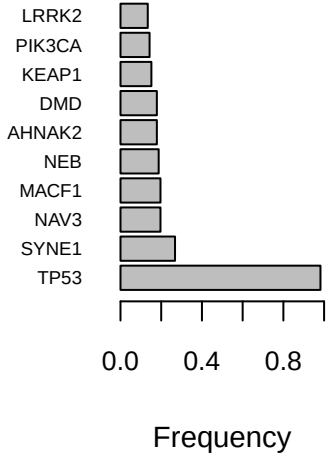


Cluster−race association

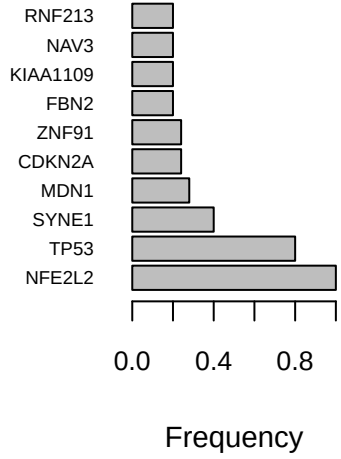


Fisher's test p−value: 0.8071

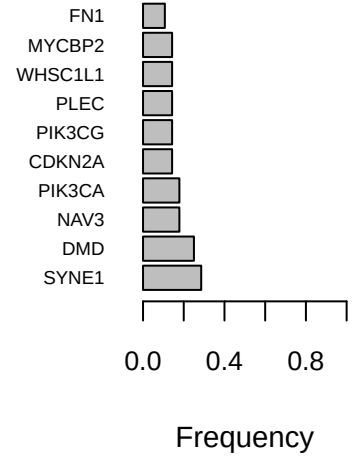
cluster 1



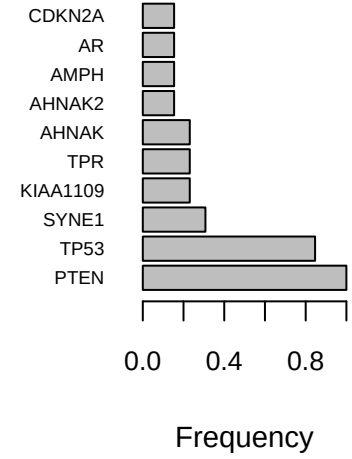
cluster 2



cluster 3

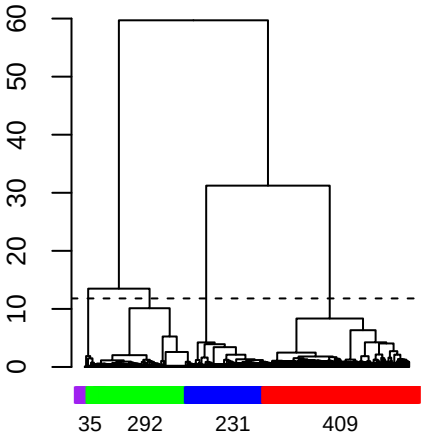


cluster 4



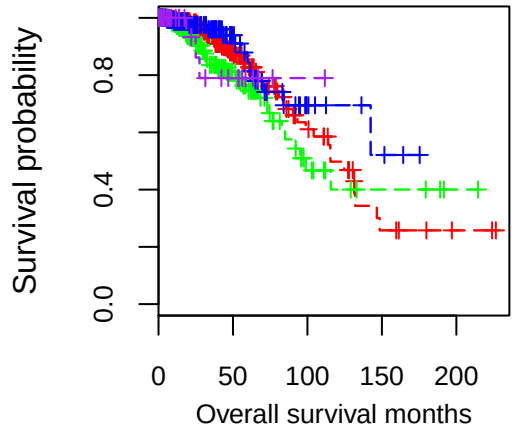
BRCA − results of xGeneModel

Mutation−based clustering

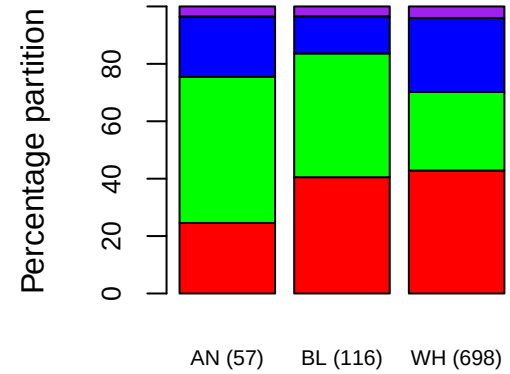


Cluster−survival association

Cox−PH p−value: 0.0027 (C2 vs others)

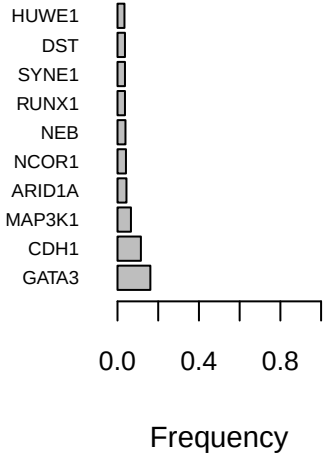


Cluster−race association

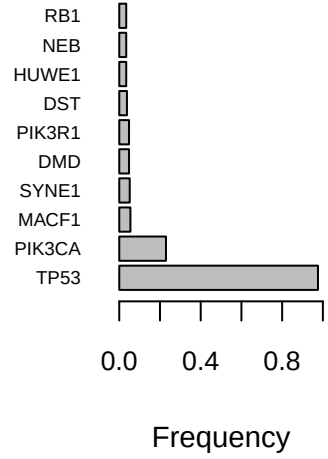


Fisher's test p−value: 0.0005

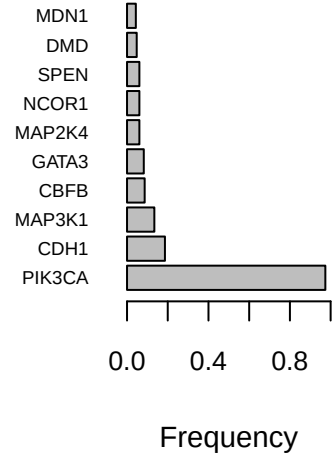
cluster 1



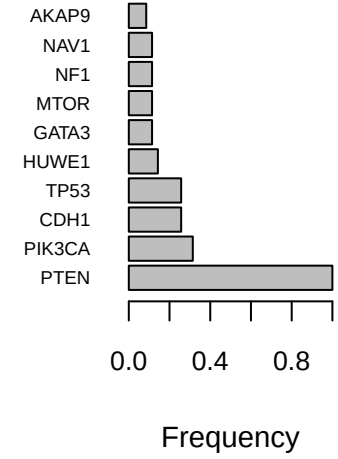
cluster 2



cluster 3

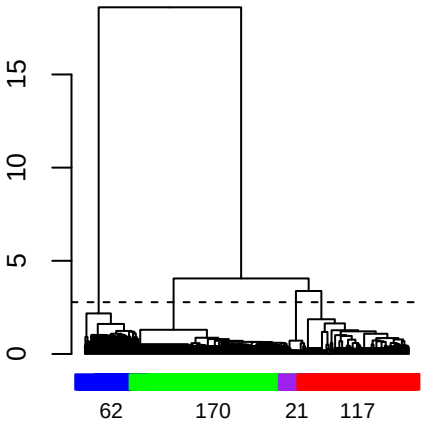


cluster 4



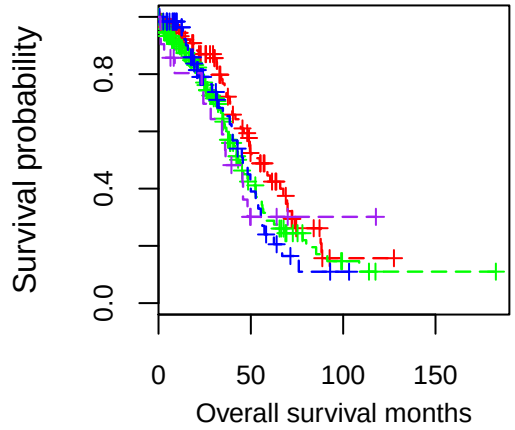
OV − results of xGeneModel

Mutation−based clustering

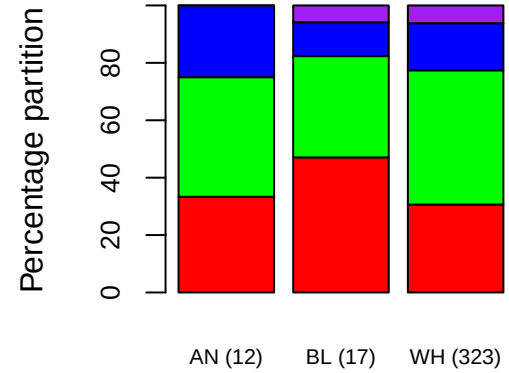


Cluster−survival association

Cox−PH p−value: 0.0158 (C1 vs others)

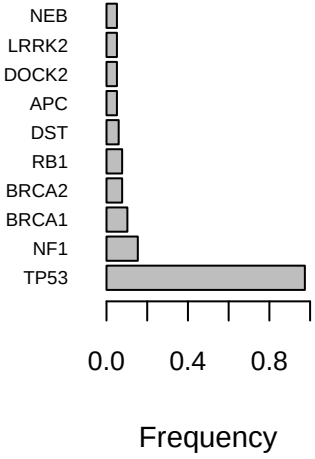


Cluster−race association

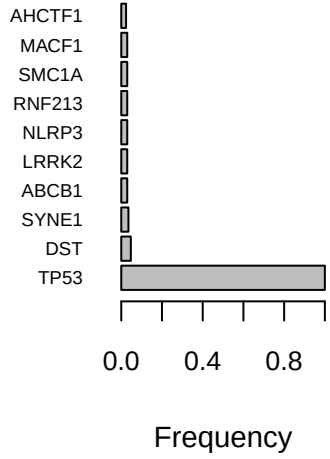


Fisher's test p−value: 0.7866

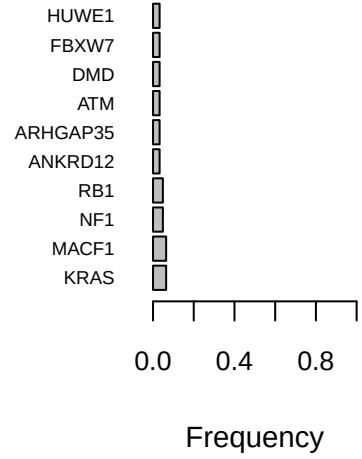
cluster 1



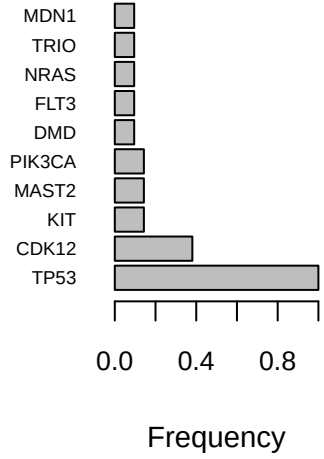
cluster 2



cluster 3

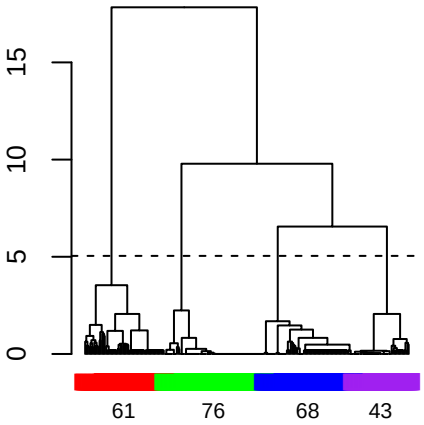


cluster 4



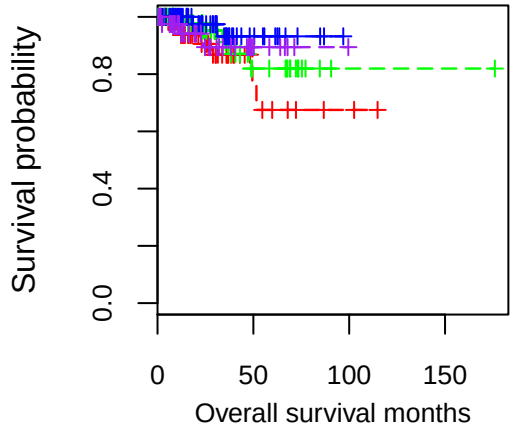
UCEC − results of xGeneModel

Mutation−based clustering

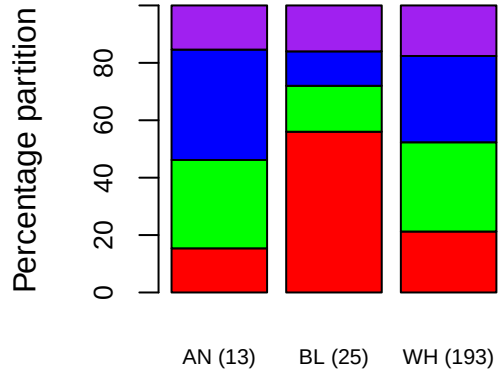


Cluster−survival association

Cox−PH p−value: 0.0862 (C1 vs others)

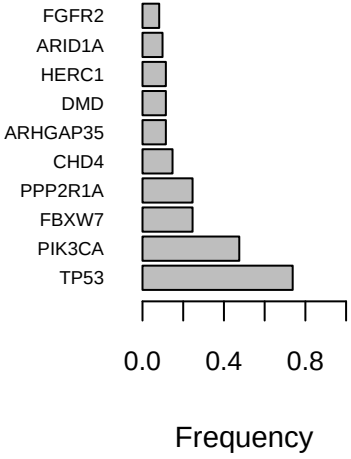


Cluster−race association

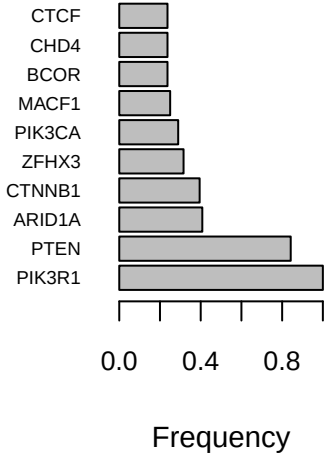


Fisher's test p−value: 0.023

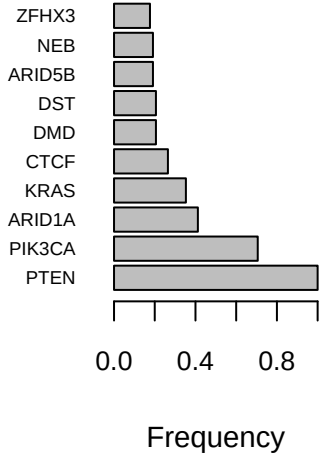
cluster 1



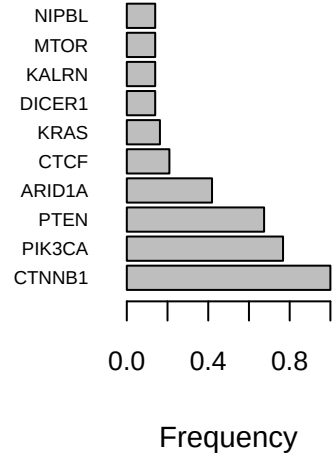
cluster 2



cluster 3

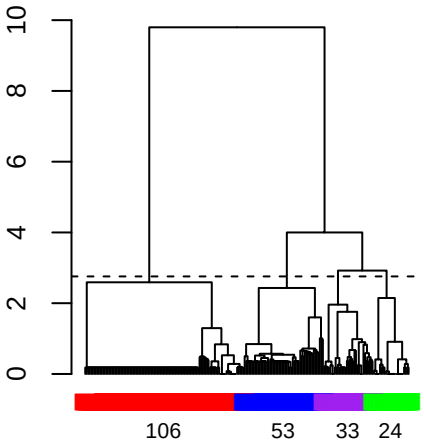


cluster 4



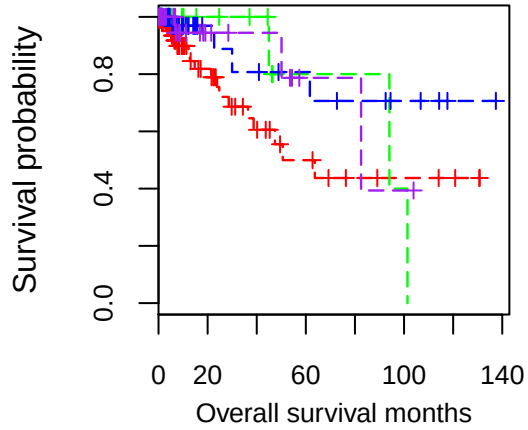
COAD − results of xGeneModel

Mutation−based clustering

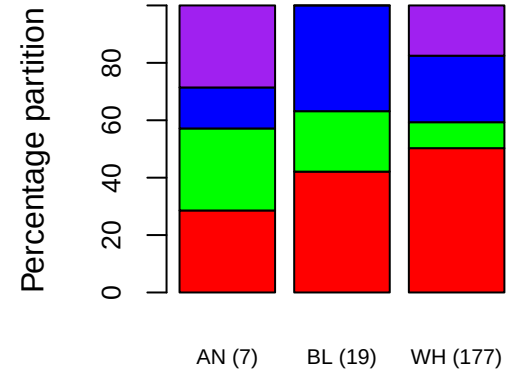


Cluster−survival association

Cox−PH p−value: 0.0394 (C1 vs others)

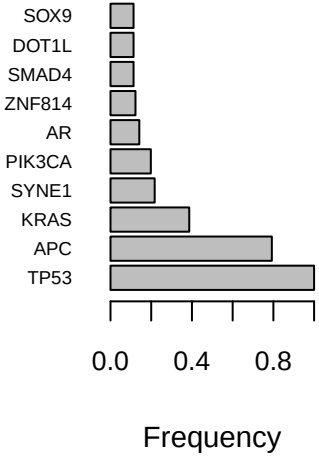


Cluster−race association

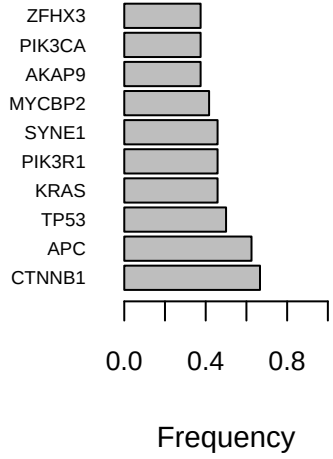


Fisher's test p−value: 0.0415

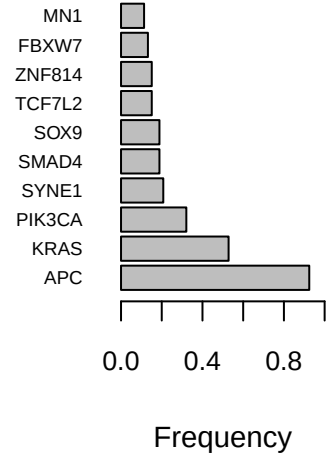
cluster 1



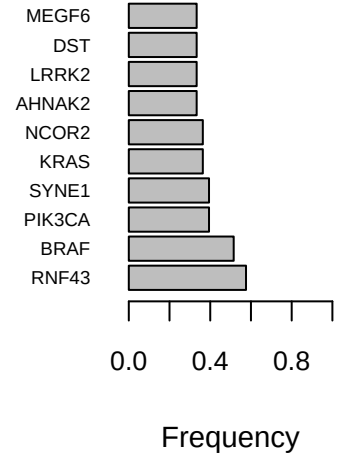
cluster 2



cluster 3

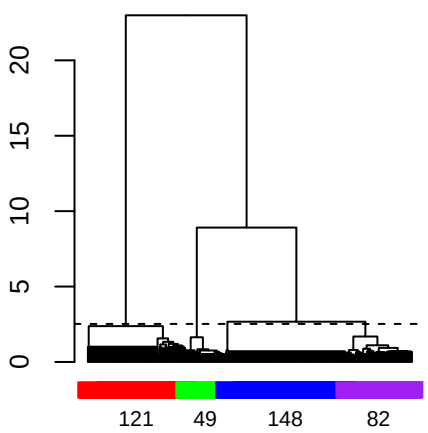


cluster 4



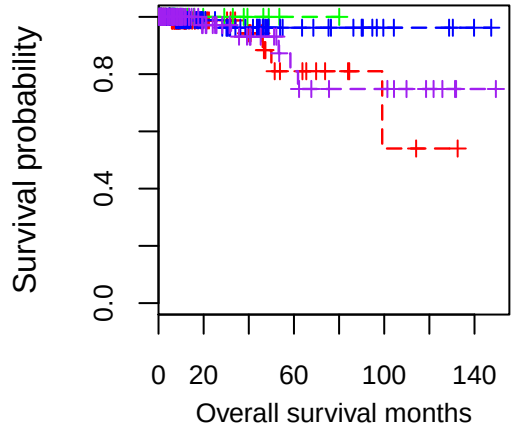
THCA − results of xGeneModel

Mutation−based clustering

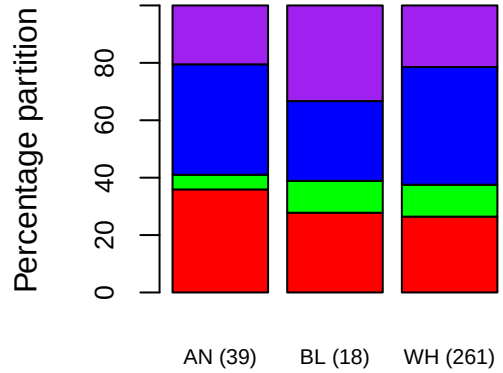


Cluster−survival association

Cox−PH p−value: 0.5565 (C1,4 vs others)

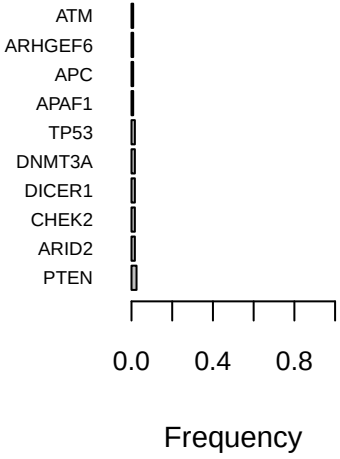


Cluster−race association

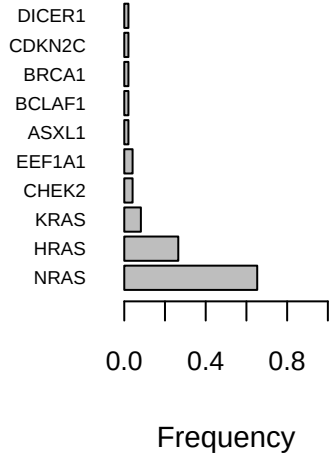


Fisher's test p−value: 0.6442

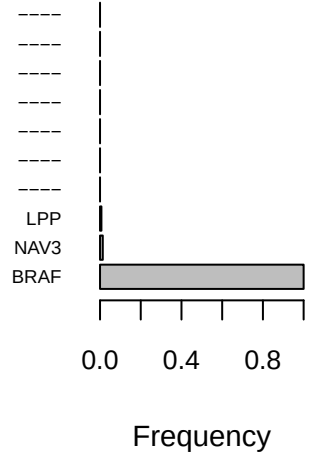
cluster 1



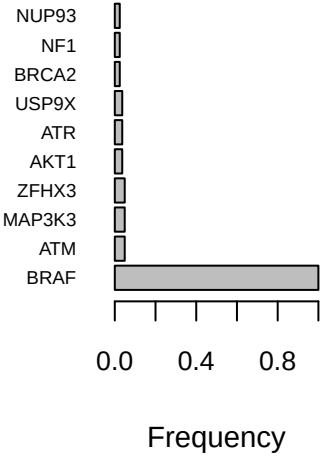
cluster 2



cluster 3

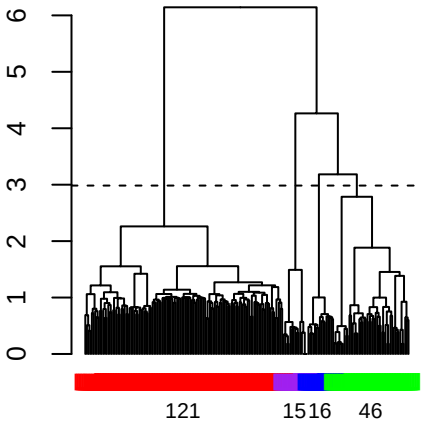


cluster 4



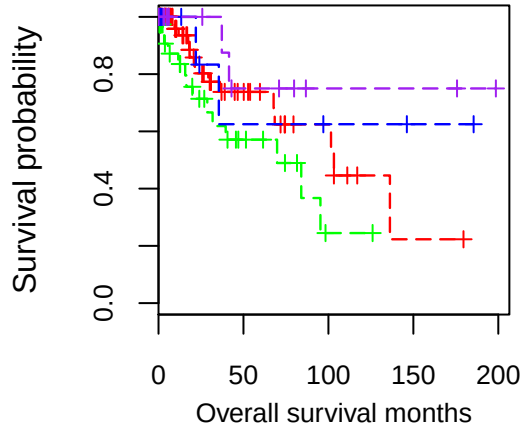
CESC − results of xGeneModel

Mutation−based clustering

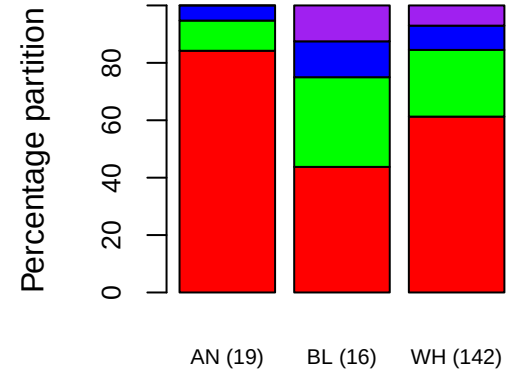


Cluster−survival association

Cox−PH p−value: 0.0329 (C2 vs others)

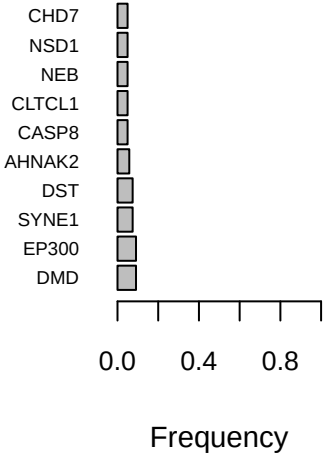


Cluster−race association

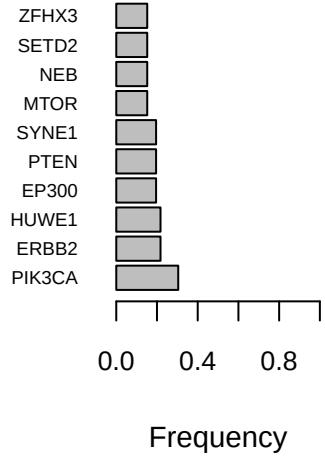


Fisher's test p−value: 0.2879

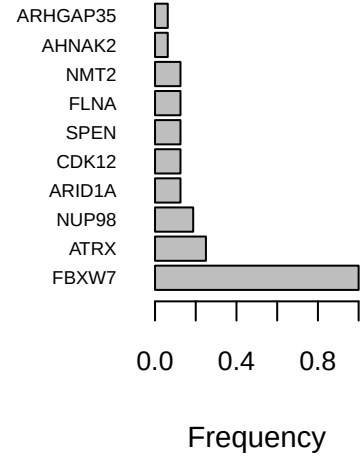
cluster 1



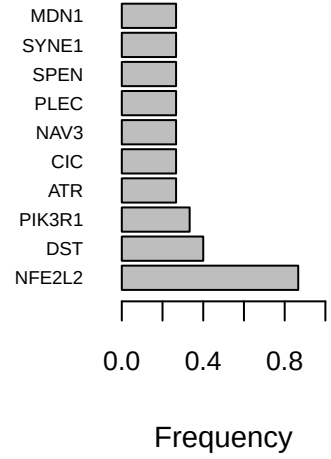
cluster 2



cluster 3

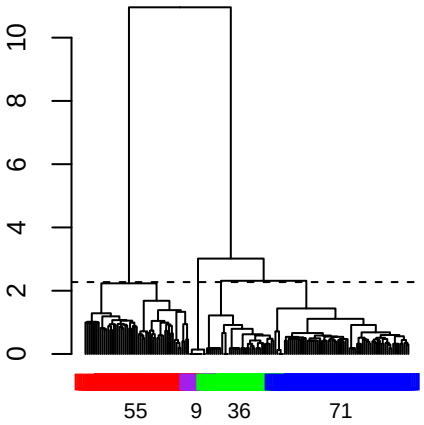


cluster 4



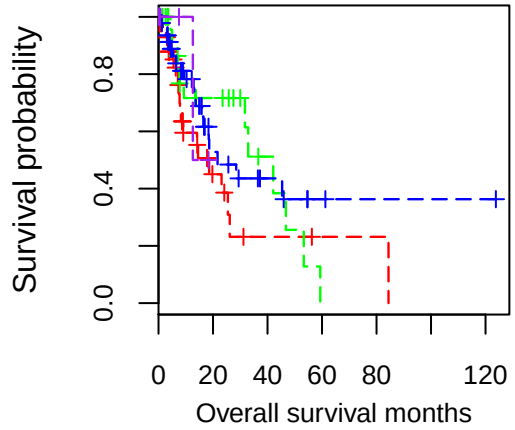
ESCA − results of xGeneModel

Mutation−based clustering

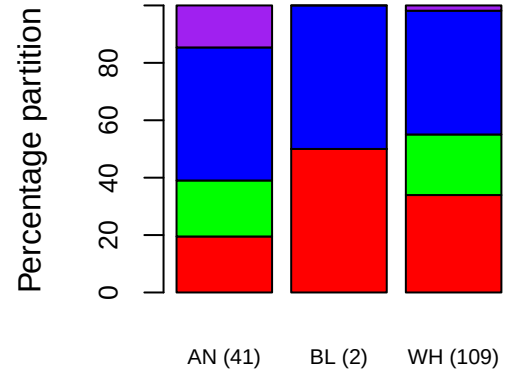


Cluster−survival association

Cox−PH p−value: 0.0716 (C1 vs others)

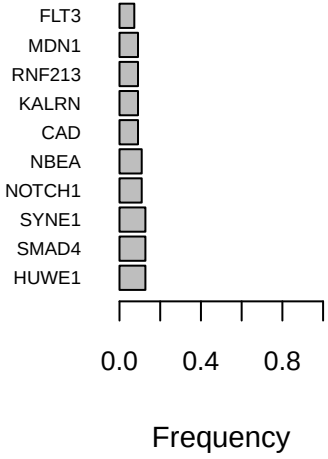


Cluster−race association

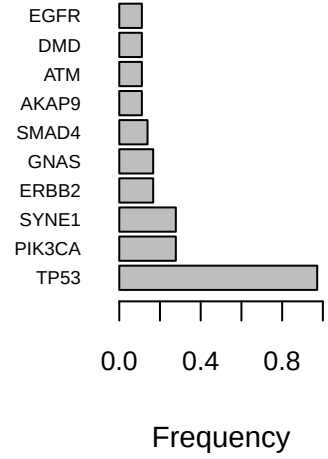


Fisher's test p−value: 0.033

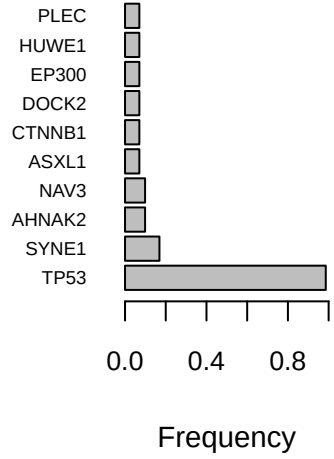
cluster 1



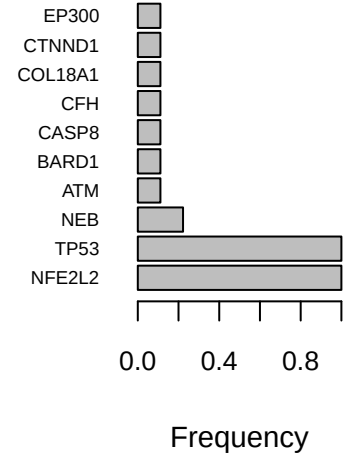
cluster 2



cluster 3

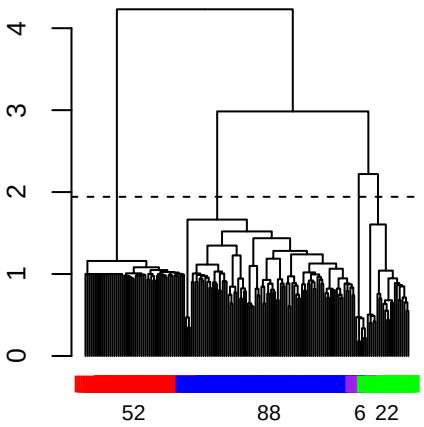


cluster 4



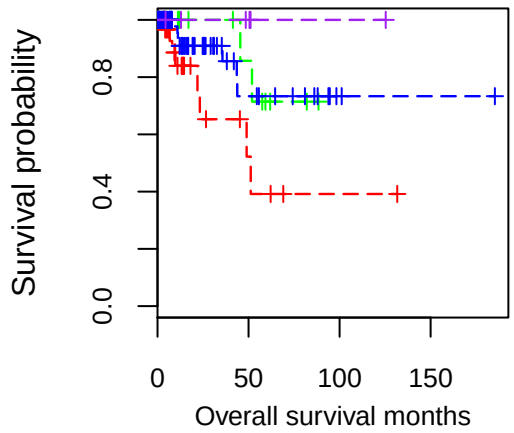
KIRP − results of xGeneModel

Mutation−based clustering

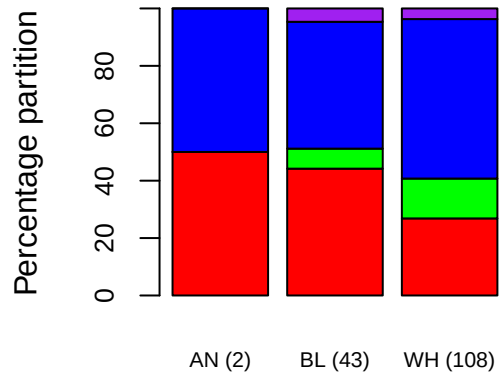


Cluster−survival association

Cox−PH p−value: 0.0162 (C1 vs others)

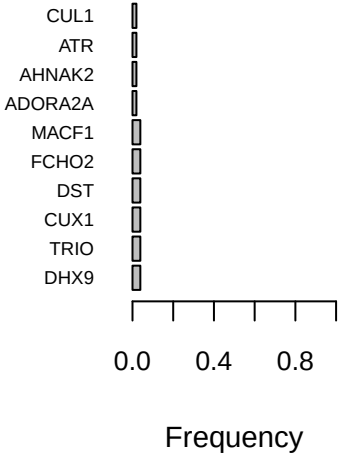


Cluster−race association

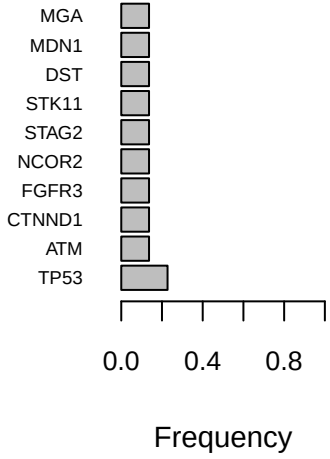


Fisher's test p−value: 0.3333

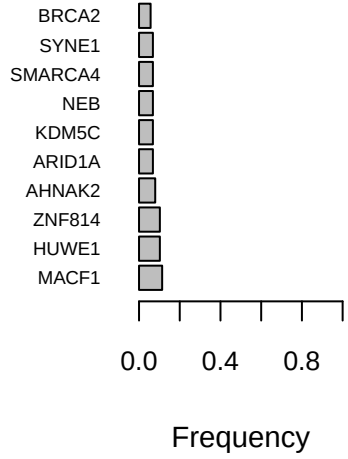
cluster 1



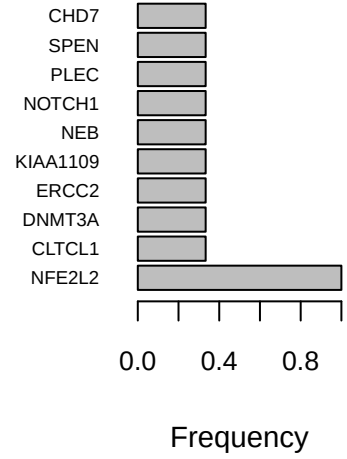
cluster 2



cluster 3

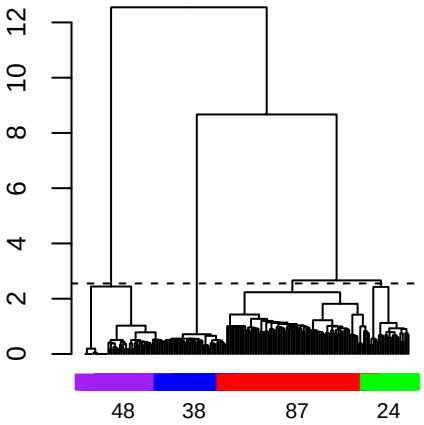


cluster 4



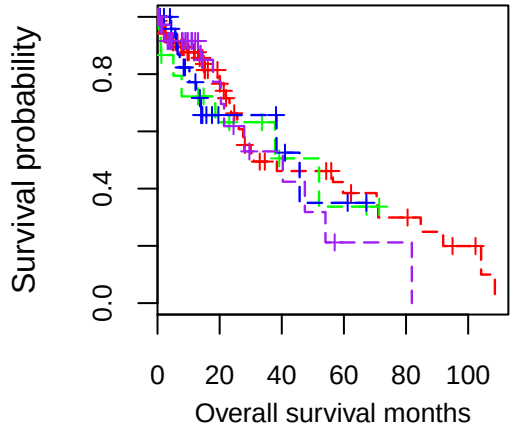
LIHC − results of xGeneModel

Mutation−based clustering

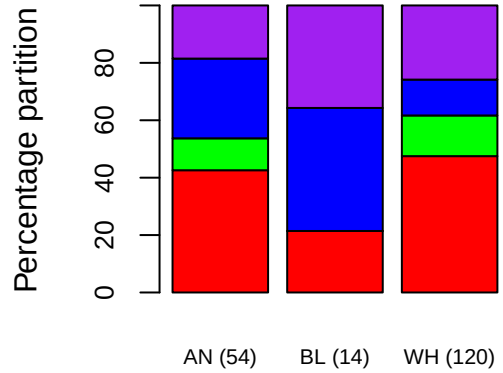


Cluster−survival association

Cox−PH p−value: 0.4364 (C1 vs others)

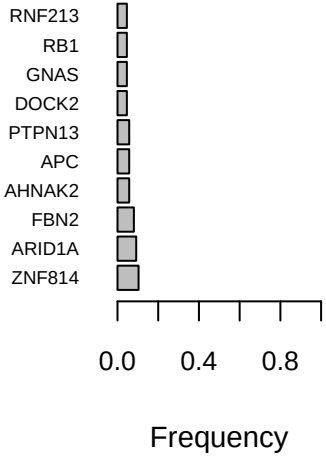


Cluster−race association

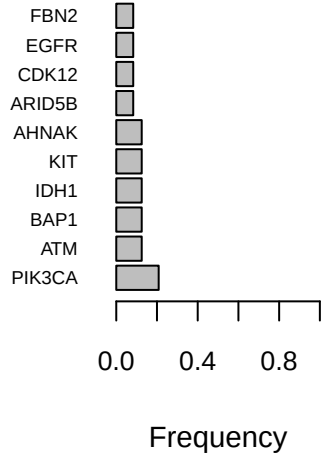


Fisher's test p−value: 0.023

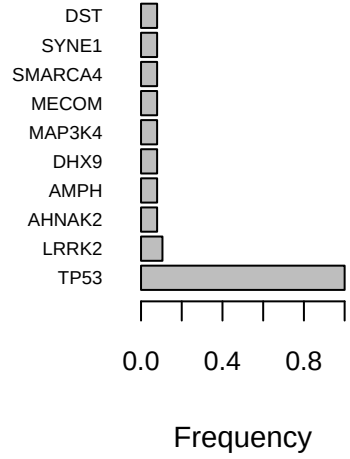
cluster 1



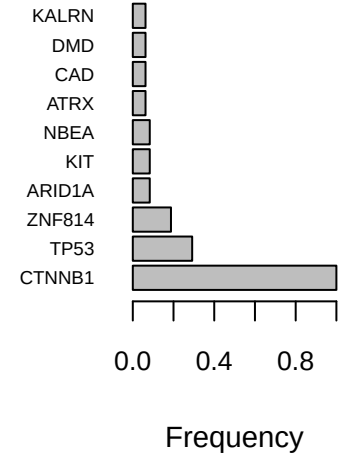
cluster 2



cluster 3

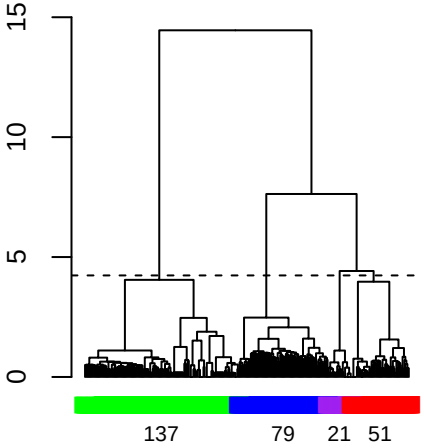


cluster 4



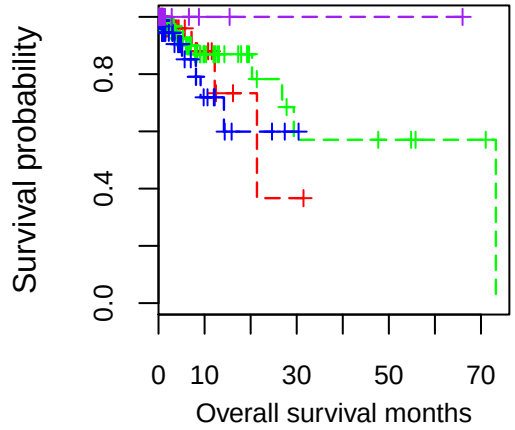
STAD − results of xGeneModel

Mutation−based clustering

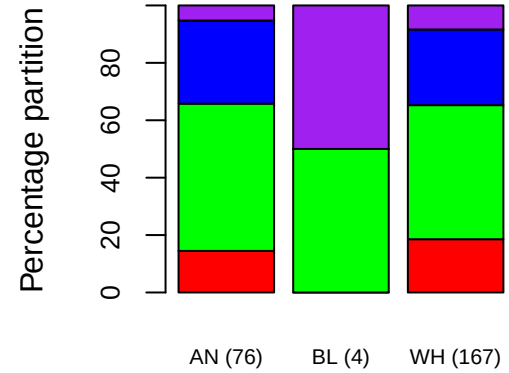


Cluster−survival association

Cox−PH p−value: 0.1443 (C3 vs others)

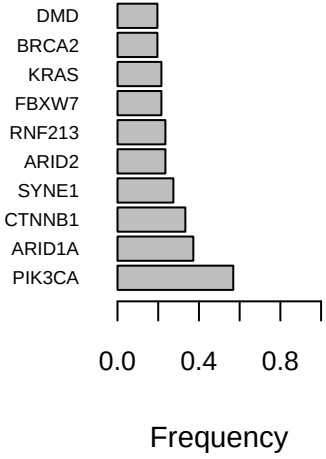


Cluster−race association

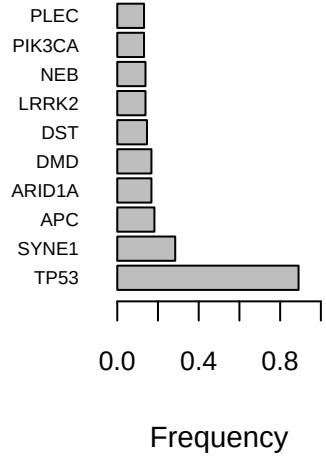


Fisher's test p−value: 0.2014

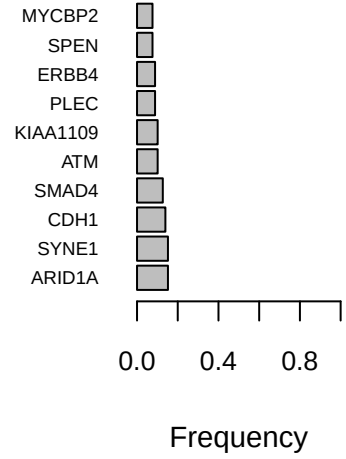
cluster 1



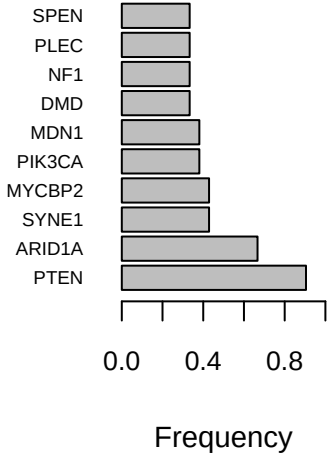
cluster 2



cluster 3

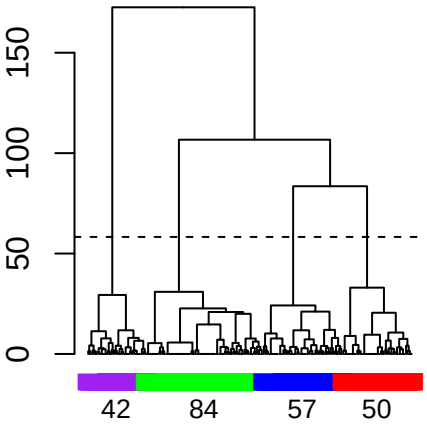


cluster 4



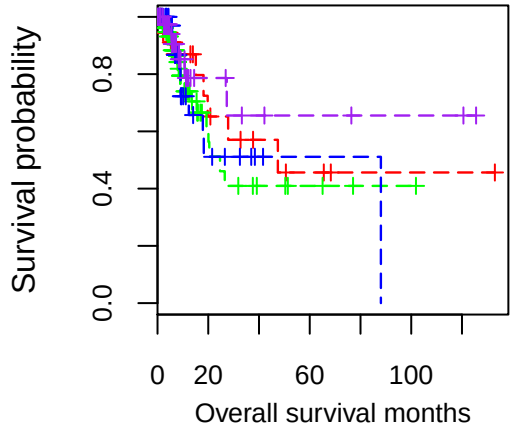
BLCA − results of ccpwModel

Mutation−based clustering

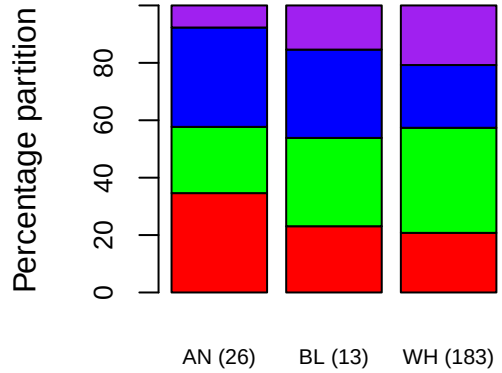


Cluster−survival association

Cox−PH p−value: 0.1921 (C4 vs others)

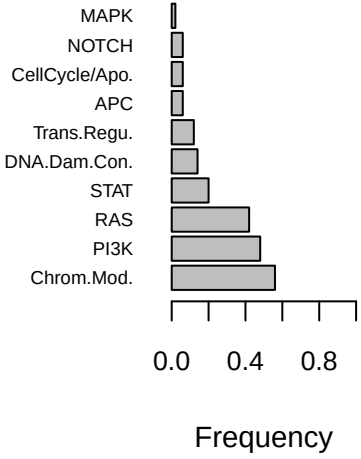


Cluster−race association

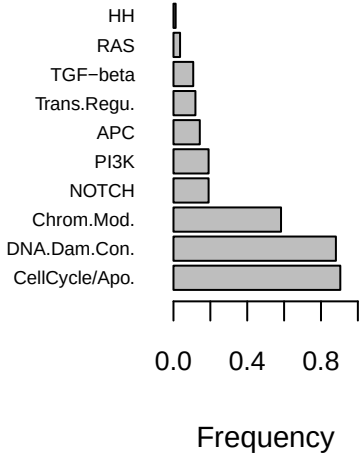


Fisher's test p−value: 0.2884

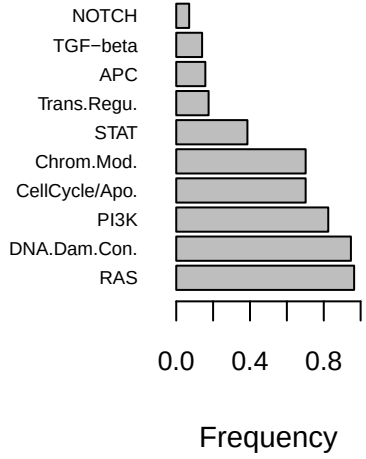
cluster 1



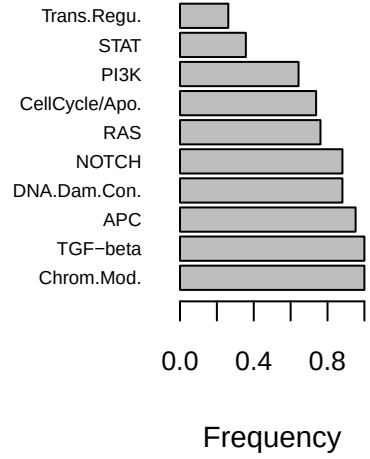
cluster 2



cluster 3

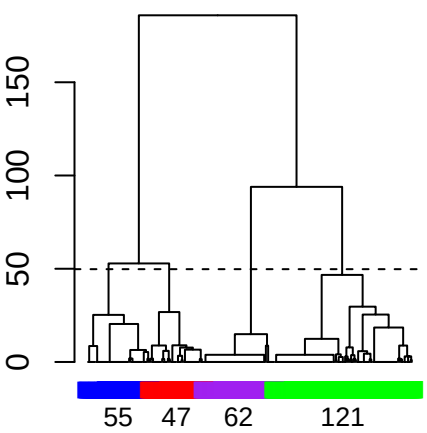


cluster 4



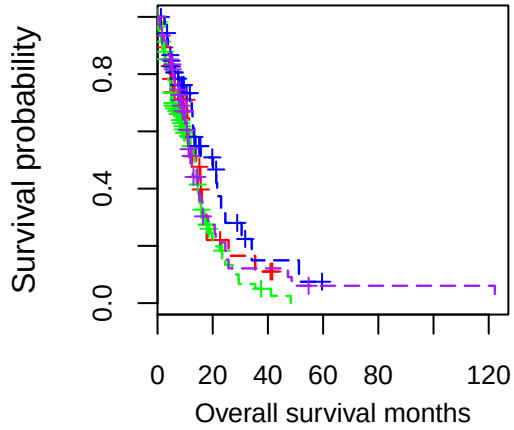
GBM − results of ccpwModel

Mutation−based clustering

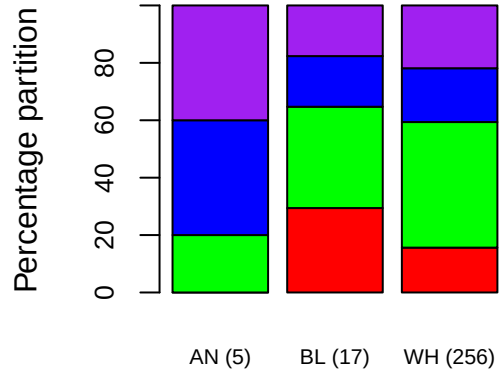


Cluster−survival association

Cox−PH p−value: 0.1261 (C1,3 vs others)

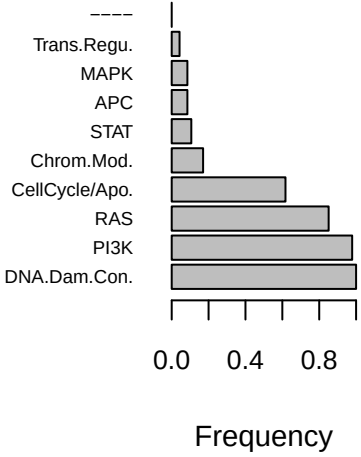


Cluster−race association

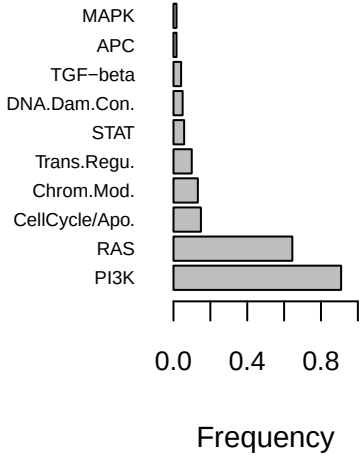


Fisher's test p−value: 0.4478

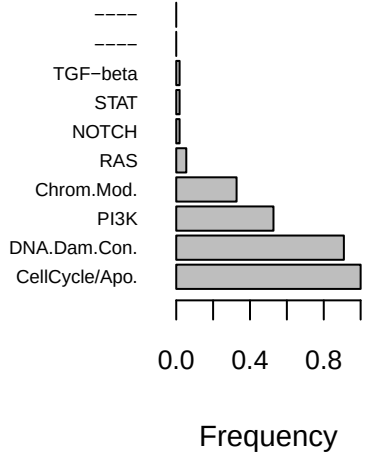
cluster 1



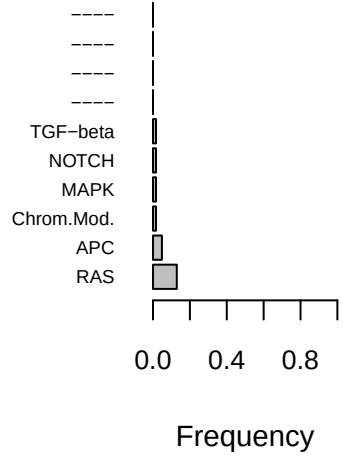
cluster 2



cluster 3

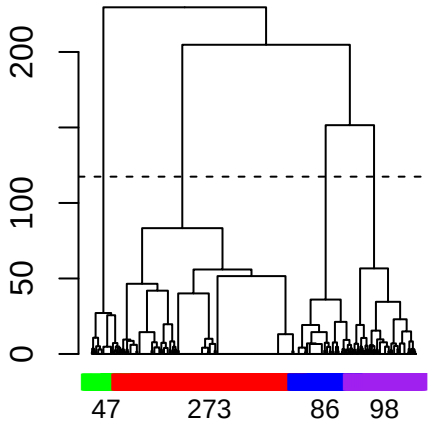


cluster 4



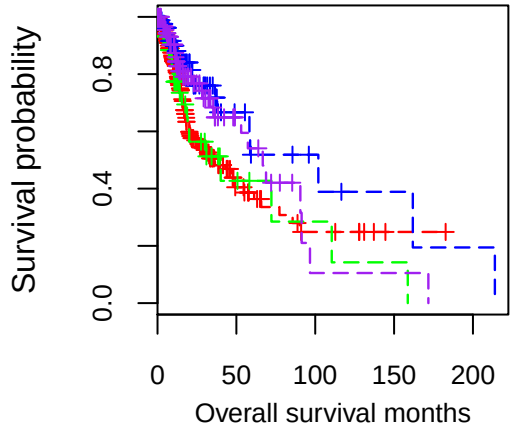
HNSC − results of ccpwModel

Mutation−based clustering

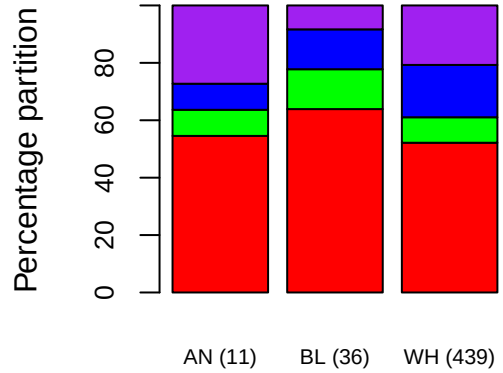


Cluster−survival association

Cox−PH p−value: 0.0058 (C1,2 vs others)

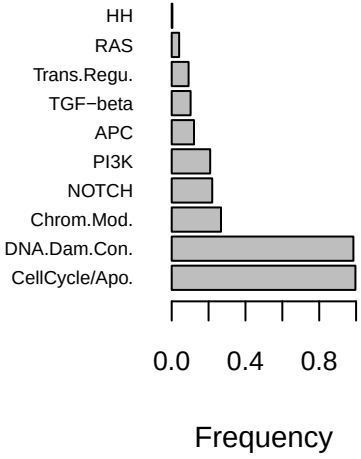


Cluster−race association

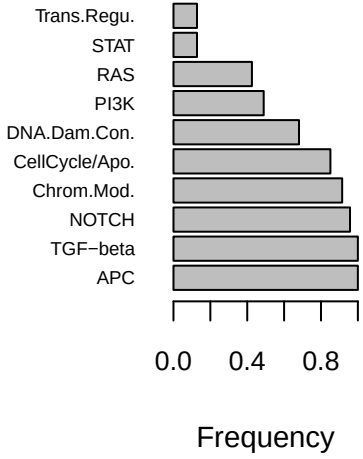


Fisher's test p−value: 0.4173

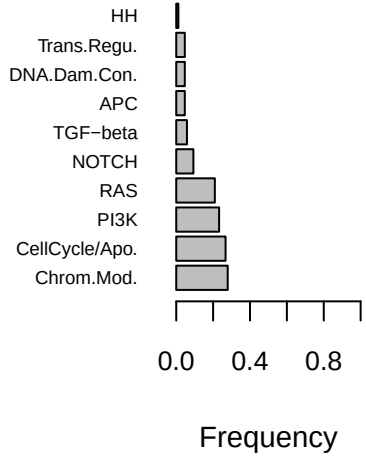
cluster 1



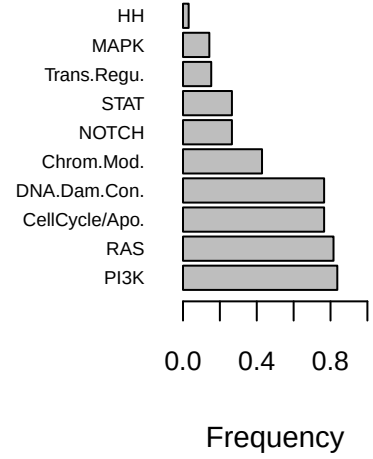
cluster 2



cluster 3

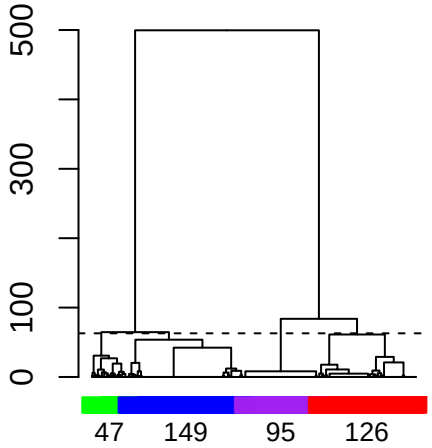


cluster 4



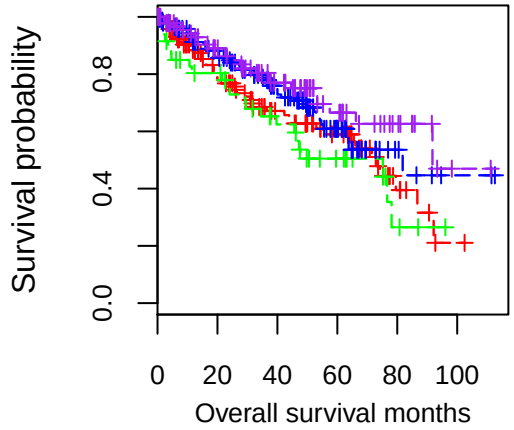
KIRC − results of ccpwModel

Mutation−based clustering

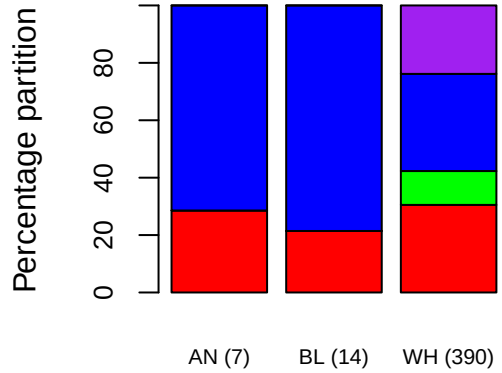


Cluster−survival association

Cox−PH p−value: 0.017 (C1,2 vs others)

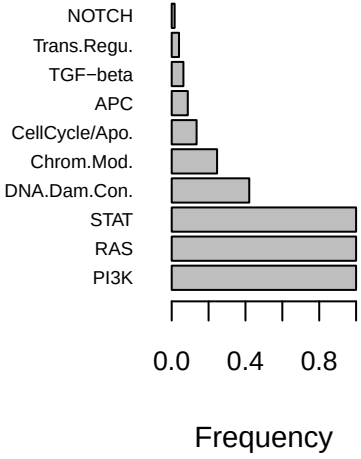


Cluster−race association

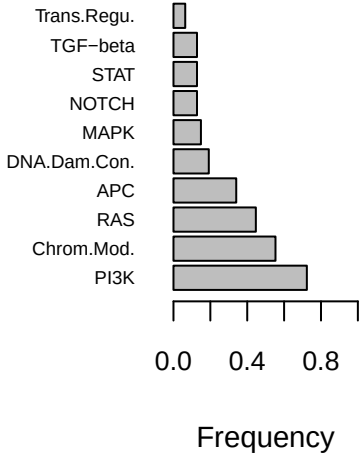


Fisher's test p−value: 0.008

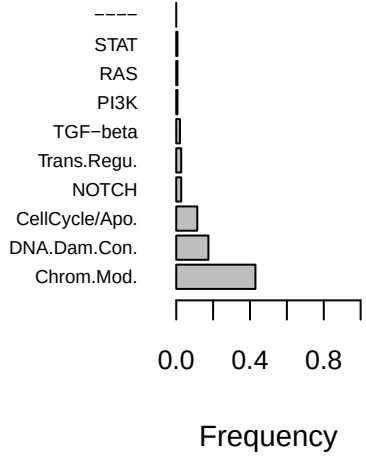
cluster 1



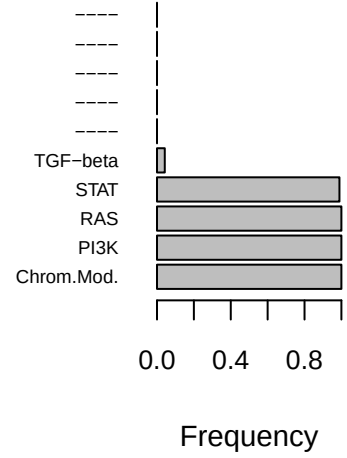
cluster 2



cluster 3

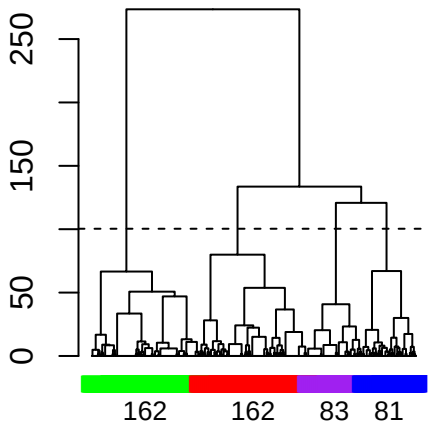


cluster 4



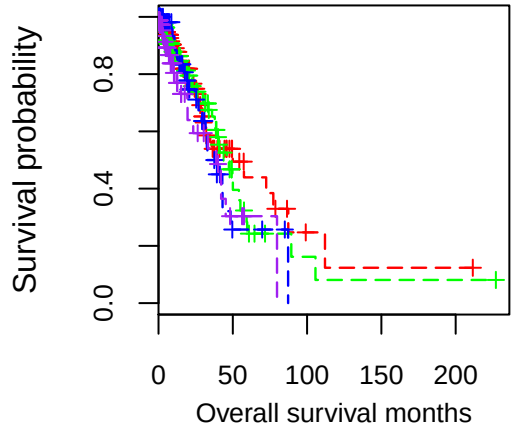
LUAD − results of ccpwModel

Mutation−based clustering

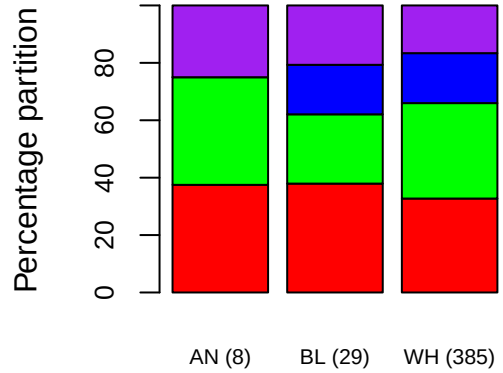


Cluster−survival association

Cox−PH p−value: 0.0687 (C1,2 vs others)

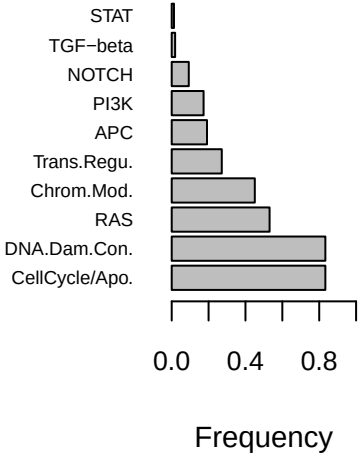


Cluster−race association

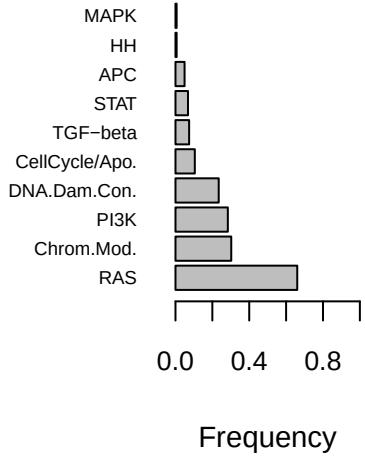


Fisher's test p−value: 0.8136

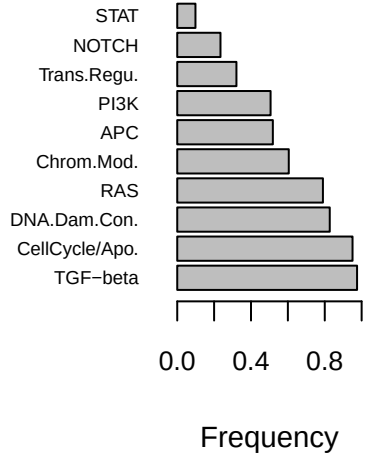
cluster 1



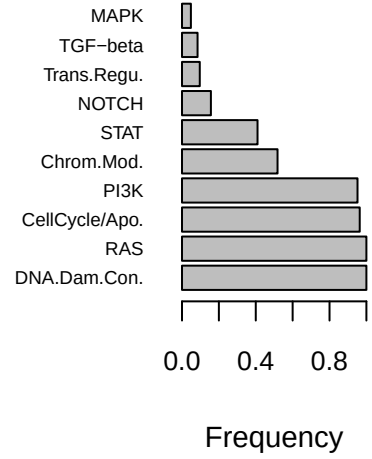
cluster 2



cluster 3

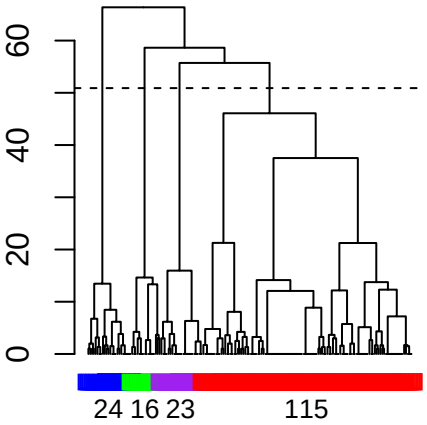


cluster 4



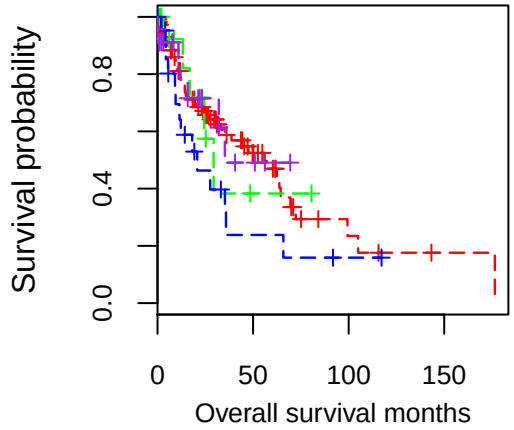
LUSC − results of ccpwModel

Mutation−based clustering

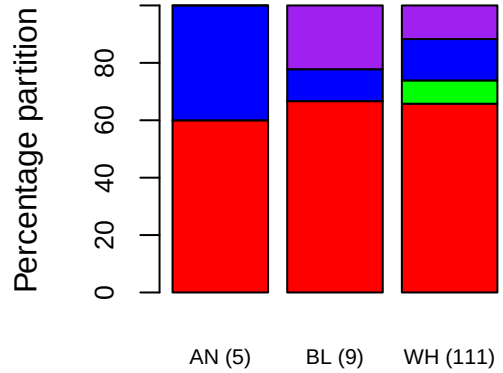


Cluster−survival association

Cox−PH p−value: 0.051 (C3 vs others)

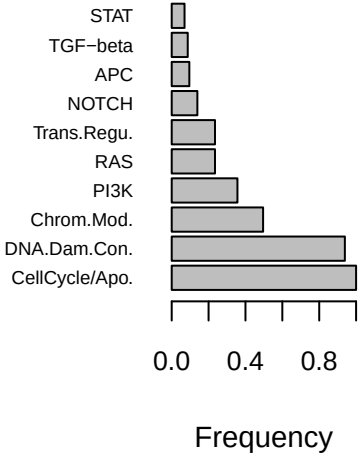


Cluster−race association

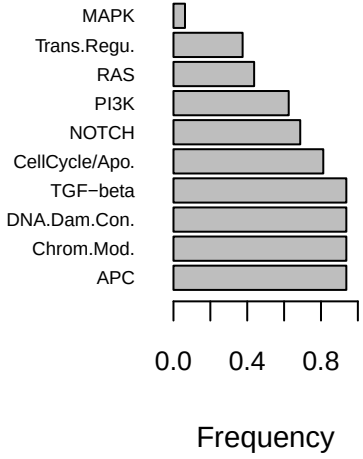


Fisher's test p−value: 0.6742

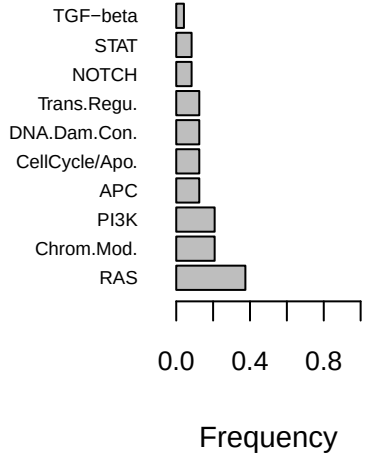
cluster 1



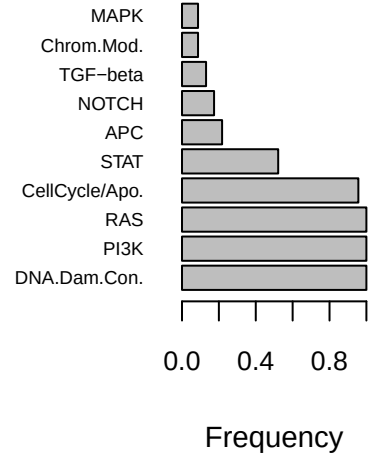
cluster 2



cluster 3

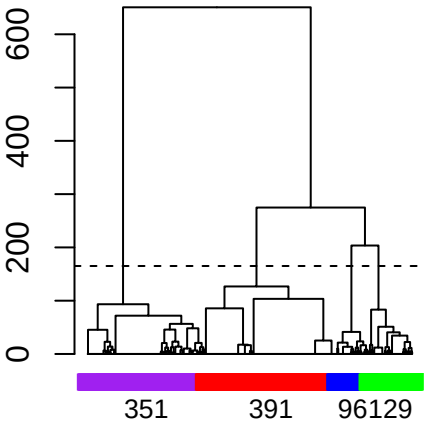


cluster 4



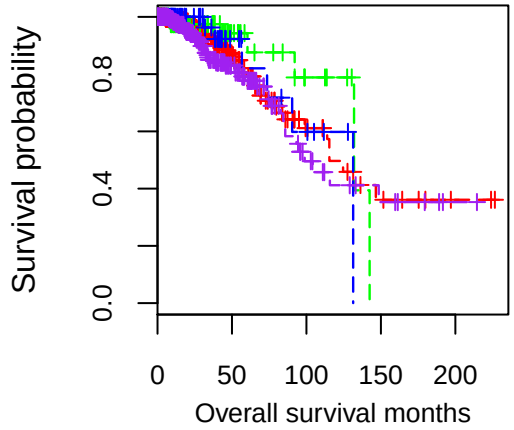
BRCA − results of ccpwModel

Mutation−based clustering

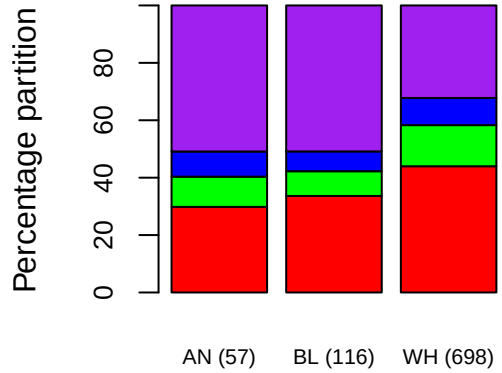


Cluster−survival association

Cox−PH p−value: 0.0233 (C4 vs others)

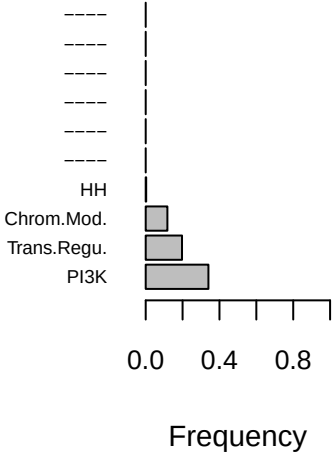


Cluster−race association

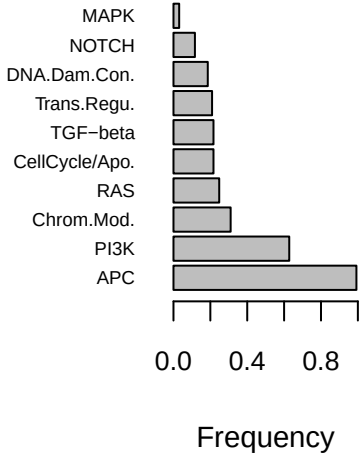


Fisher's test p−value: 0.0025

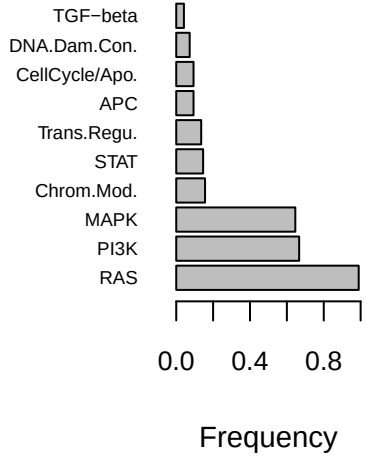
cluster 1



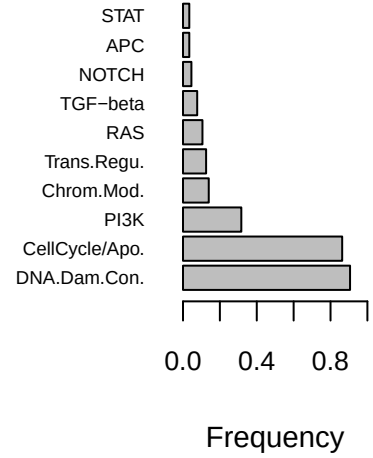
cluster 2



cluster 3

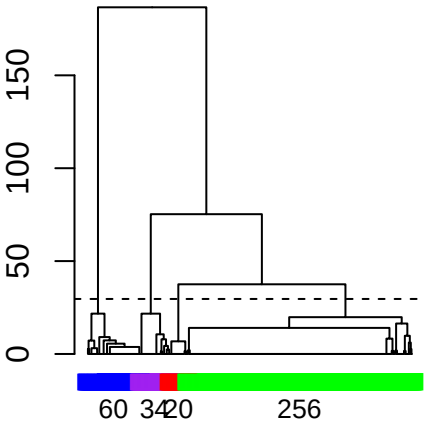


cluster 4



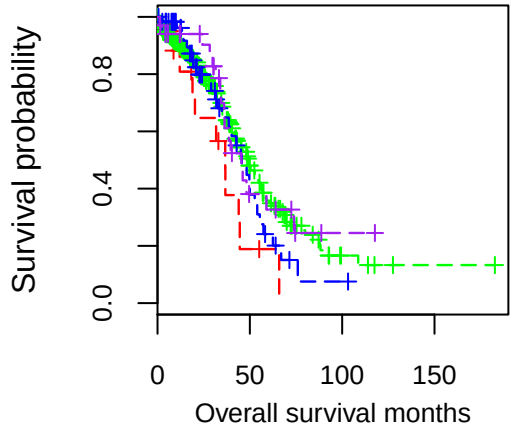
OV − results of ccpwModel

Mutation−based clustering

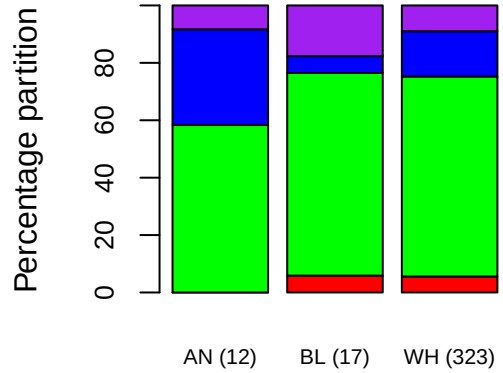


Cluster−survival association

Cox−PH p−value: 0.0742 (C1 vs others)

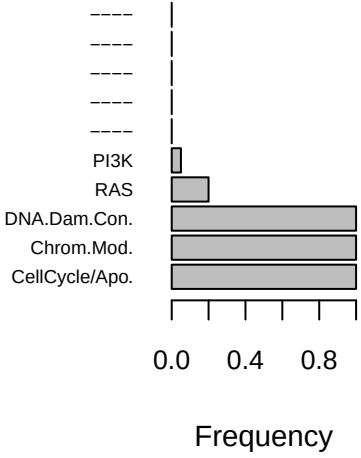


Cluster−race association

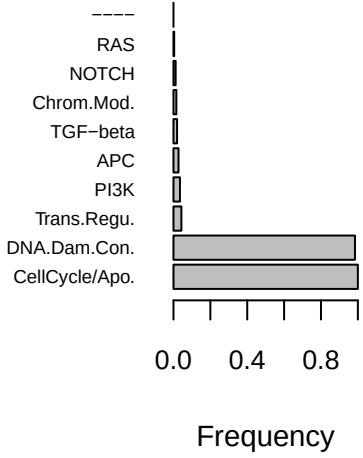


Fisher's test p−value: 0.4303

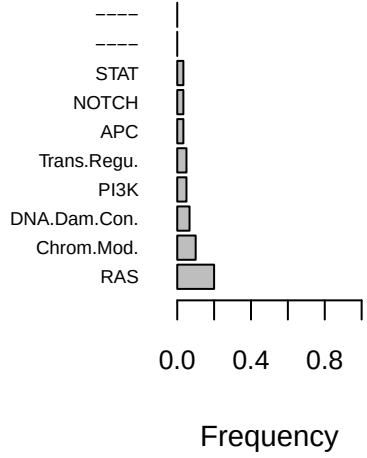
cluster 1



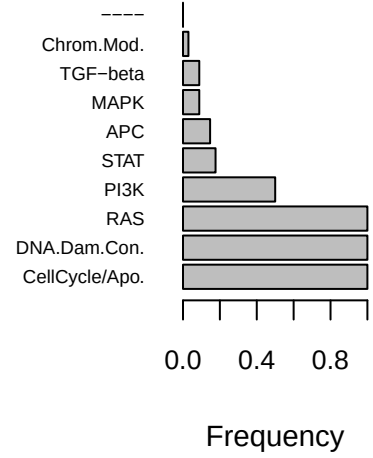
cluster 2



cluster 3

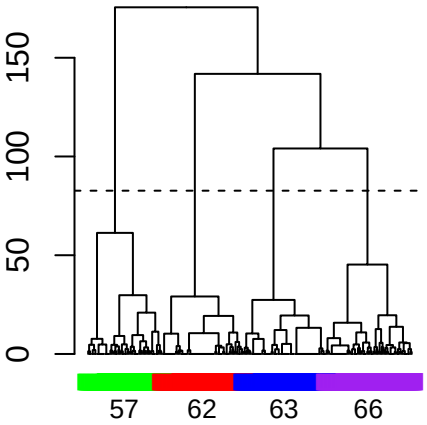


cluster 4



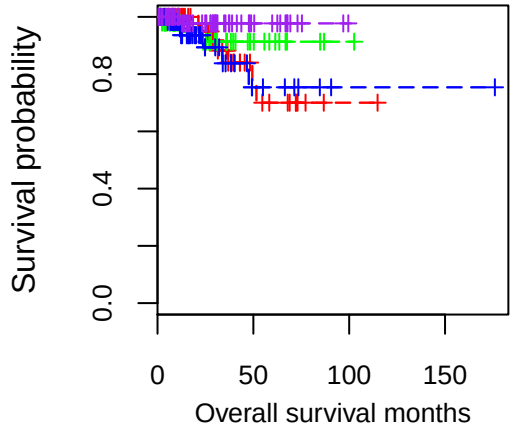
UCEC − results of ccpwModel

Mutation−based clustering

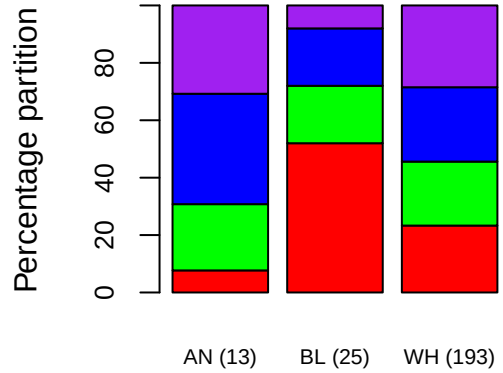


Cluster−survival association

Cox−PH p−value: 0.0417 (C1,3 vs others)

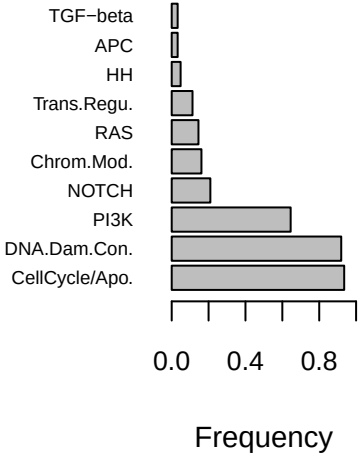


Cluster−race association

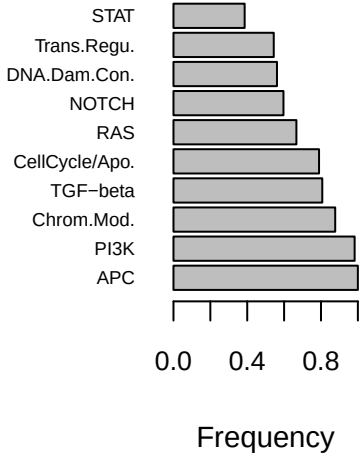


Fisher's test p−value: 0.0375

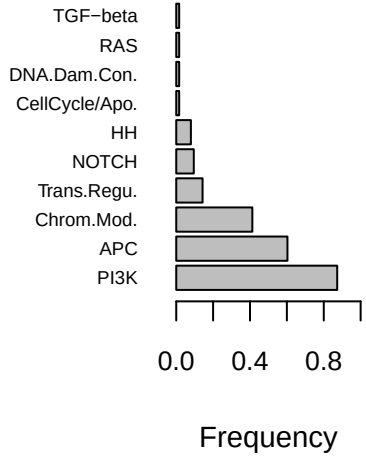
cluster 1



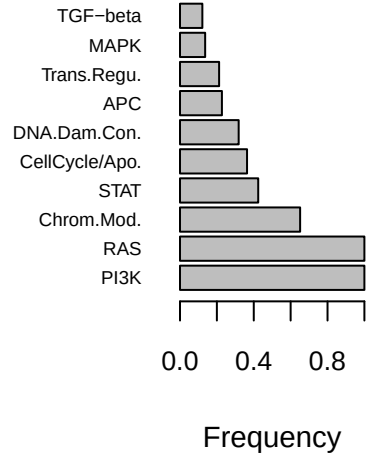
cluster 2



cluster 3

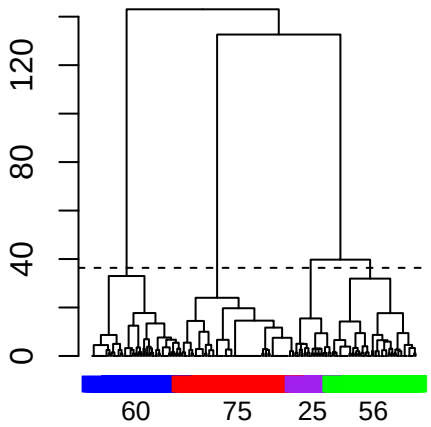


cluster 4



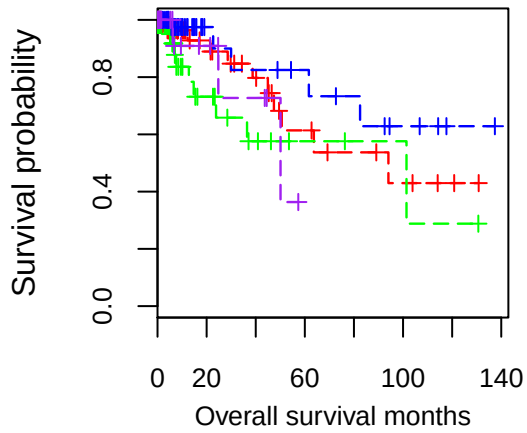
COAD − results of ccpwModel

Mutation−based clustering

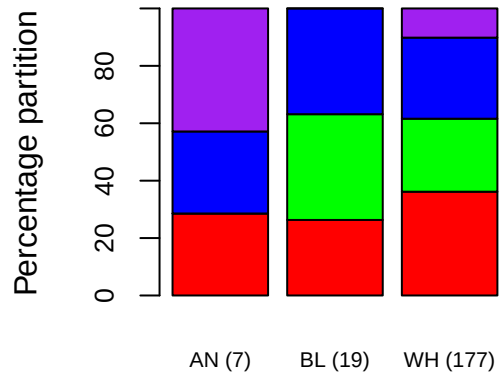


Cluster−survival association

Cox−PH p−value: 0.0368 (C2 vs others)

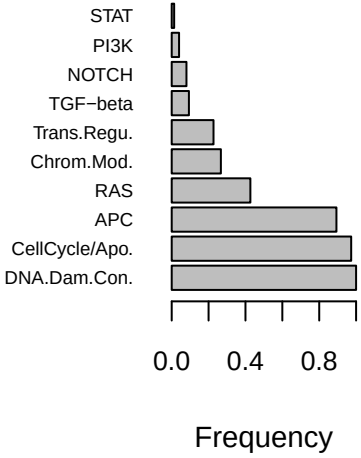


Cluster−race association

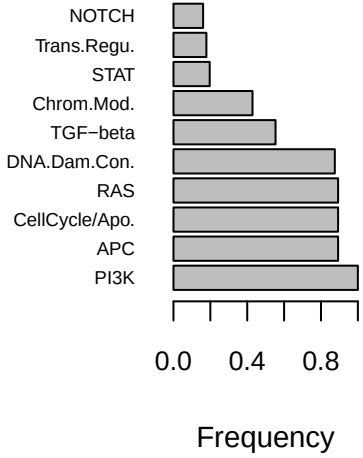


Fisher's test p−value: 0.0715

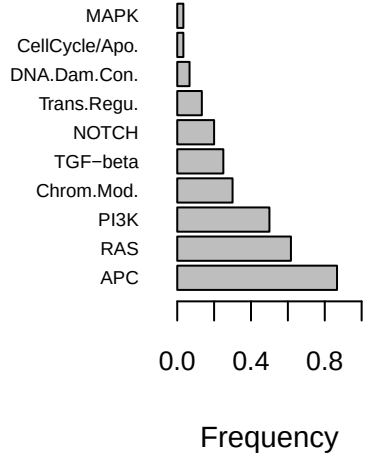
cluster 1



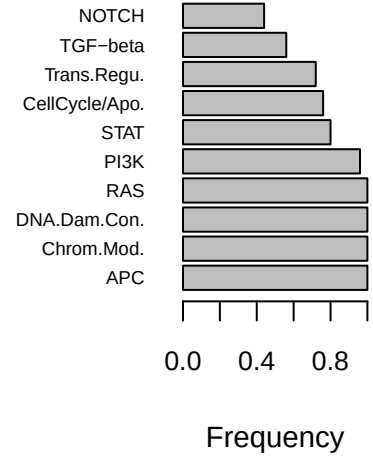
cluster 2



cluster 3

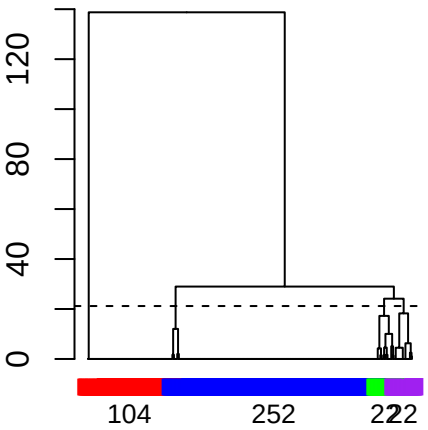


cluster 4



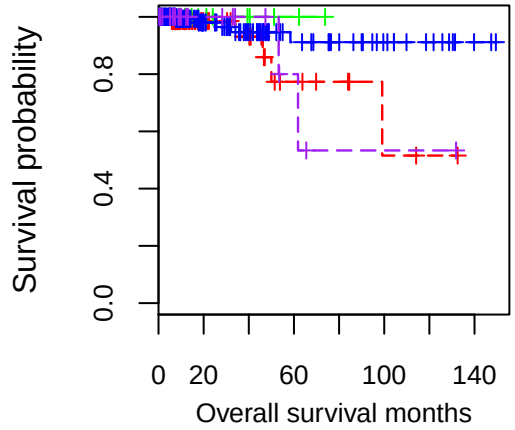
THCA − results of ccpwModel

Mutation−based clustering

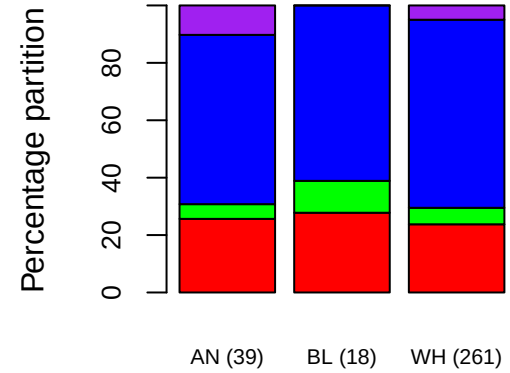


Cluster−survival association

Cox−PH p−value: 0.6407 (C1,4 vs others)

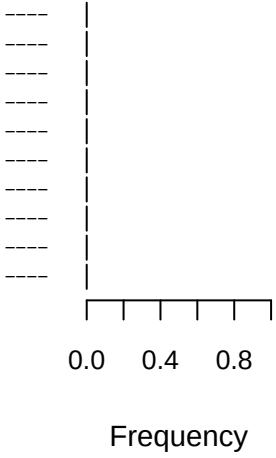


Cluster−race association

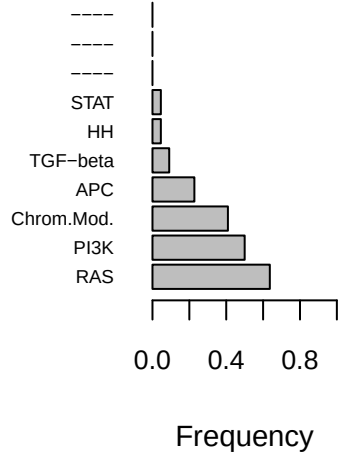


Fisher's test p−value: 0.6342

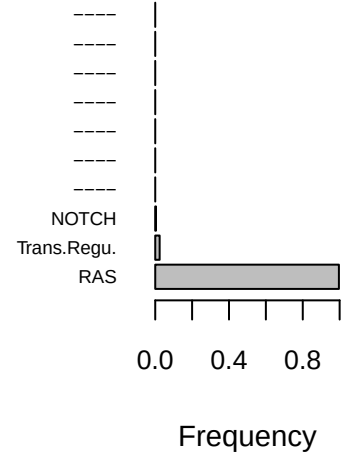
cluster 1



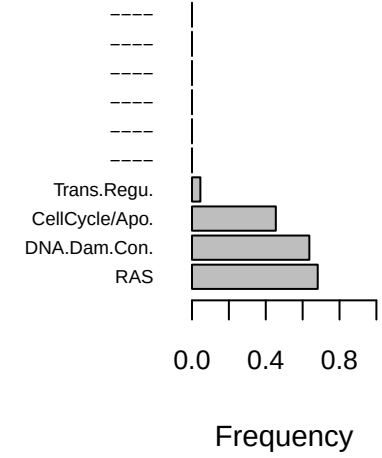
cluster 2



cluster 3

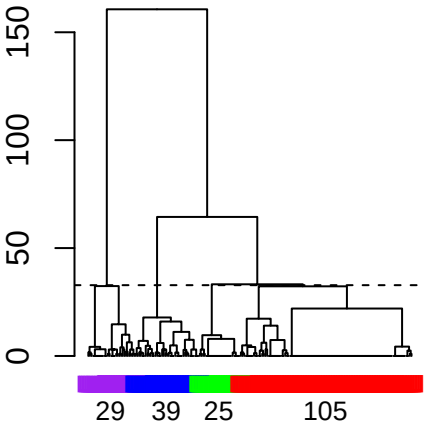


cluster 4



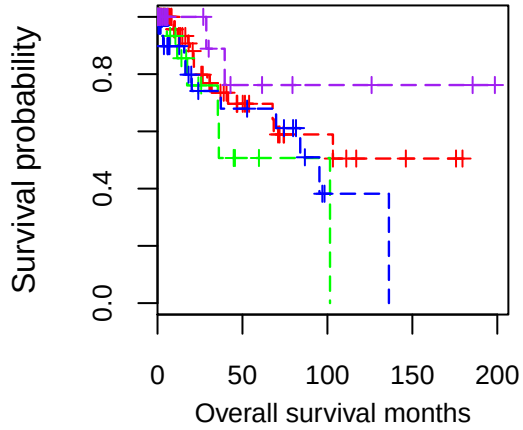
CESC − results of ccpwModel

Mutation−based clustering

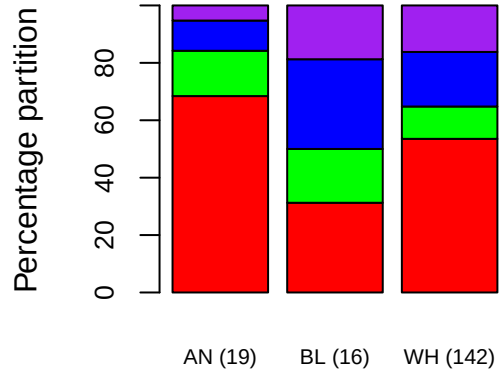


Cluster−survival association

Cox−PH p−value: 0.0554 (C1,4 vs others)

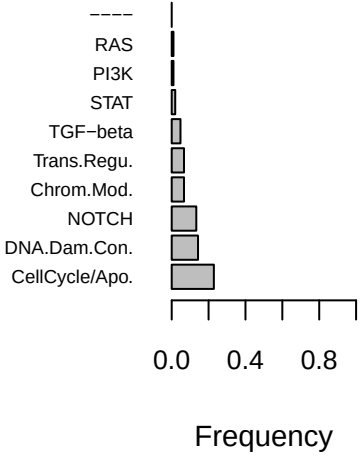


Cluster−race association

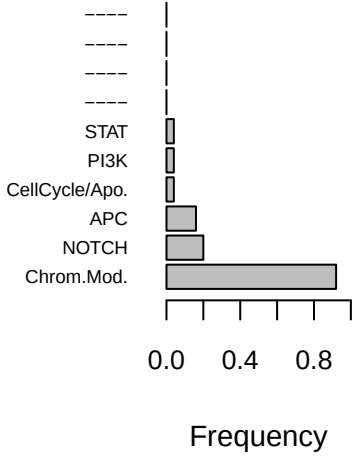


Fisher's test p−value: 0.3053

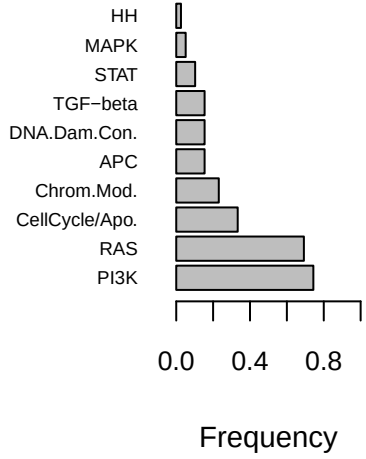
cluster 1



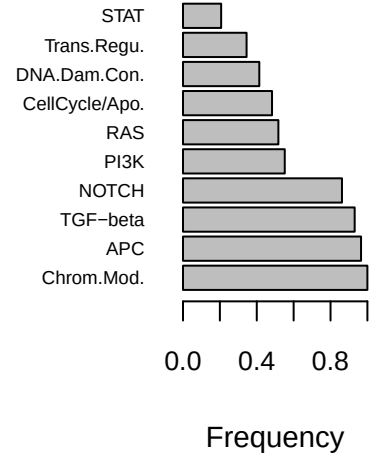
cluster 2



cluster 3

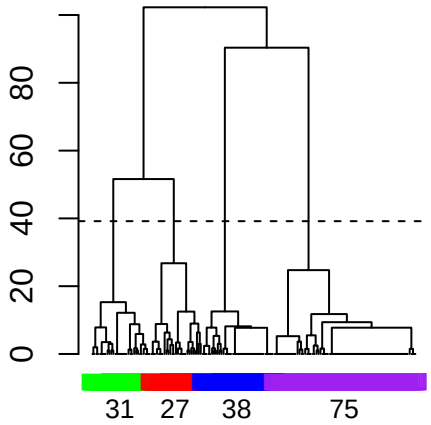


cluster 4



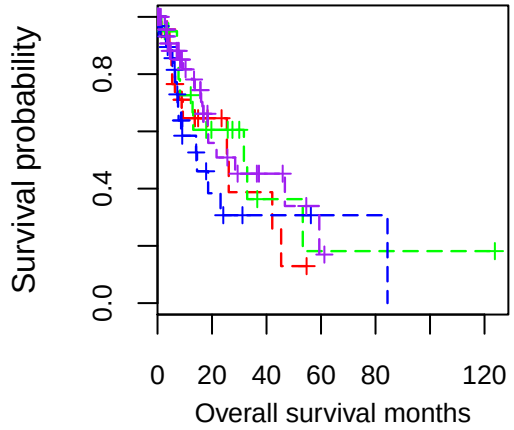
ESCA − results of ccpwModel

Mutation−based clustering

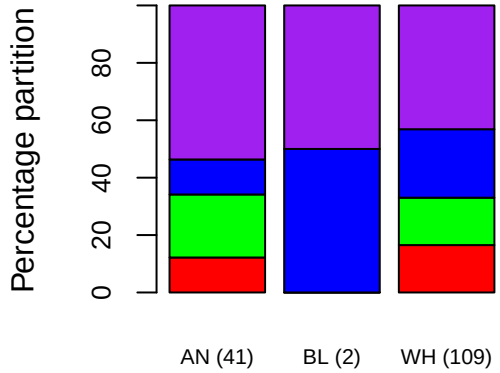


Cluster−survival association

Cox−PH p−value: 0.1222 (C2,4 vs others)

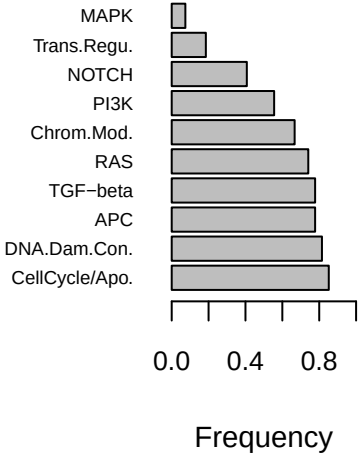


Cluster−race association

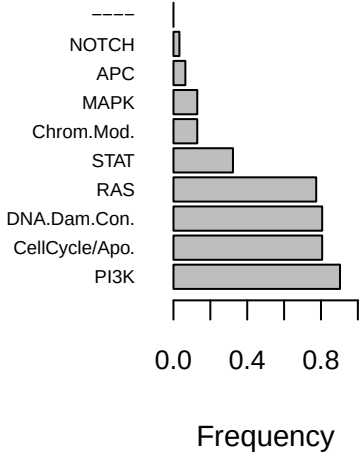


Fisher's test p−value: 0.4833

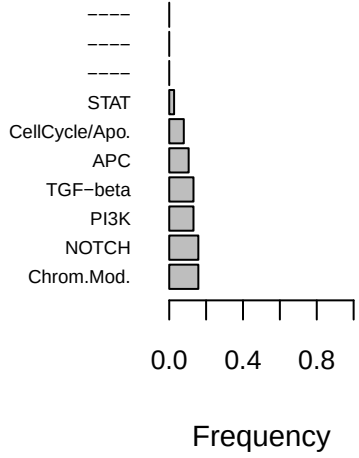
cluster 1



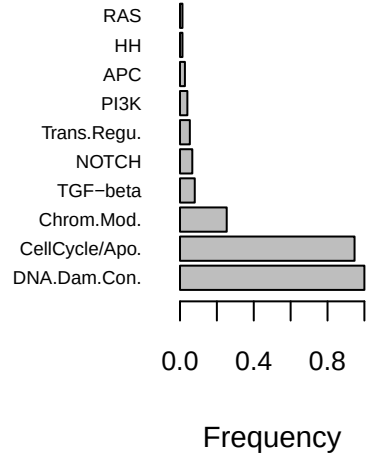
cluster 2



cluster 3

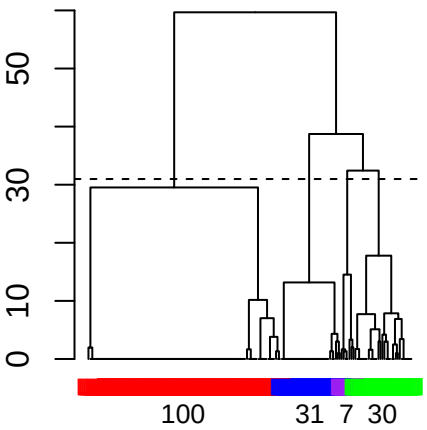


cluster 4



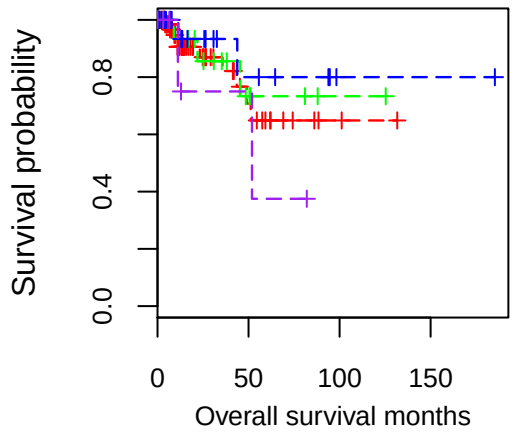
KIRP − results of ccpwModel

Mutation−based clustering

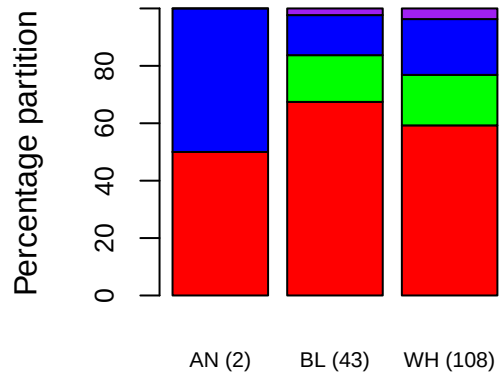


Cluster−survival association

Cox−PH p−value: 0.2643 (C4 vs others)

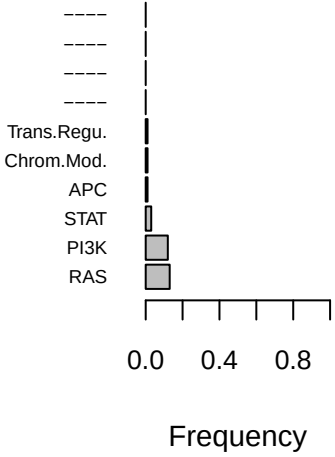


Cluster−race association

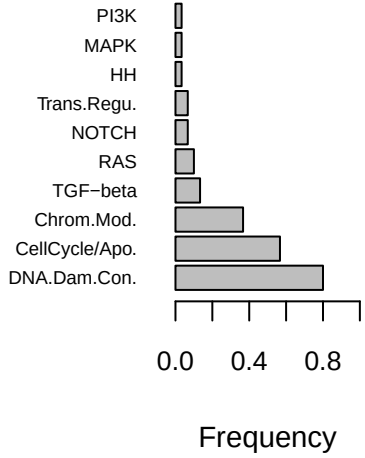


Fisher's test p−value: 0.7756

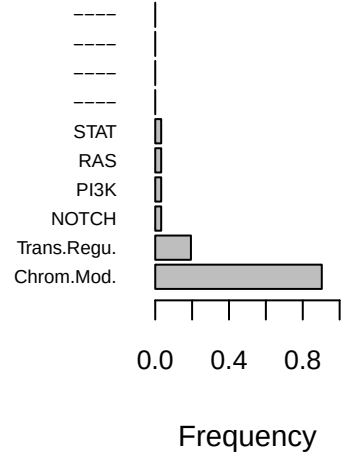
cluster 1



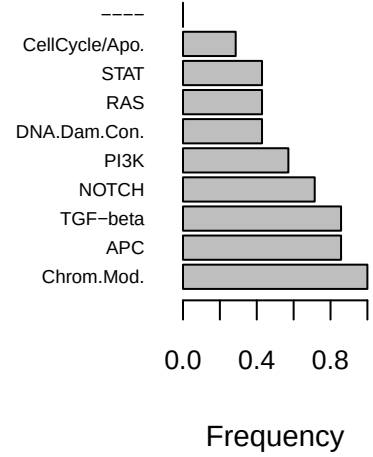
cluster 2



cluster 3

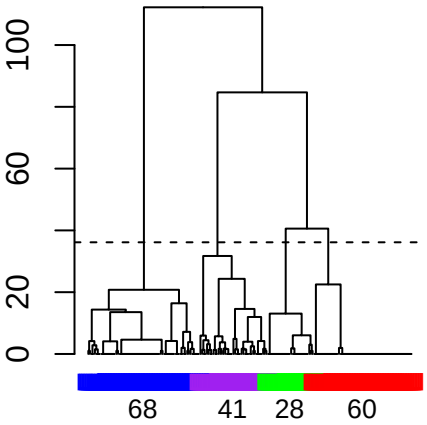


cluster 4



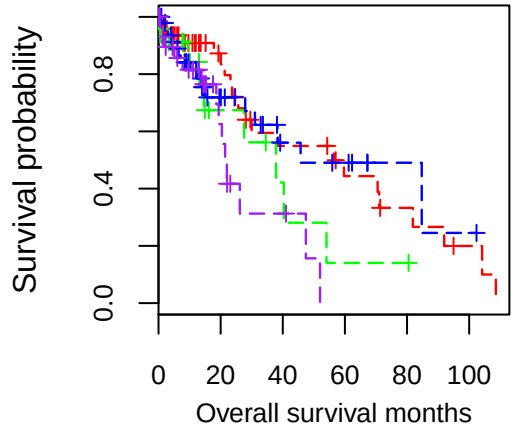
LIHC − results of ccpwModel

Mutation−based clustering

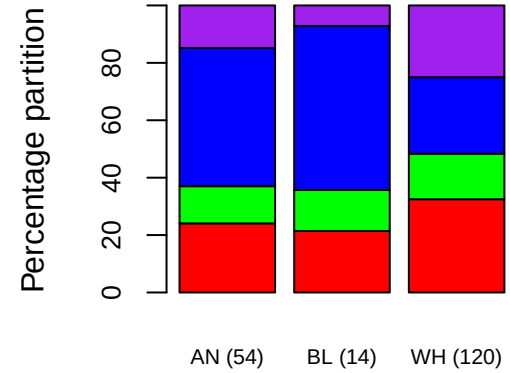


Cluster−survival association

Cox−PH p−value: 0.0107 (C1,3 vs others)

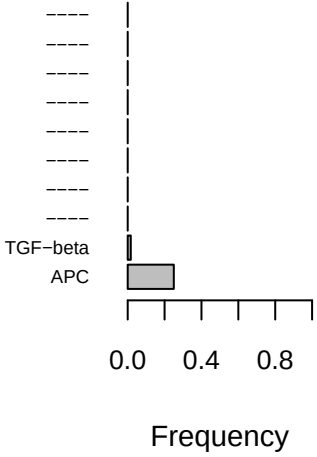


Cluster−race association

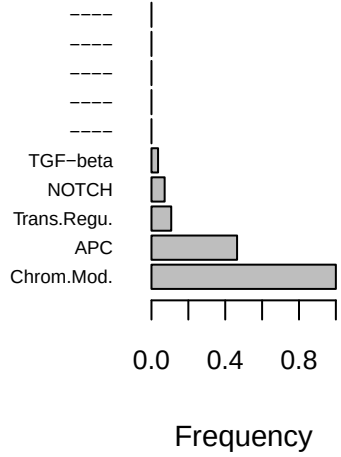


Fisher's test p−value: 0.0775

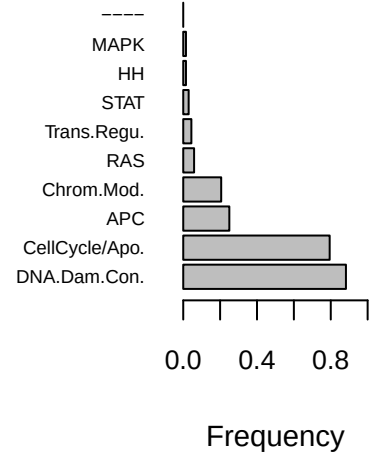
cluster 1



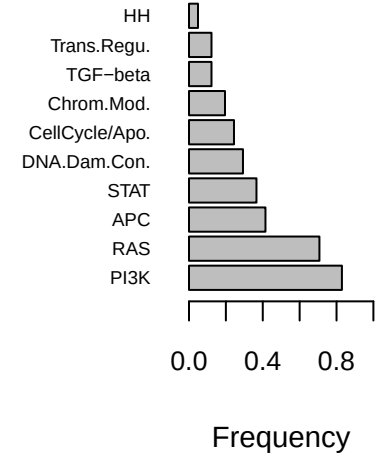
cluster 2



cluster 3

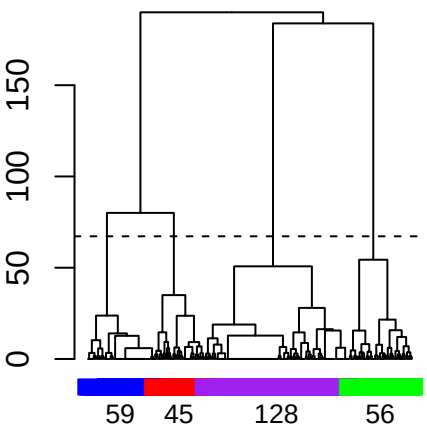


cluster 4



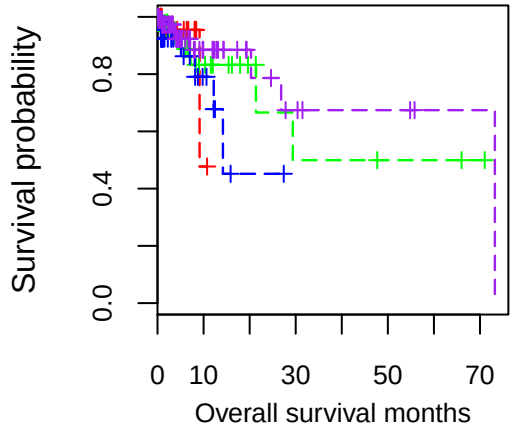
STAD − results of ccpwModel

Mutation−based clustering

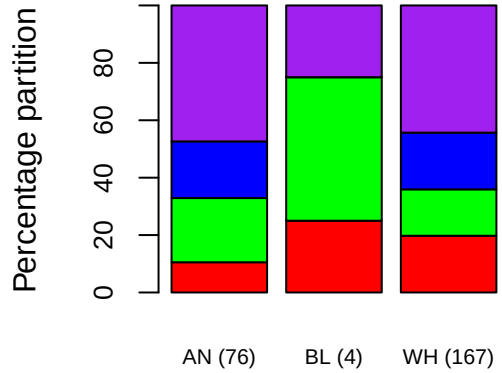


Cluster−survival association

Cox−PH p−value: 0.072 (C3 vs others)

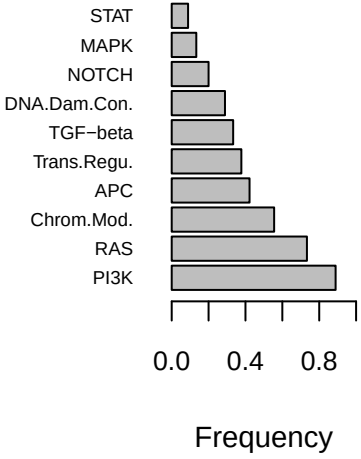


Cluster−race association

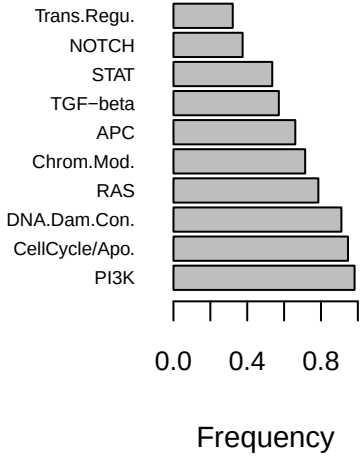


Fisher's test p−value: 0.2529

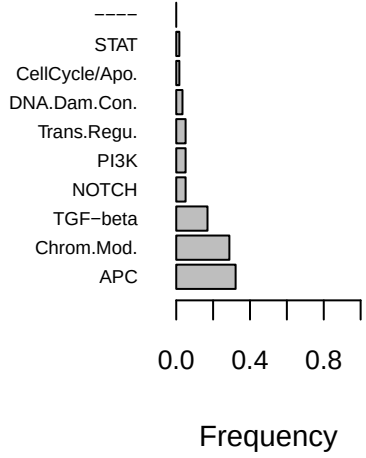
cluster 1



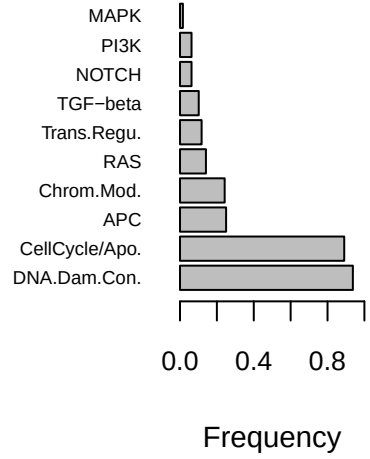
cluster 2



cluster 3

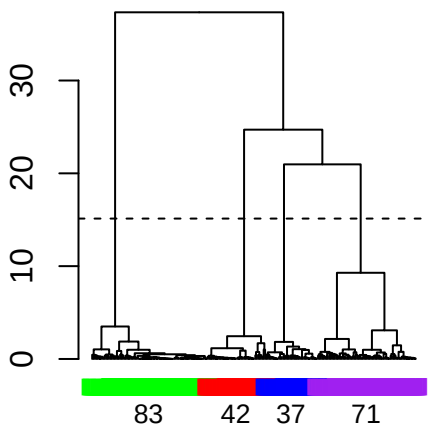


cluster 4



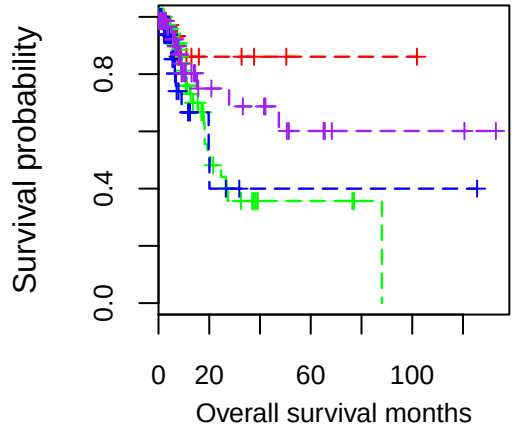
BLCA − results of nbsModel

Mutation−based clustering

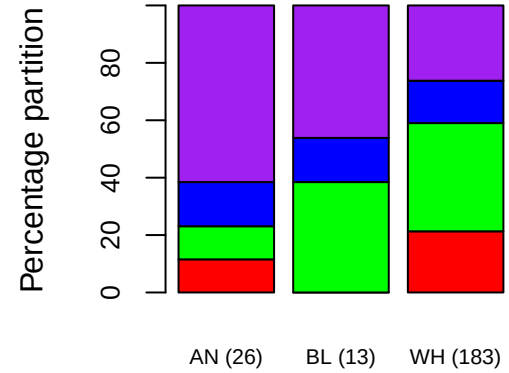


Cluster−survival association

Cox−PH p−value: 0.0276 (C2,3 vs others)

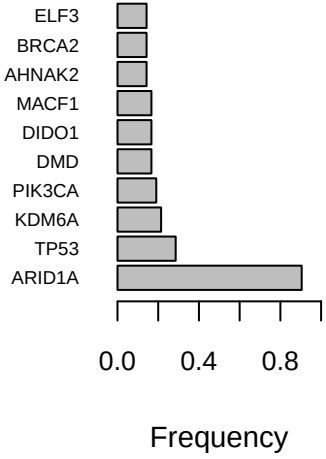


Cluster−race association

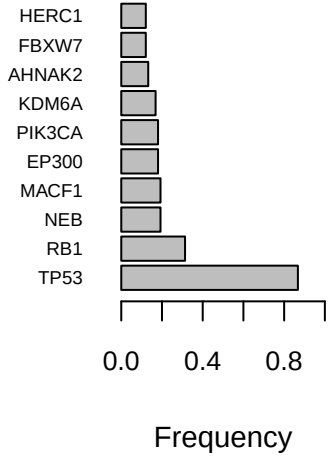


Fisher's test p−value: 0.0025

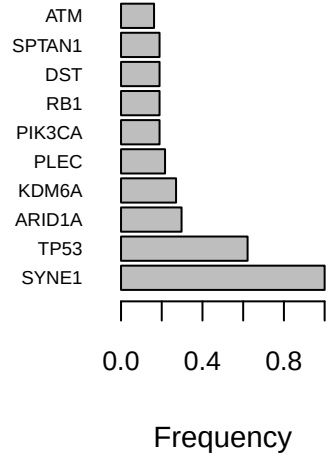
cluster 1



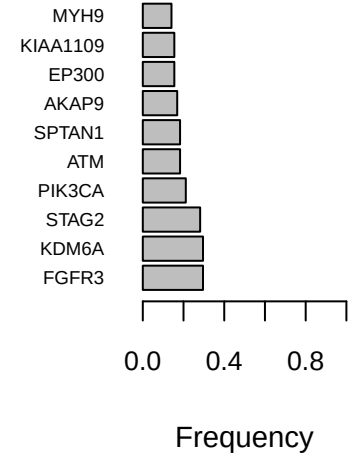
cluster 2



cluster 3

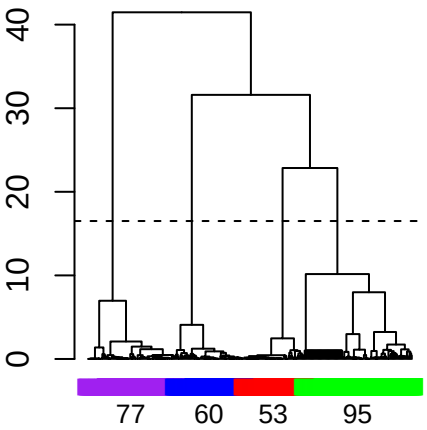


cluster 4



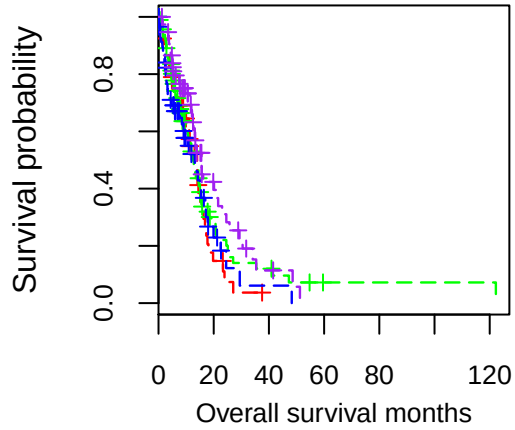
GBM − results of nbsModel

Mutation−based clustering

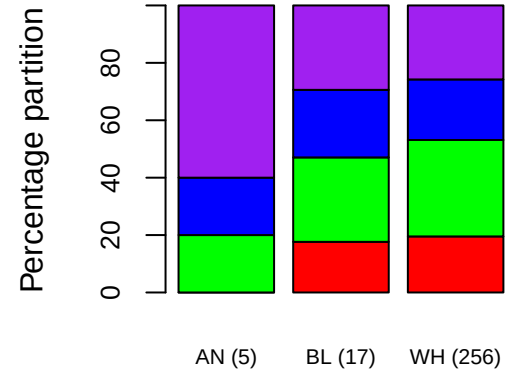


Cluster−survival association

Cox−PH p−value: 0.0656 (C1,3 vs others)

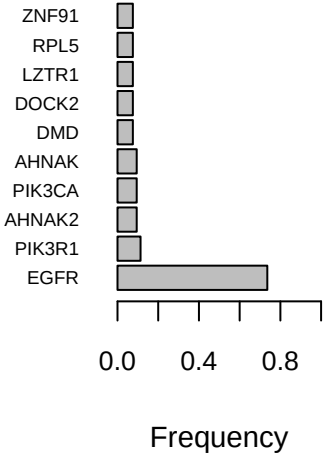


Cluster−race association

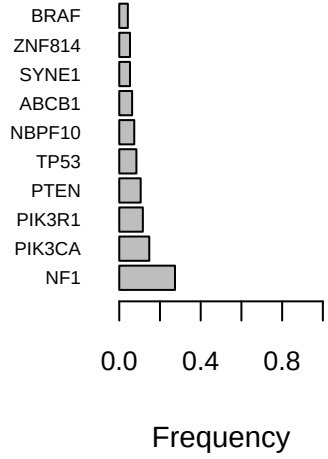


Fisher's test p−value: 0.8401

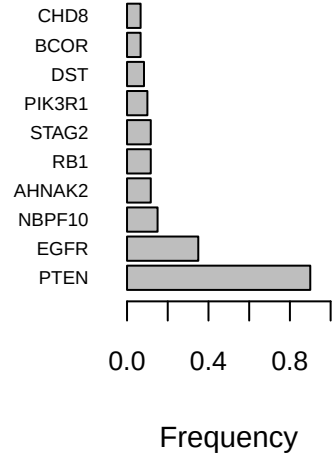
cluster 1



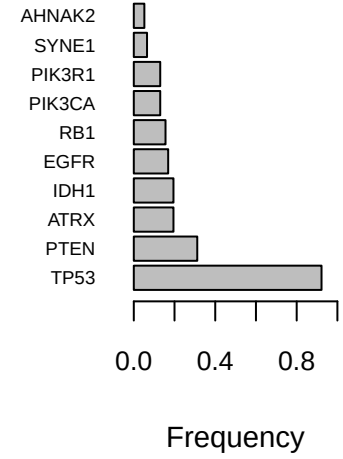
cluster 2



cluster 3

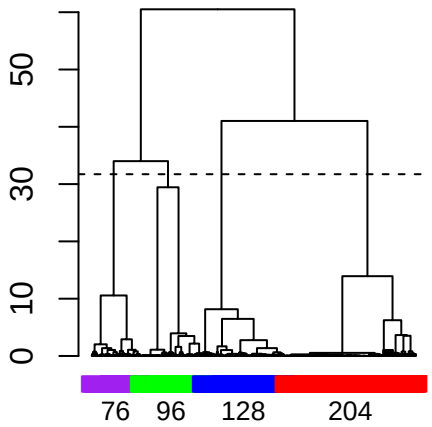


cluster 4



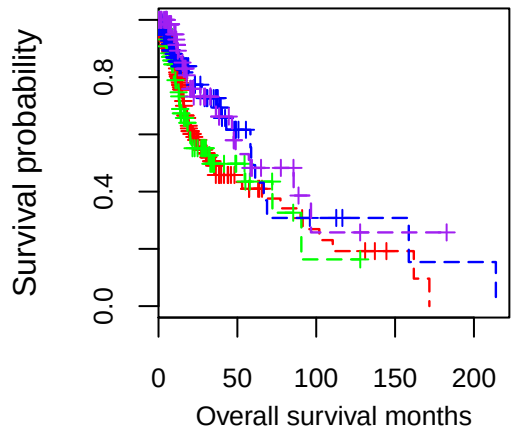
HNSC − results of nbsModel

Mutation−based clustering

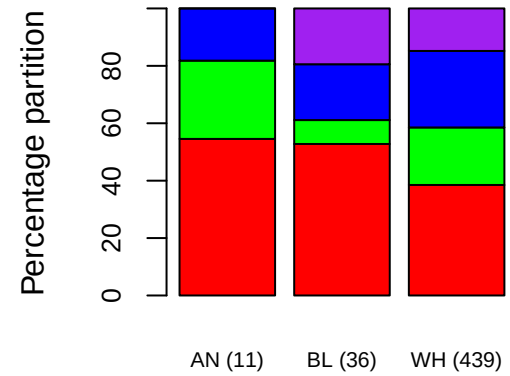


Cluster−survival association

Cox−PH p−value: 0.0028 (C1,2 vs others)

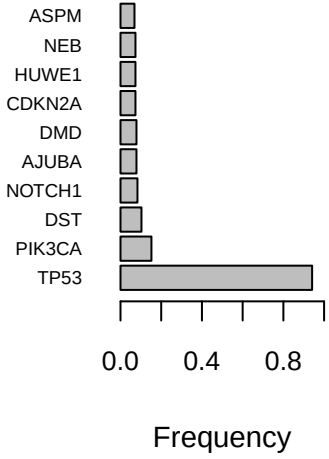


Cluster−race association

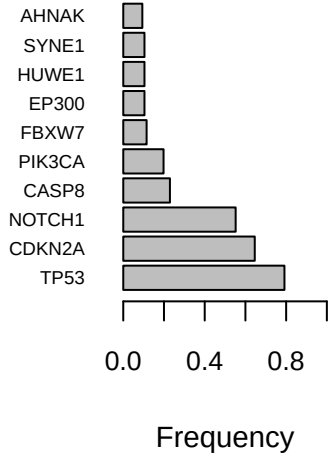


Fisher's test p−value: 0.2089

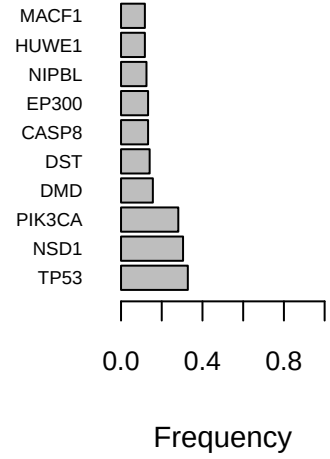
cluster 1



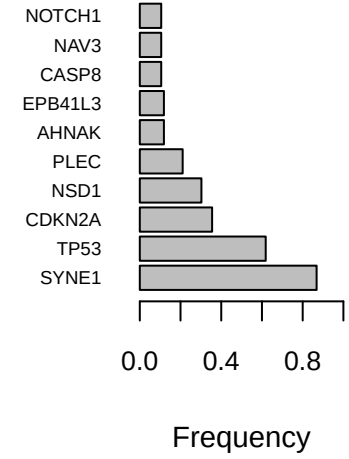
cluster 2



cluster 3

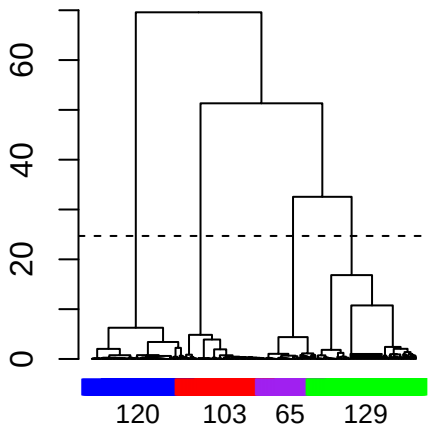


cluster 4



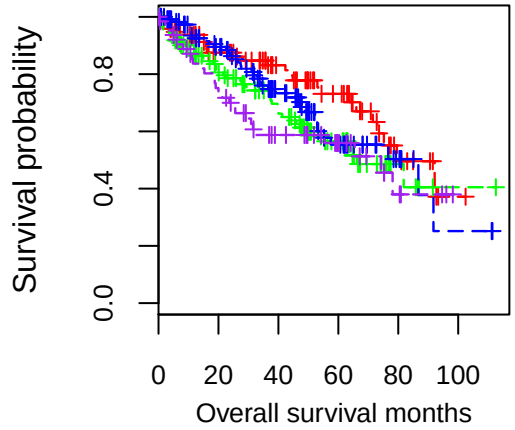
KIRC − results of nbsModel

Mutation−based clustering

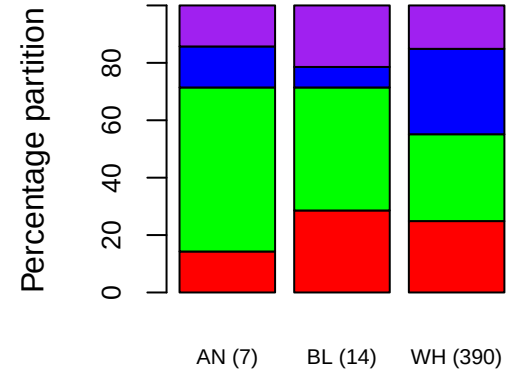


Cluster−survival association

Cox−PH p−value: 0.0176 (C1,3 vs others)

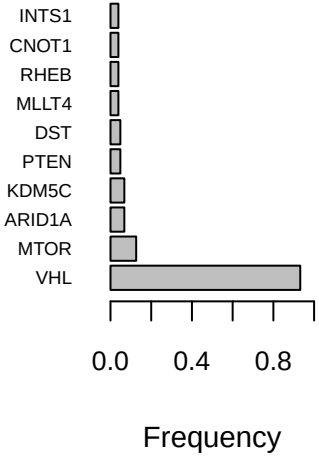


Cluster−race association

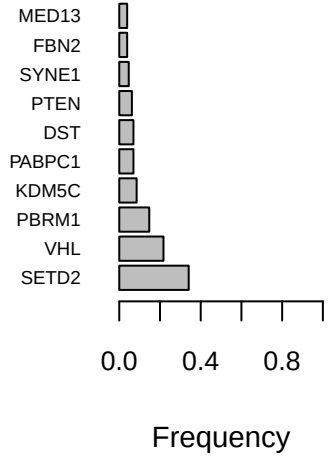


Fisher's test p−value: 0.3713

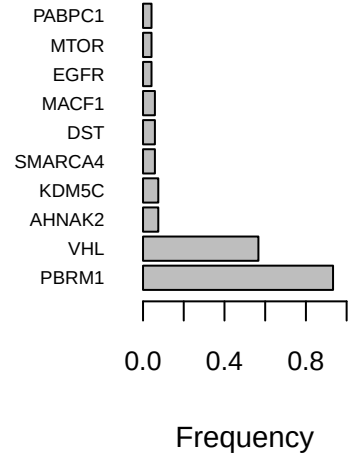
cluster 1



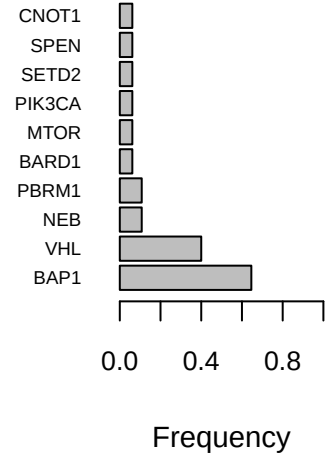
cluster 2



cluster 3

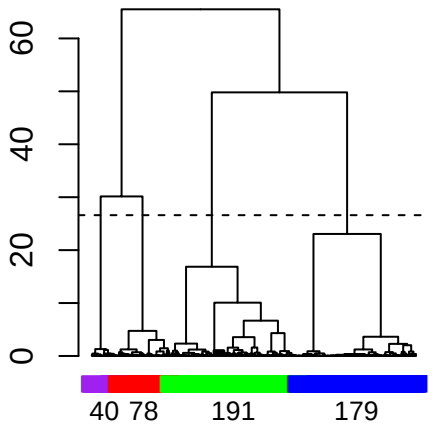


cluster 4



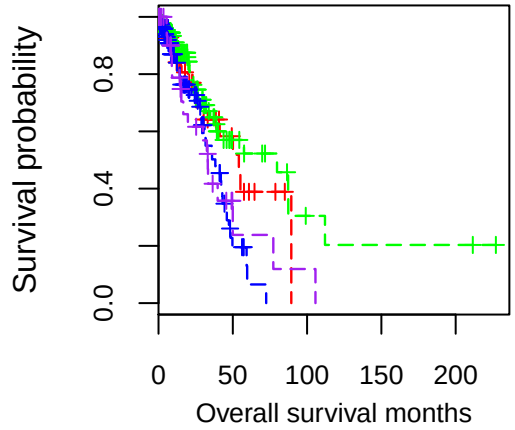
LUAD − results of nbsModel

Mutation−based clustering

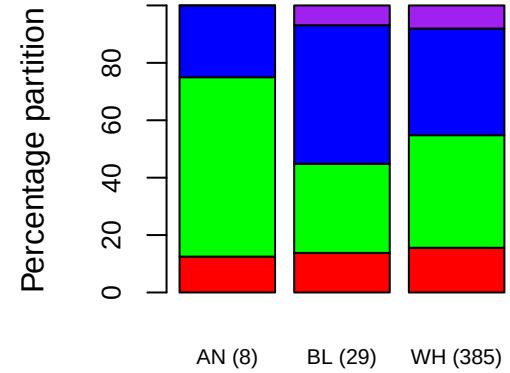


Cluster−survival association

Cox−PH p−value: 0.0005 (C1,2 vs others)

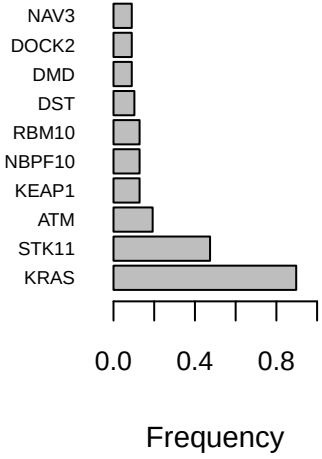


Cluster−race association

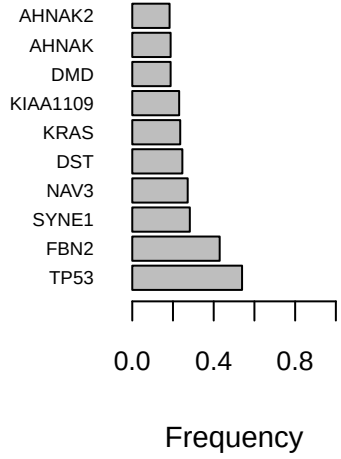


Fisher's test p−value: 0.8311

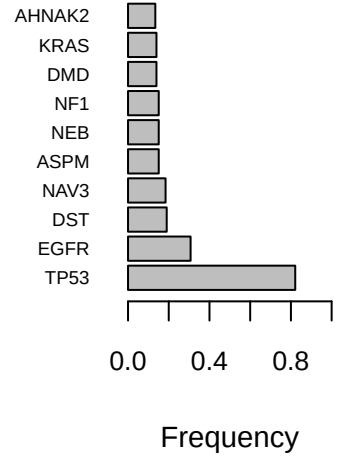
cluster 1



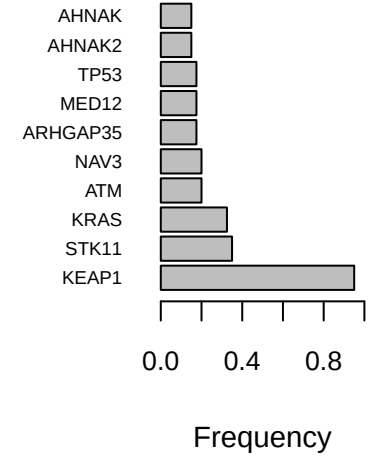
cluster 2



cluster 3

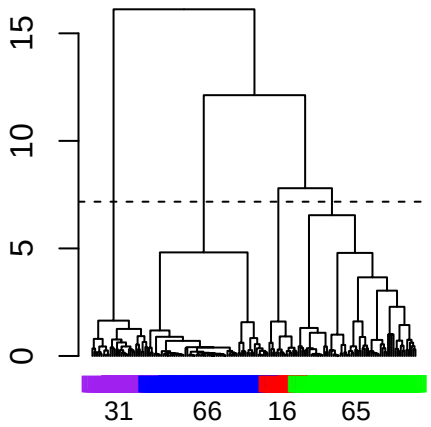


cluster 4



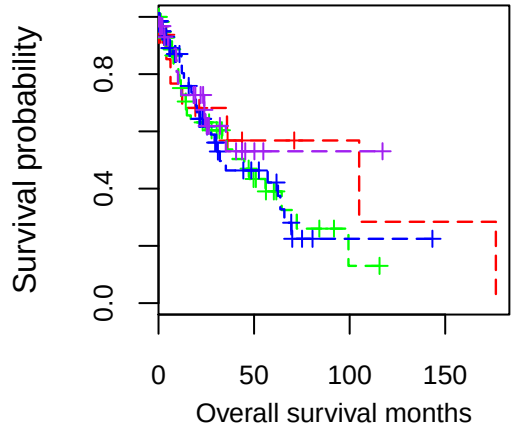
LUSC − results of nbsModel

Mutation−based clustering

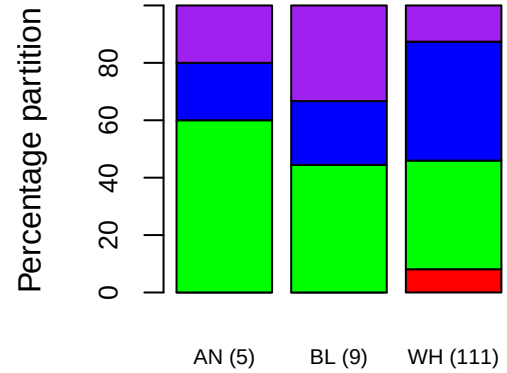


Cluster−survival association

Cox−PH p−value: 0.4998 (C2,3 vs others)

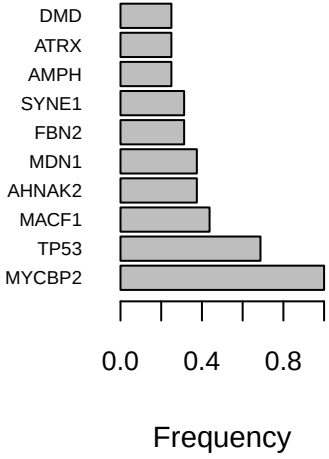


Cluster−race association

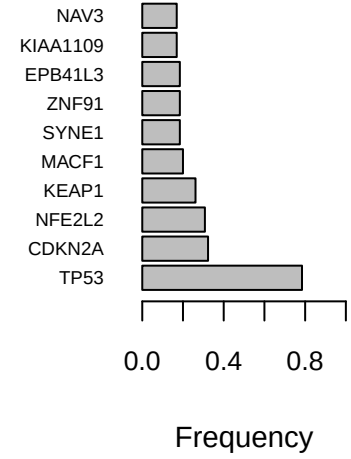


Fisher's test p−value: 0.4583

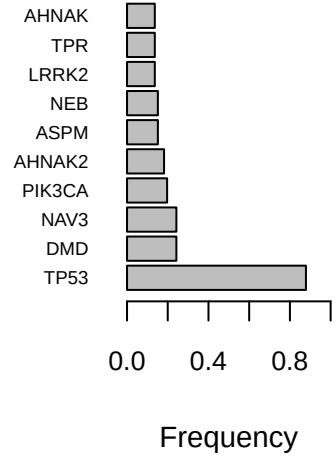
cluster 1



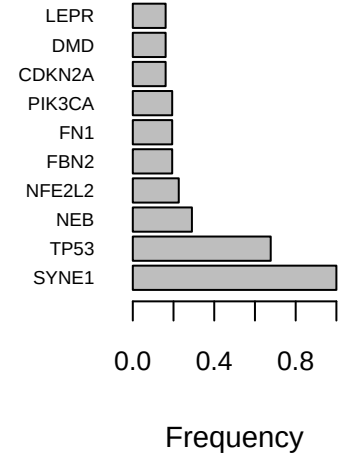
cluster 2



cluster 3

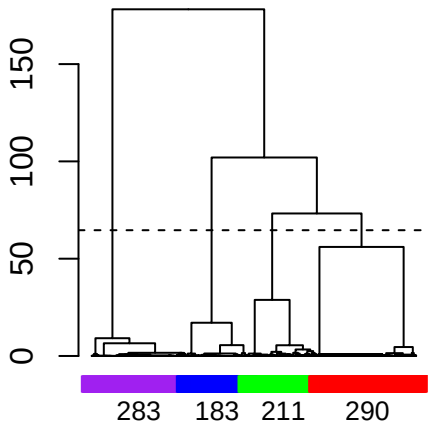


cluster 4



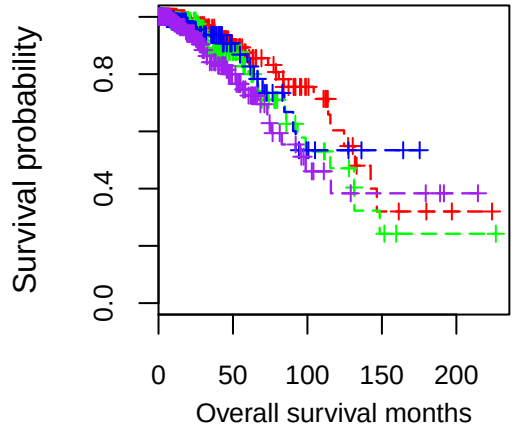
BRCA − results of nbsModel

Mutation−based clustering

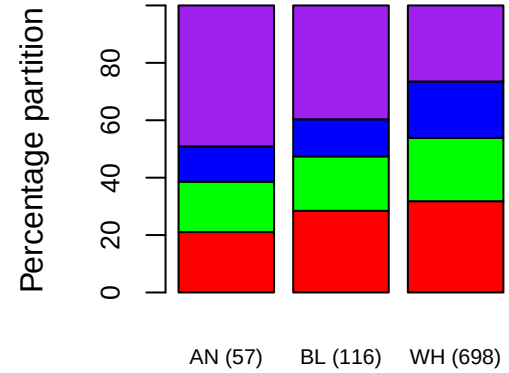


Cluster−survival association

Cox−PH p−value: 0.0025 (C1,3 vs others)

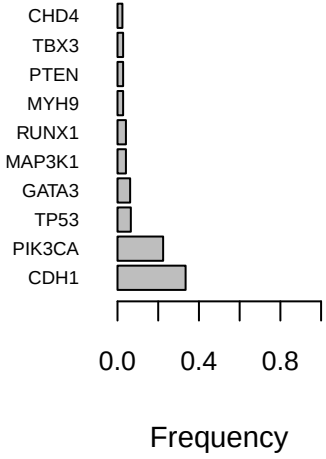


Cluster−race association

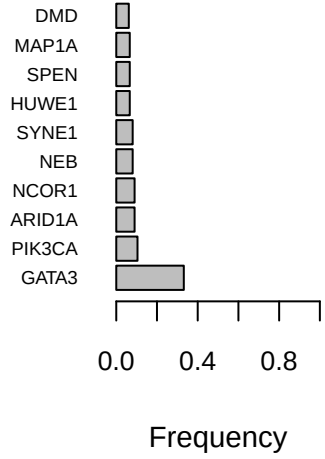


Fisher's test p−value: 0.003

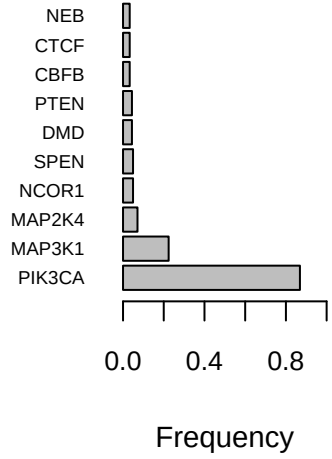
cluster 1



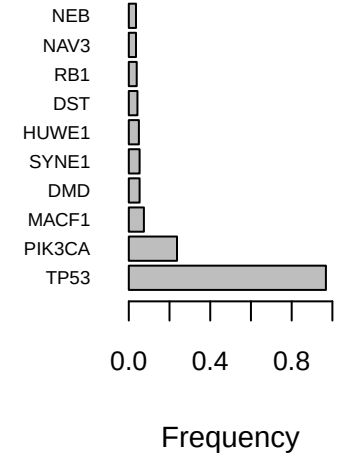
cluster 2



cluster 3

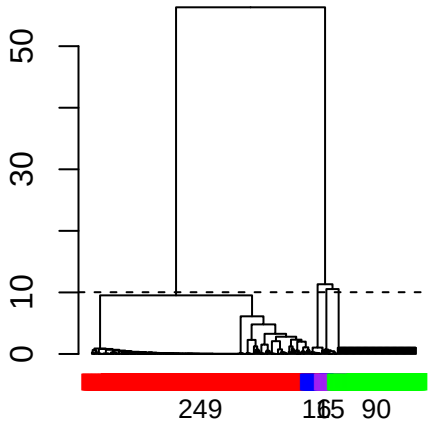


cluster 4



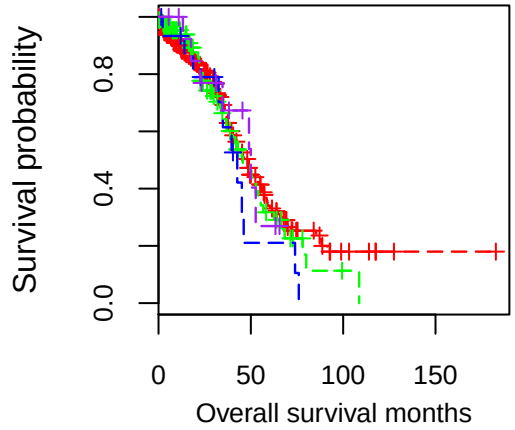
OV − results of nbsModel

Mutation−based clustering

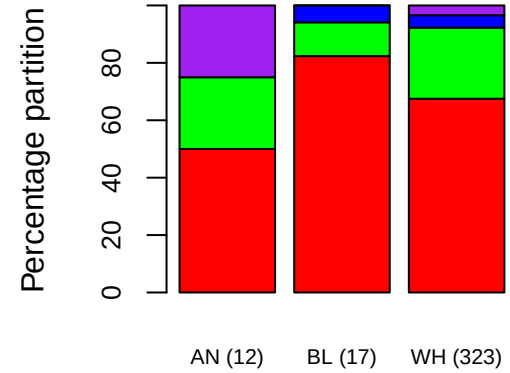


Cluster−survival association

Cox−PH p−value: 0.1832 (C1,4 vs others)

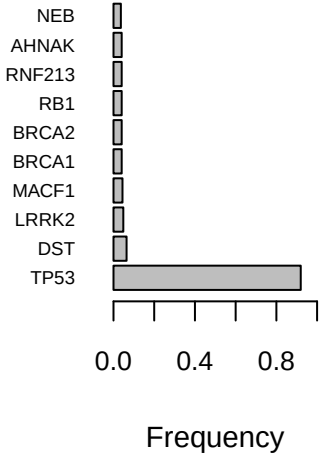


Cluster−race association

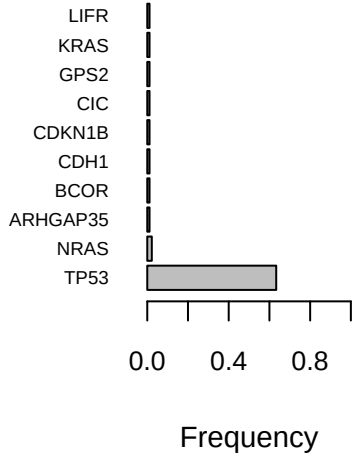


Fisher's test p−value: 0.0695

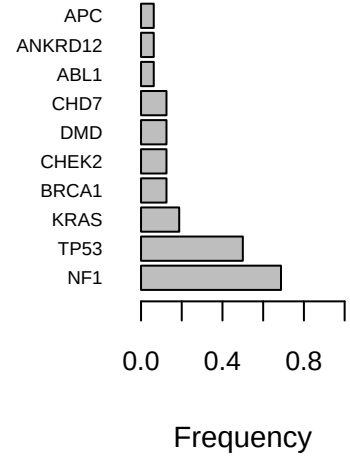
cluster 1



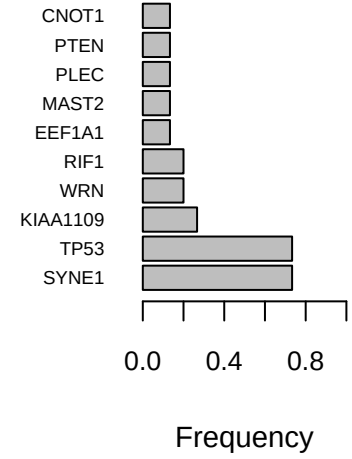
cluster 2



cluster 3

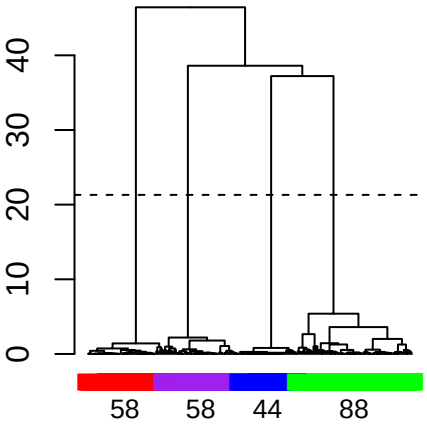


cluster 4



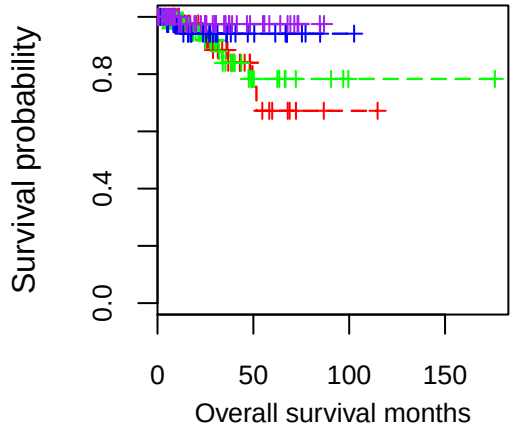
UCEC − results of nbsModel

Mutation−based clustering

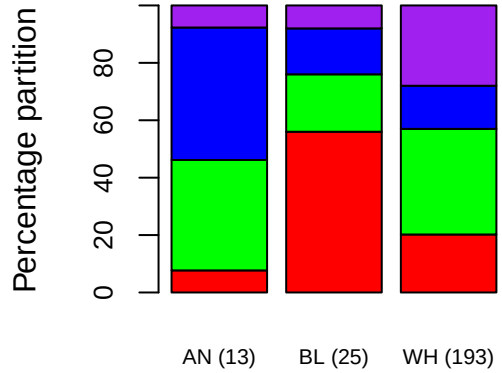


Cluster−survival association

Cox−PH p−value: 0.0406 (C3,4 vs others)

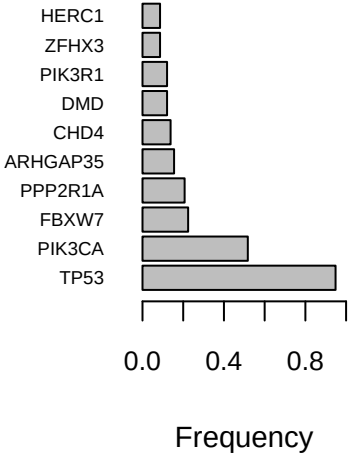


Cluster−race association

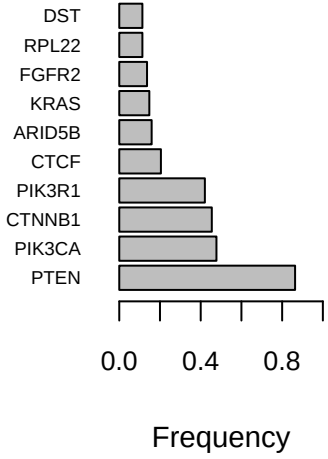


Fisher's test p−value: 0.0005

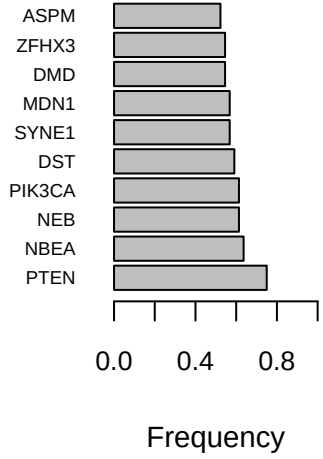
cluster 1



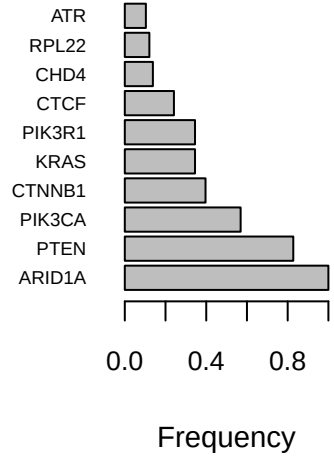
cluster 2



cluster 3

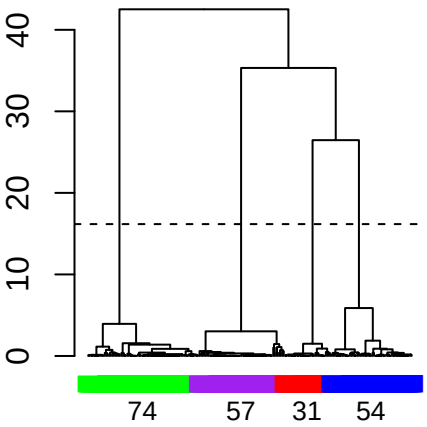


cluster 4



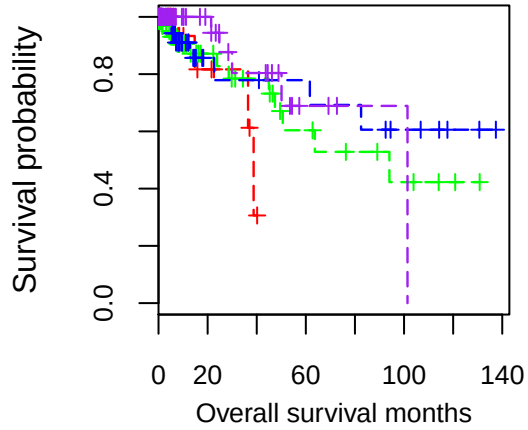
COAD − results of nbsModel

Mutation−based clustering

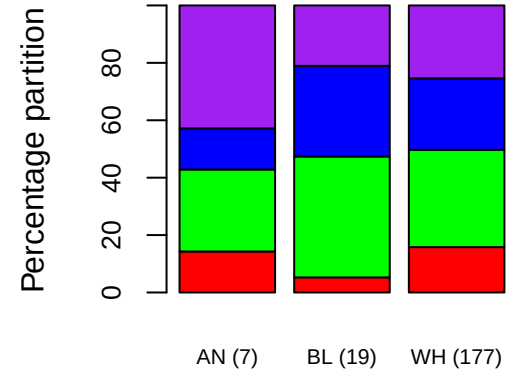


Cluster−survival association

Cox−PH p−value: 0.1833 (C1,2 vs others)

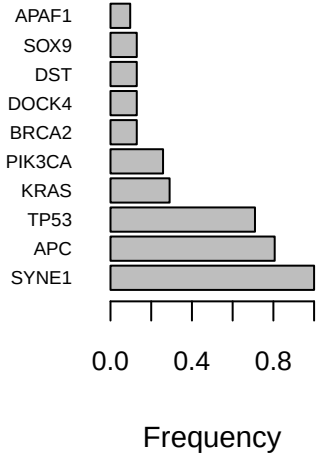


Cluster−race association

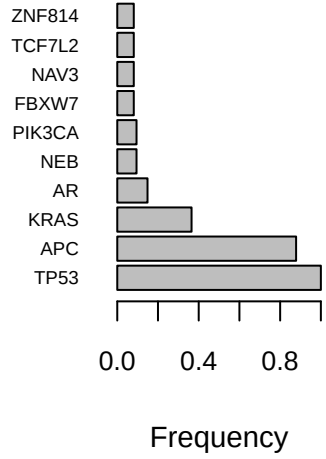


Fisher's test p−value: 0.7996

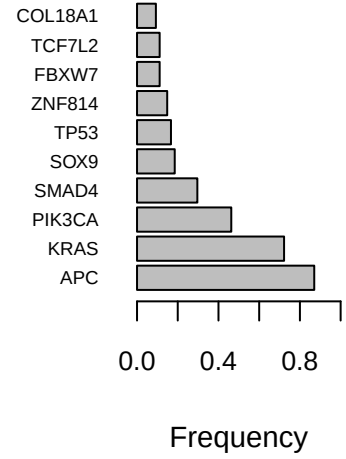
cluster 1



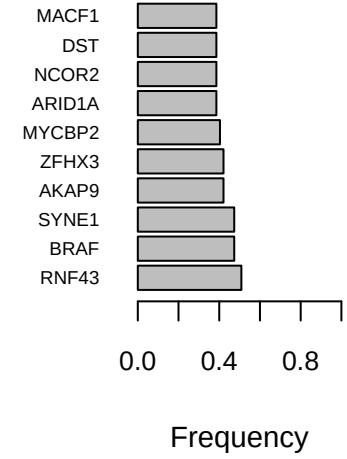
cluster 2



cluster 3

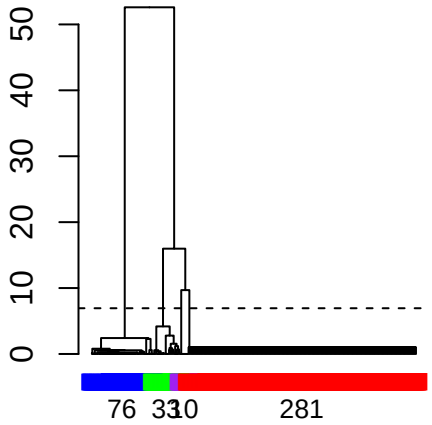


cluster 4



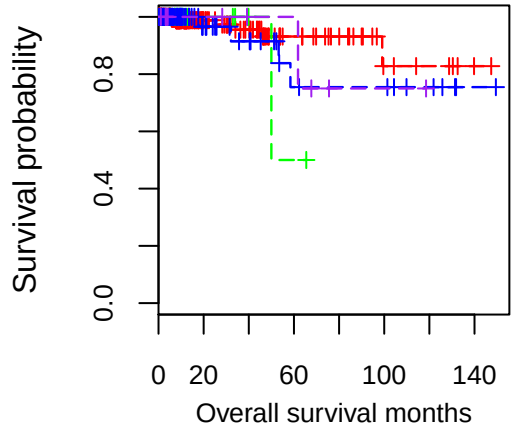
THCA − results of nbsModel

Mutation−based clustering

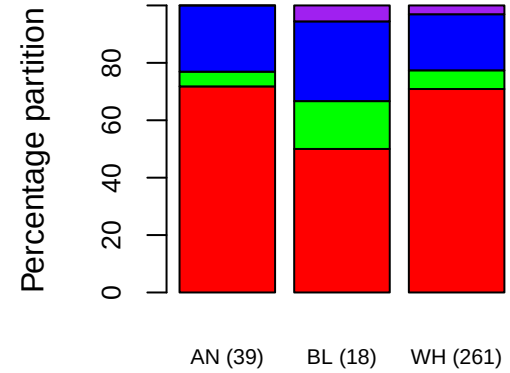


Cluster−survival association

Cox−PH p−value: 0.2934 (C2,3 vs others)

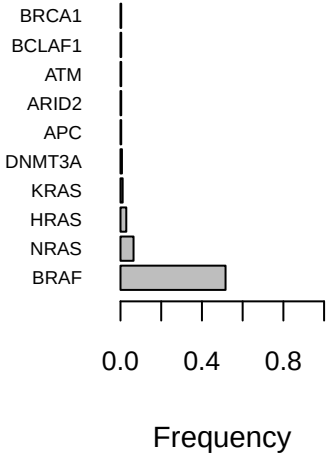


Cluster−race association

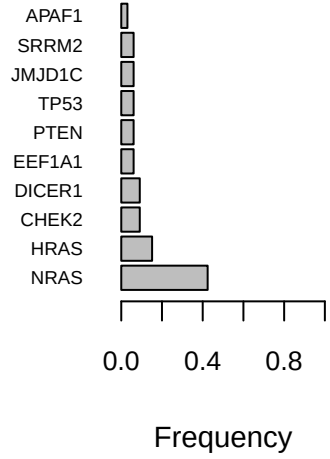


Fisher's test p−value: 0.3113

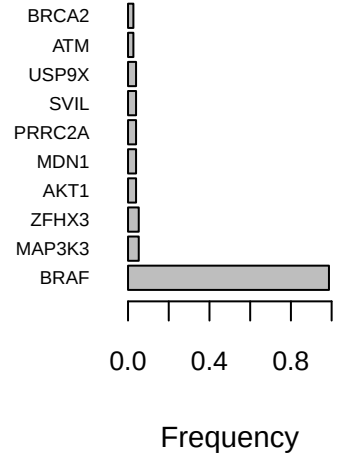
cluster 1



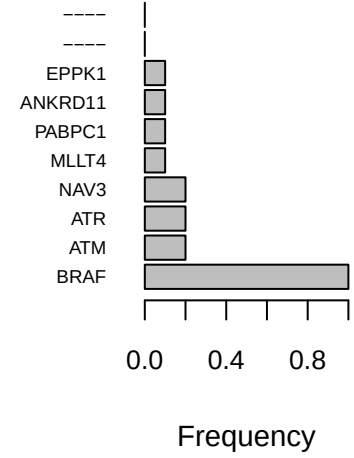
cluster 2



cluster 3

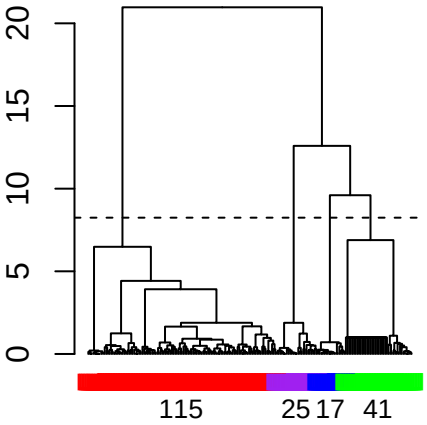


cluster 4



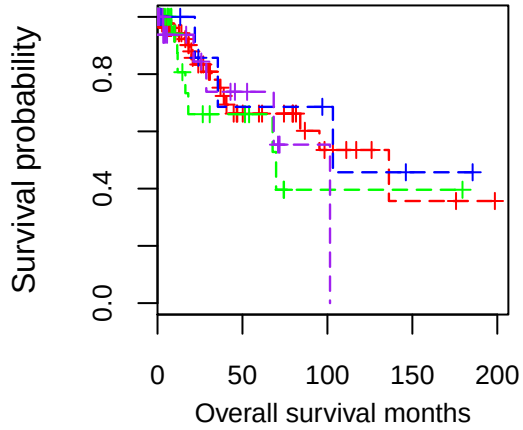
CESC − results of nbsModel

Mutation−based clustering

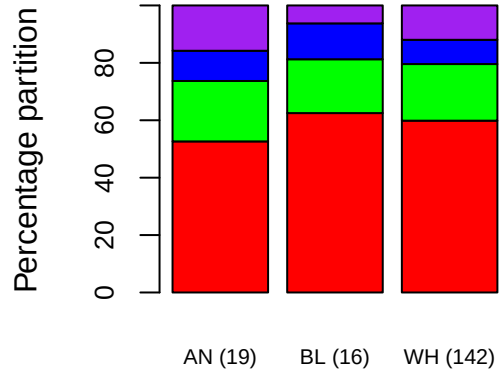


Cluster−survival association

Cox−PH p−value: 0.233 (C2 vs others)

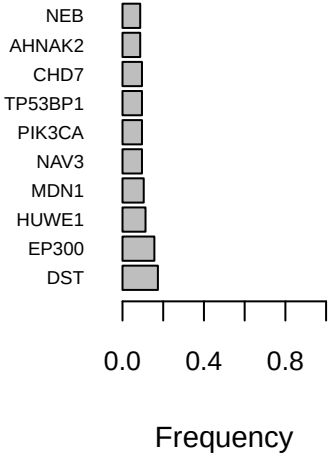


Cluster−race association

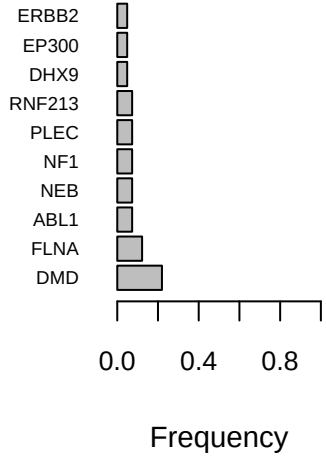


Fisher's test p−value: 0.9585

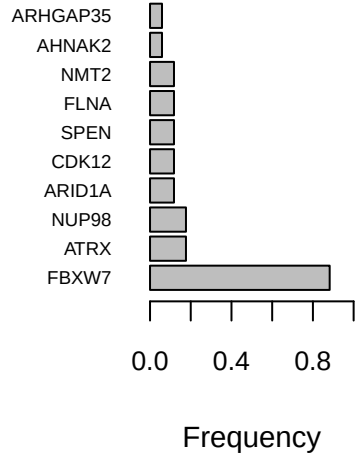
cluster 1



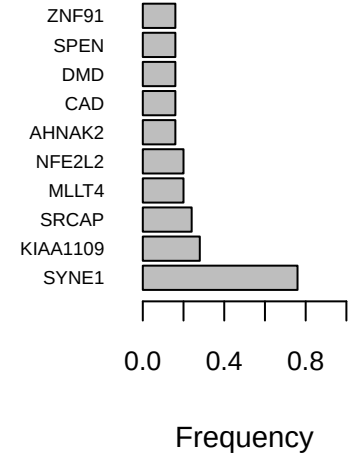
cluster 2



cluster 3

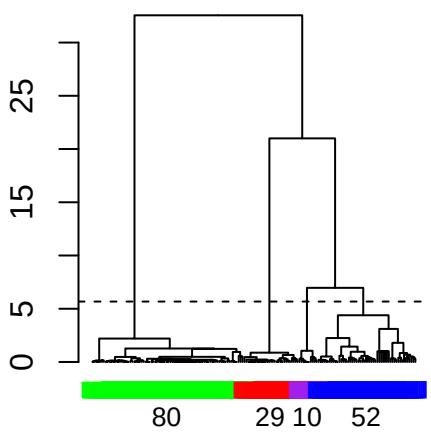


cluster 4



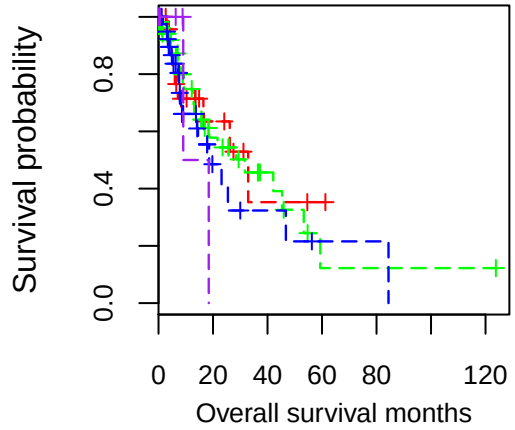
ESCA − results of nbsModel

Mutation−based clustering

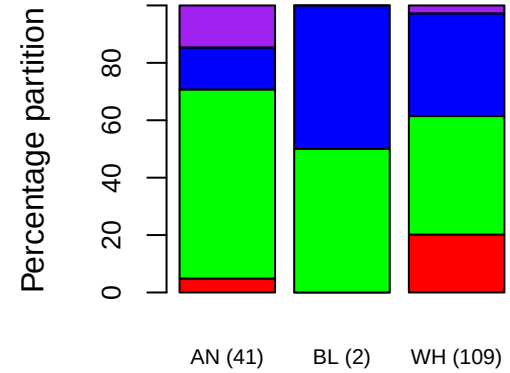


Cluster−survival association

Cox−PH p−value: 0.2622 (C1,2 vs others)

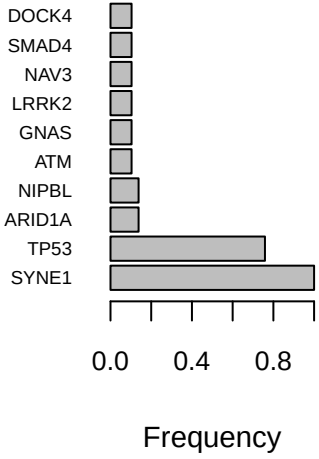


Cluster−race association

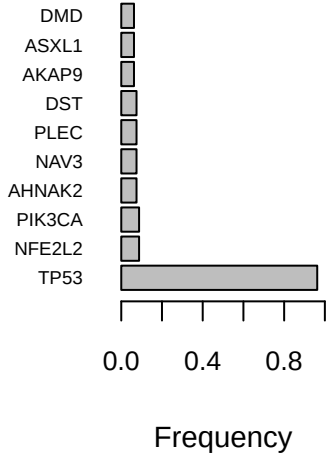


Fisher's test p−value: 0.0015

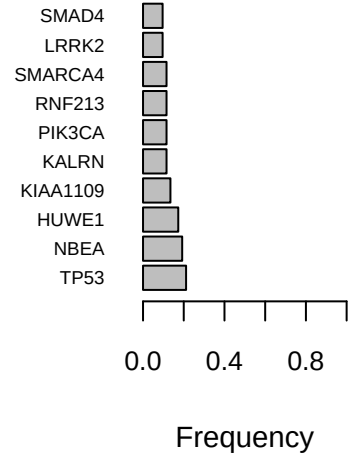
cluster 1



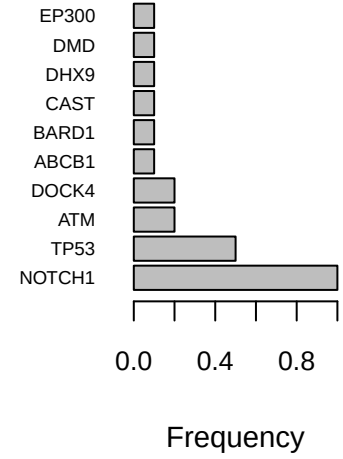
cluster 2



cluster 3

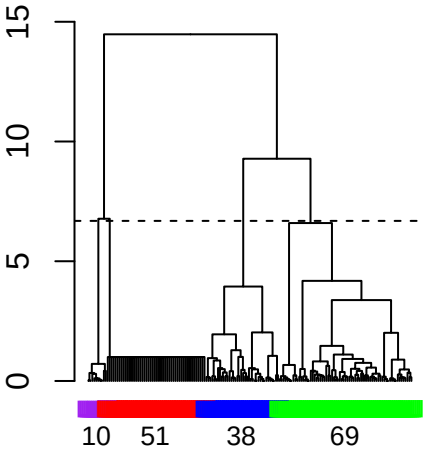


cluster 4



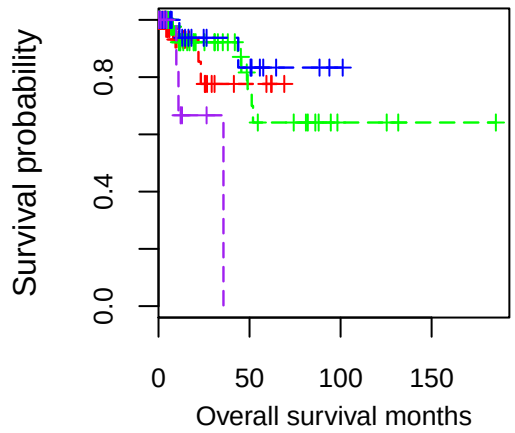
KIRP − results of nbsModel

Mutation−based clustering

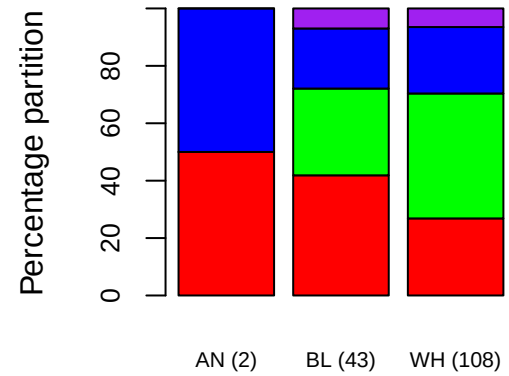


Cluster−survival association

Cox−PH p−value: 0.0093 (C4 vs others)

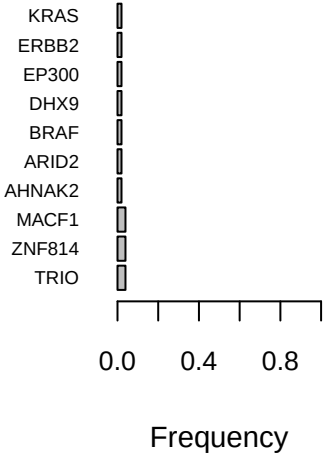


Cluster−race association

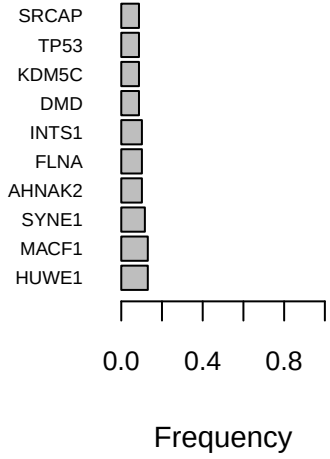


Fisher's test p−value: 0.3623

cluster 1

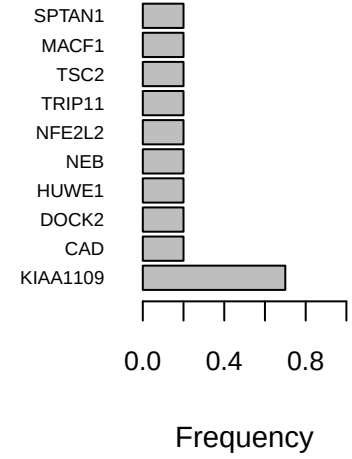


cluster 2



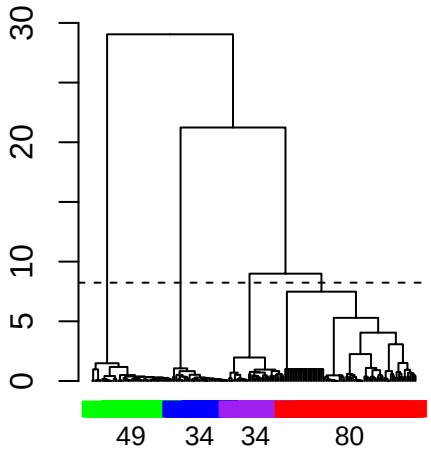
cluster 3

cluster 4

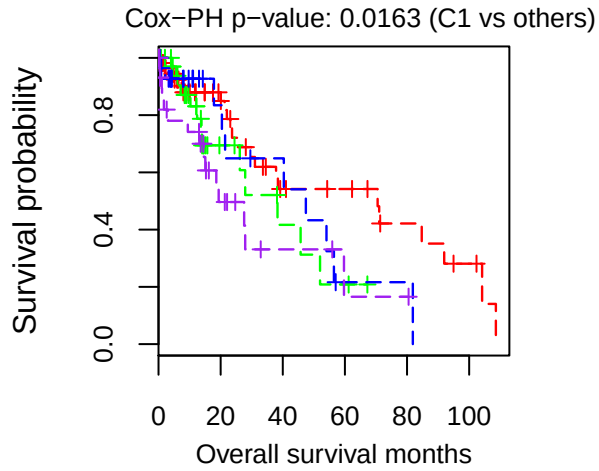


LIHC − results of nbsModel

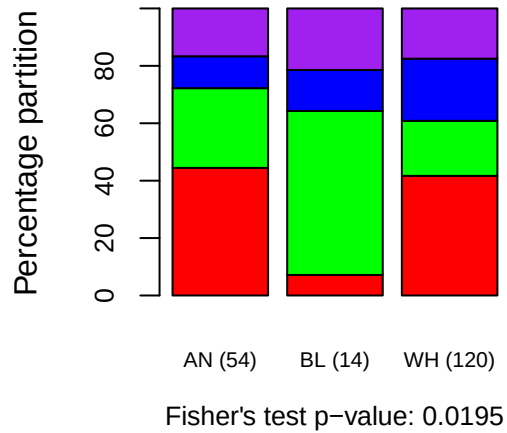
Mutation−based clustering



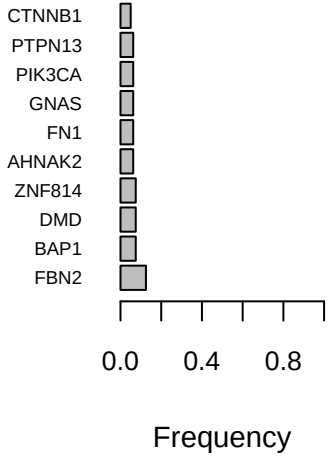
Cluster−survival association



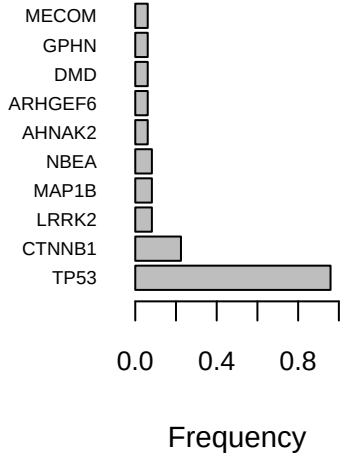
Cluster−race association



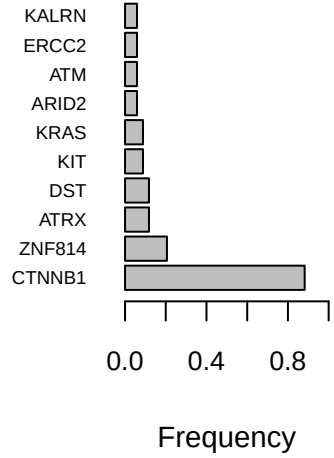
cluster 1



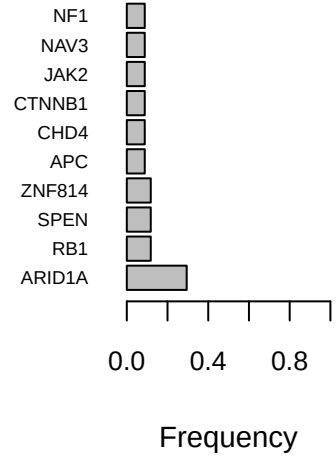
cluster 2



cluster 3

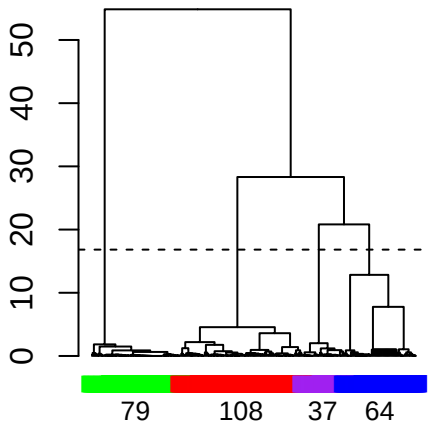


cluster 4



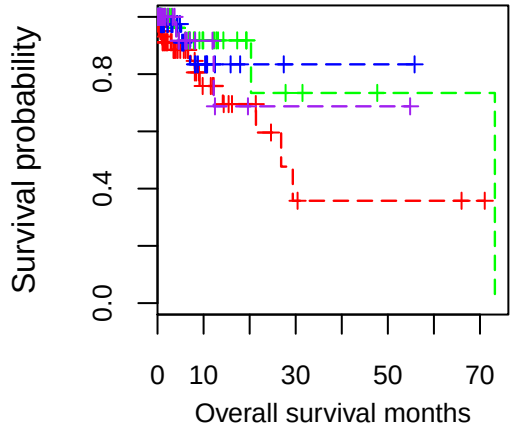
STAD − results of nbsModel

Mutation−based clustering

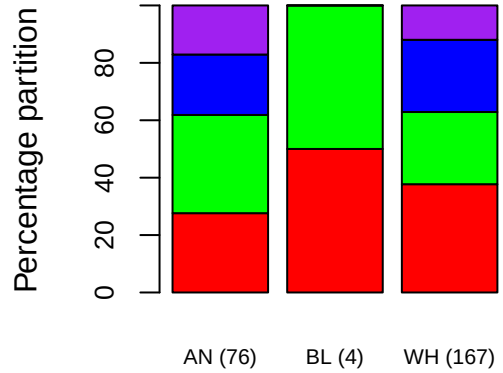


Cluster−survival association

Cox−PH p−value: 0.0273 (C1 vs others)

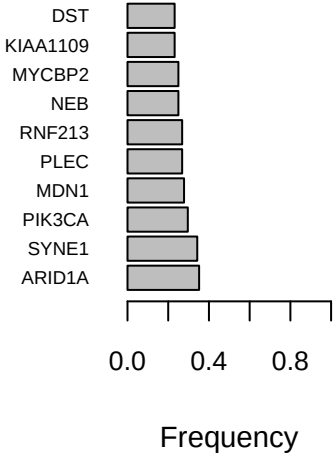


Cluster−race association

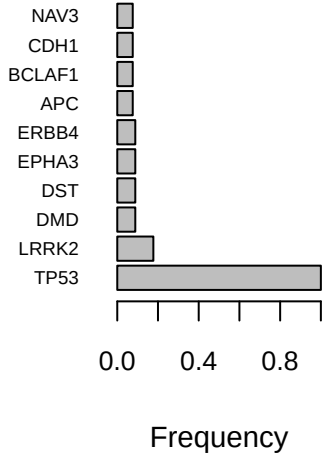


Fisher's test p−value: 0.3138

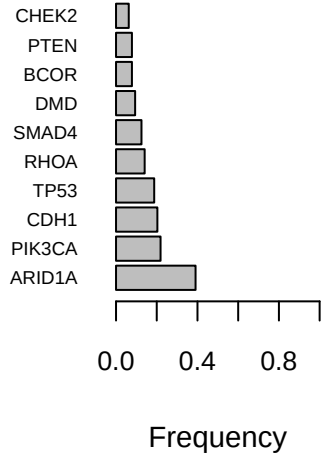
cluster 1



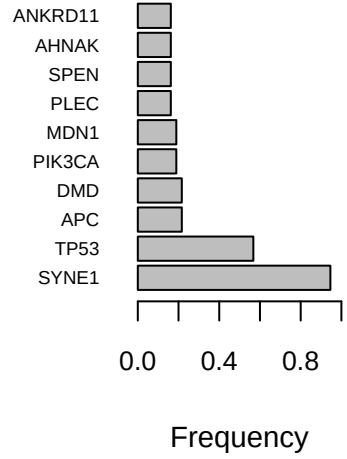
cluster 2

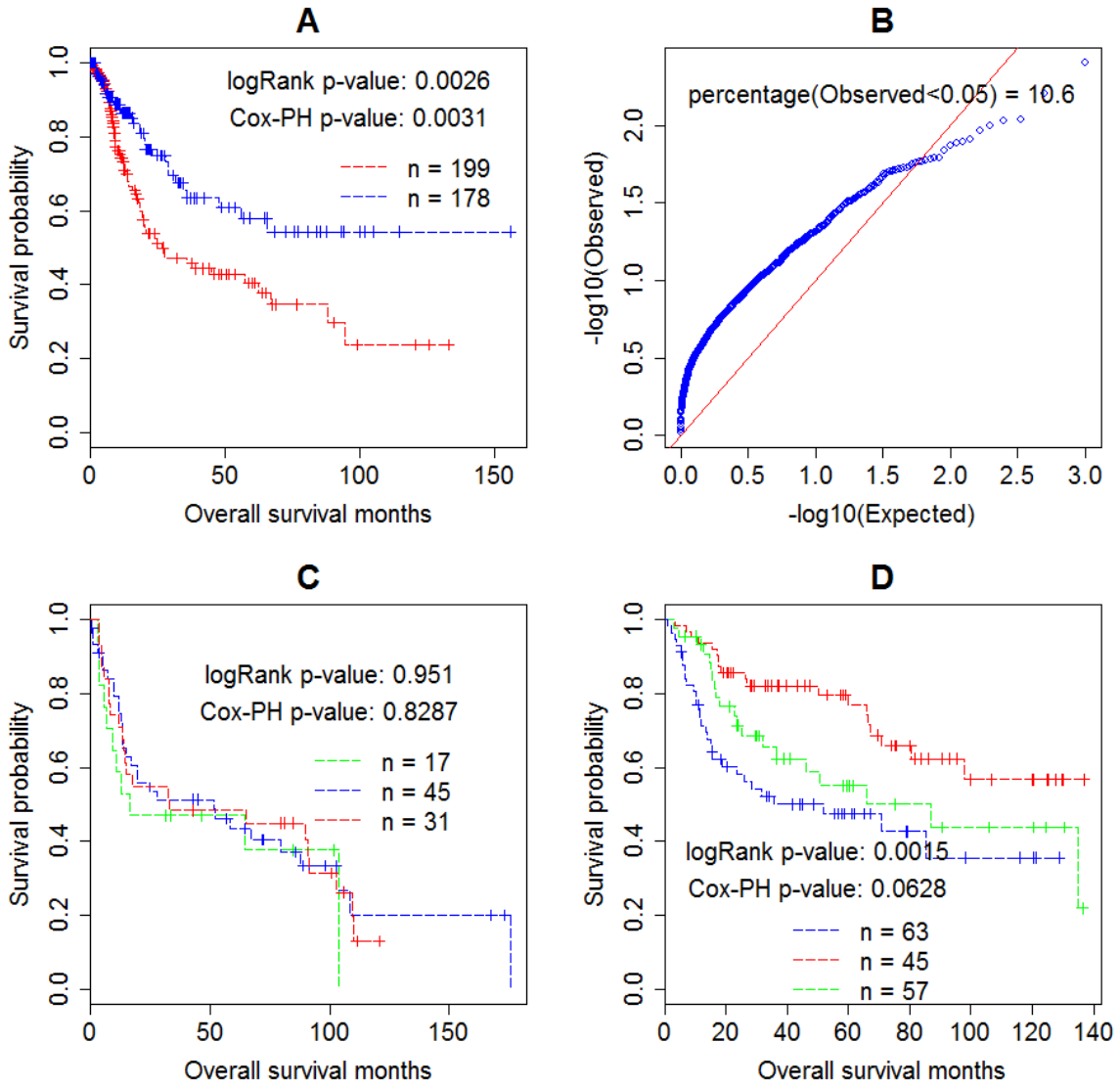


cluster 3

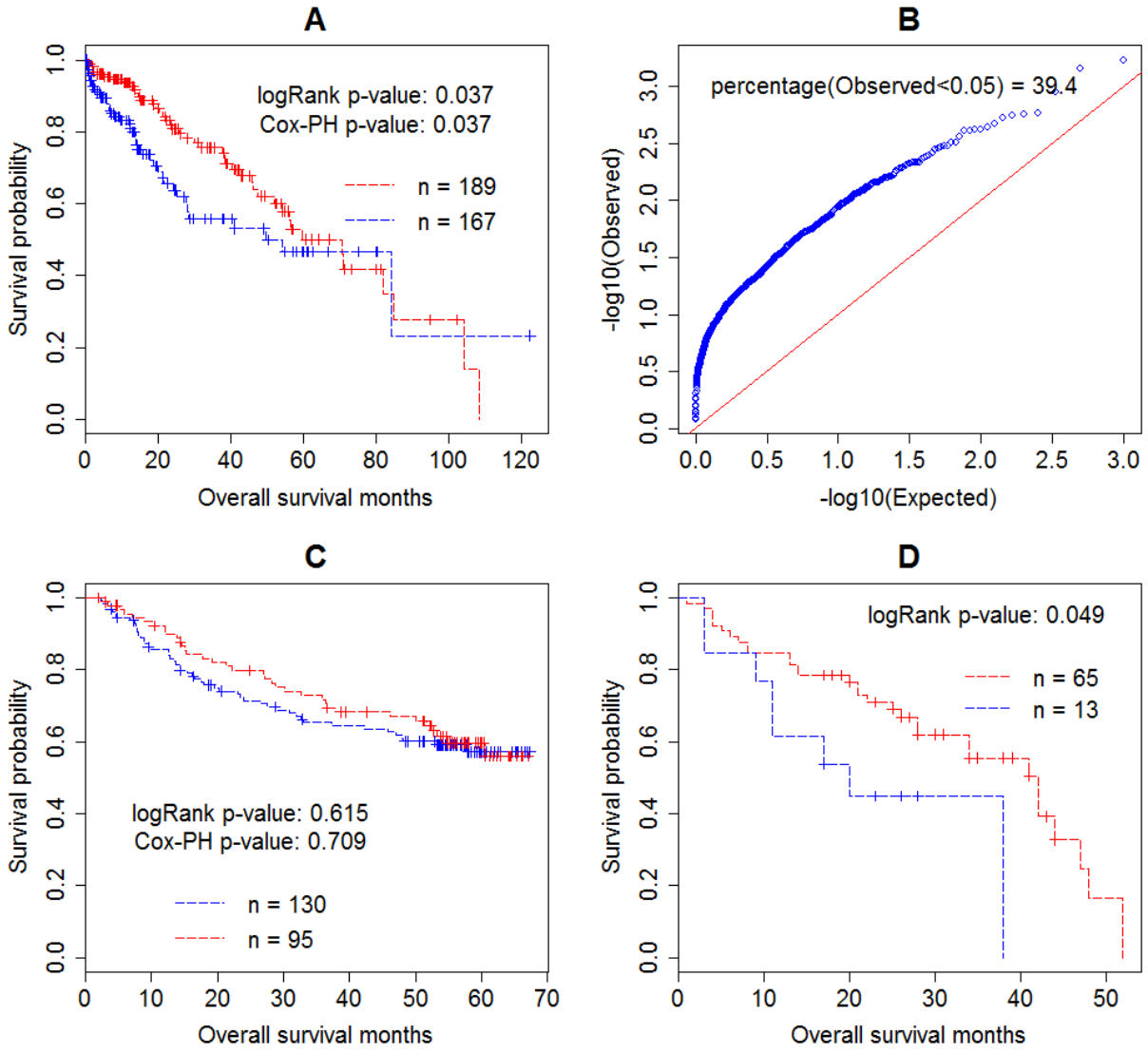


cluster 4





Supplementary Figure Set 5. Prediction strength and robustness of the prognostic signature (219 genes) identified from the result of nbsModel for bladder cancer (BLCA). A, C and D: Clustering-analysis based evaluation of the prediction strength of the signature using the enlarged TCGA dataset, Riester's dataset and Kim' dataset, respectively. P-values are calculated for the comparisons between good (red) and bad (blue) survival clusters. B: The QQ plot for the p-values obtained from 1000 tests. In each test, the SVD-based survival analysis is performed on a randomly sampled dataset that contains 75% of the patients in the enlarged TCGA data.



Supplementary Figure Set 6 Prediction strength and robustness of the prognostic signature (25 genes) identified from the clustering result of nbsModel for liver cancer (LIHC). A, C and D: Clustering analysis based evaluation of the prediction strength of the signature using the enlarged TCGA dataset, Roessler's dataset and Villa's dataset, respectively. B: The QQ plot for the p-values obtained from 1000 tests. In each test, the SVD-based survival analysis is performed on a randomly sampled dataset that contains 75% of the patients in the enlarged TCGA data.