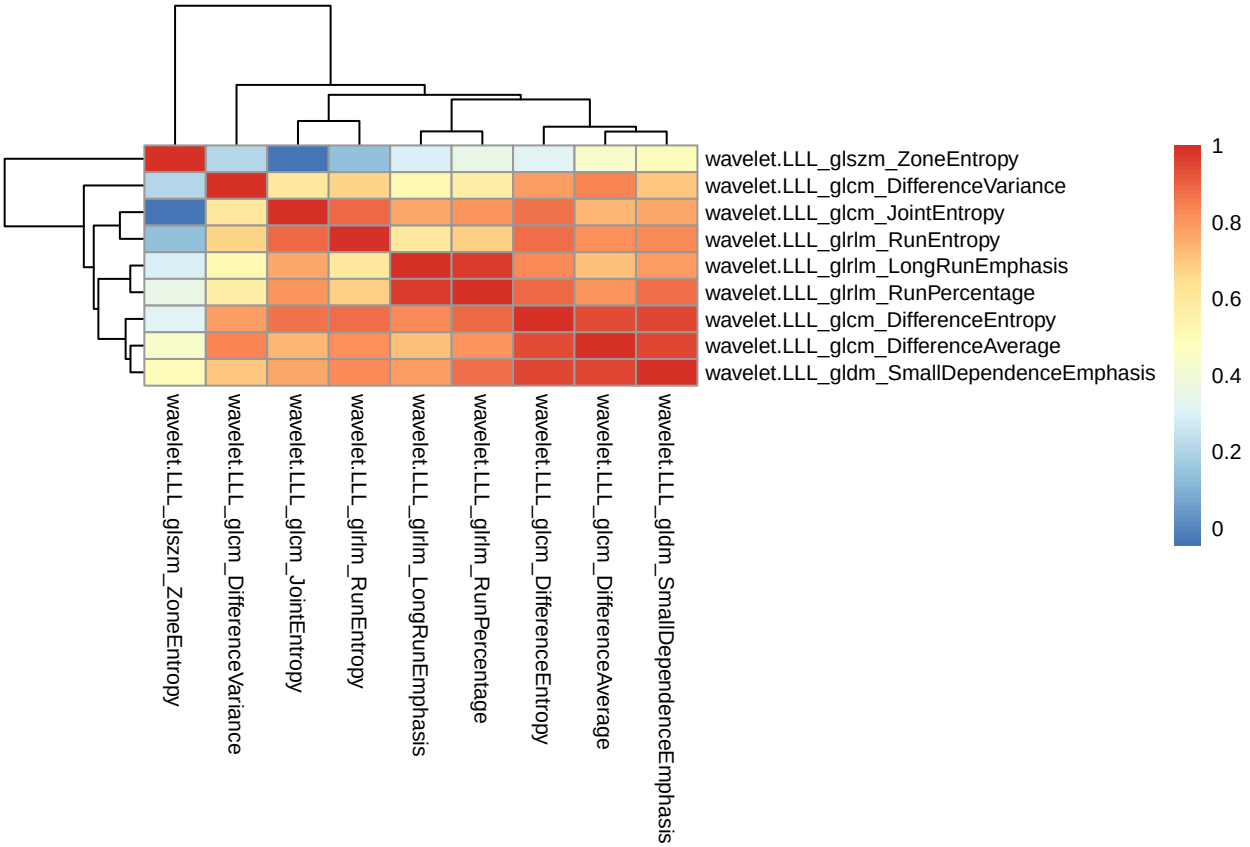
Radiomics

Loading Everything

Project Plan

What features to take?

- Not sure to take Zone Entropy or not
- Run Entropy and Zone Entropy do not follow the same patter on biplot



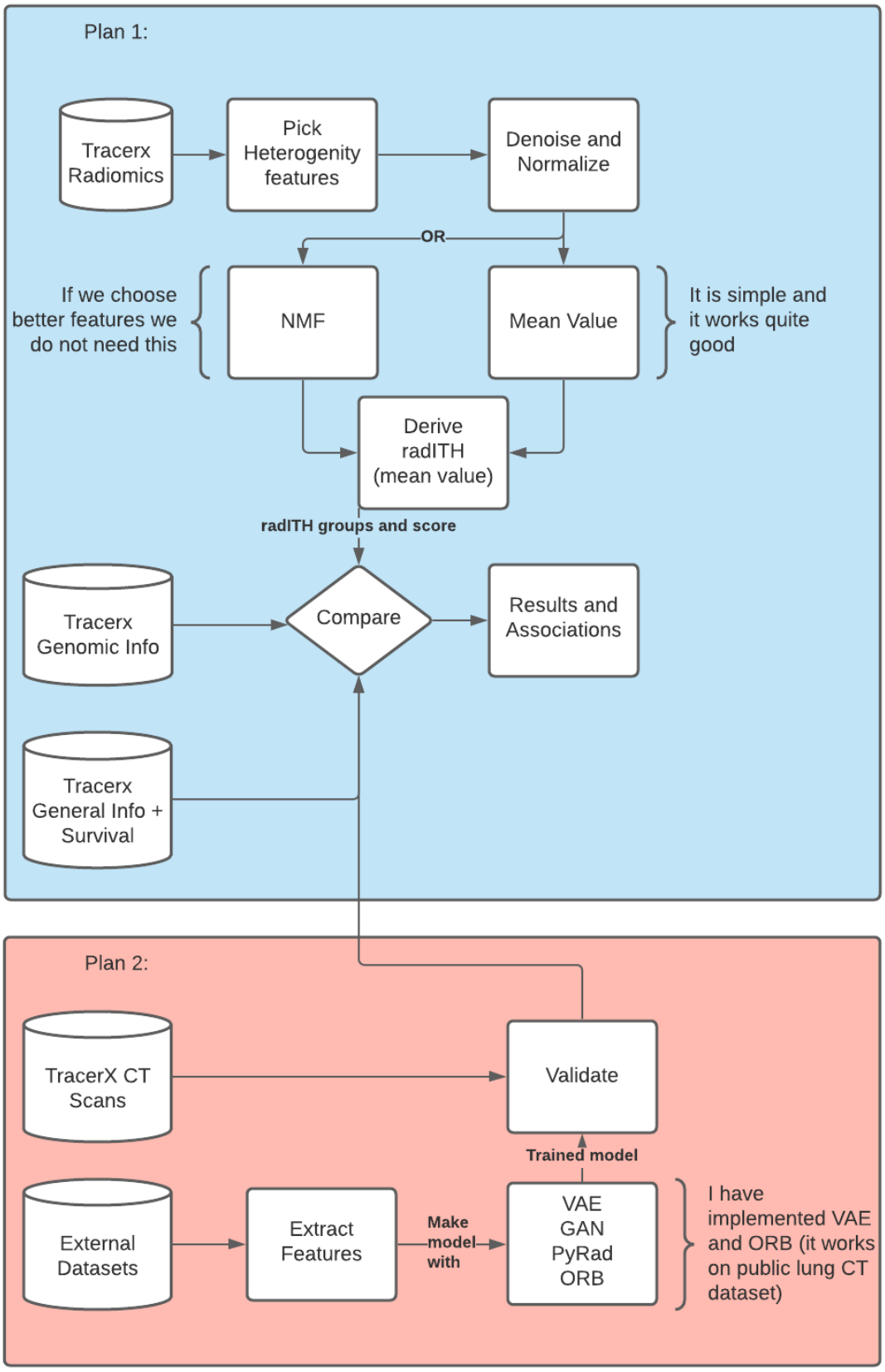
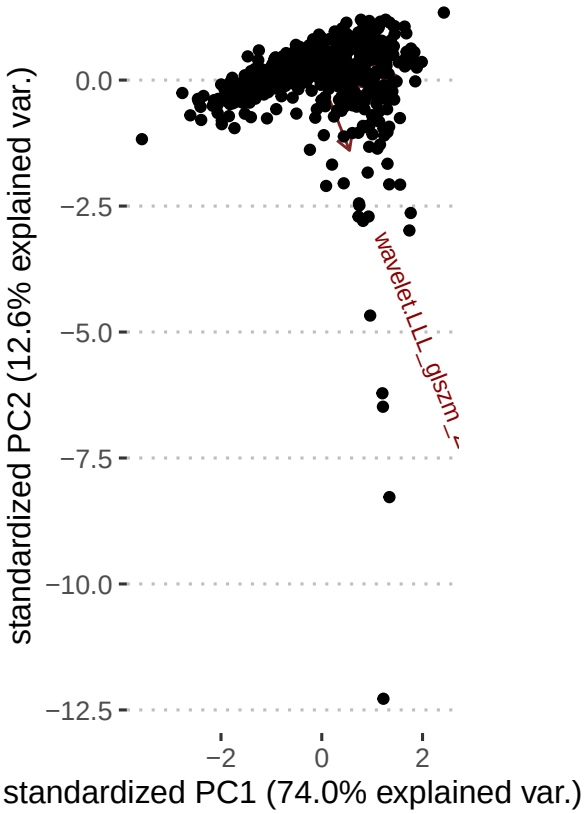


Figure 1: Project Plan



Important measures to check

- We will compare diameter to volume to ITH

```
pyrad$volume = as.numeric(pyrad$original_shape_MeshVolume)
pyrad$volume_from_pyrad = as.numeric(pyrad$original_shape_MeshVolume)
pyrad$diameter = as.numeric(pyrad$original_shape_Maximum2DDiameterSlice)
```

how to define radITH

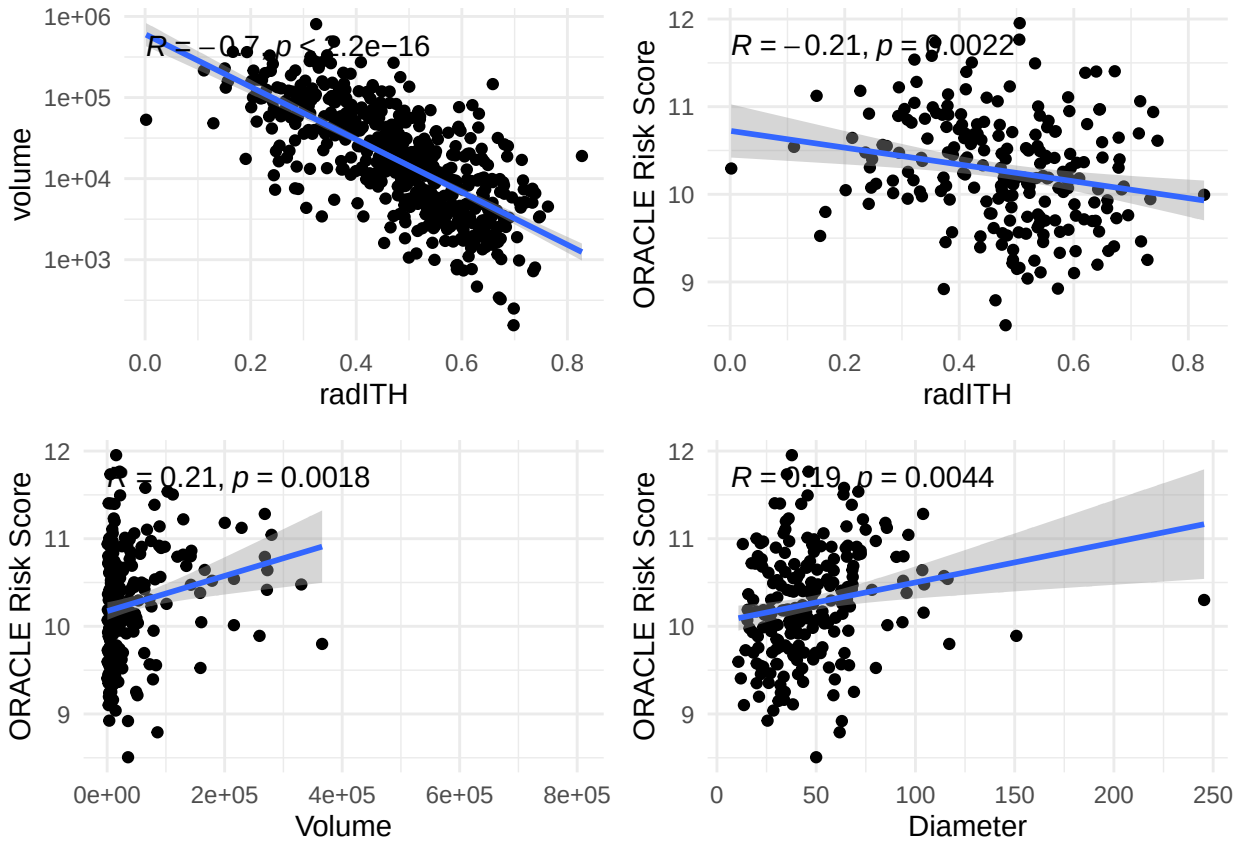
- Do we need to normalize something by volume?
- Numbers were a bit wierd when divided by volume therefore I did not divide anything with volume

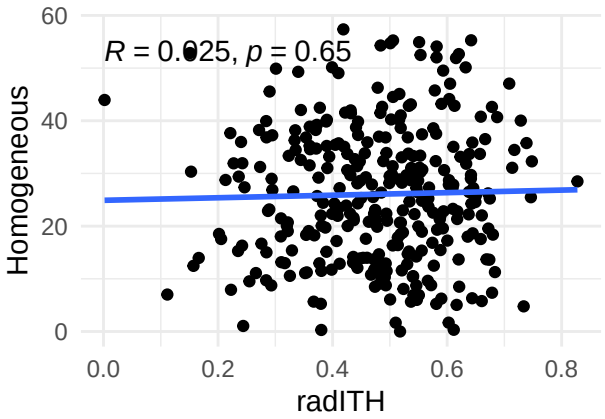
```
# weighted mean
w = 1-abs(cor(pyrad[,features_of_interest], pyrad[,"volume"]))
pyrad$radITH = apply(pyrad[,features_of_interest],1, function(x){
   weighted.mean(x[features_of_interest], w = w)
})


#pyrad$radITH = rowMeans(pyrad[,features_of_interest], na.rm = T)
```

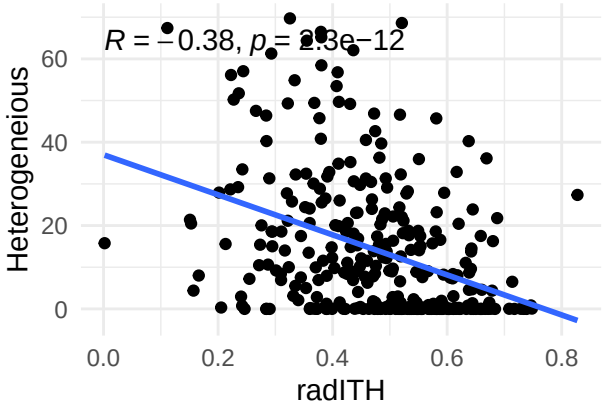
```
Q = 3


pyrad$volume_group = gtools::quantcut(pyrad$volume, q=Q, na.rm=TRUE)
pyrad$diameter_group = gtools::quantcut(pyrad$diameter, q=Q, na.rm=TRUE)
pyrad$radITH_group = gtools::quantcut(pyrad$radITH, q=Q, na.rm=TRUE)
```

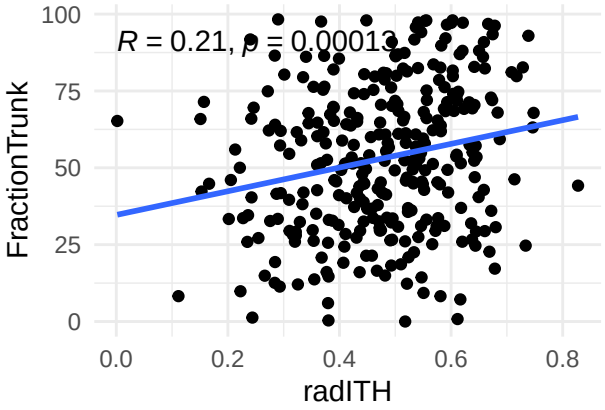
Expected correlations

- Negative cor radITH to volume









Mutations

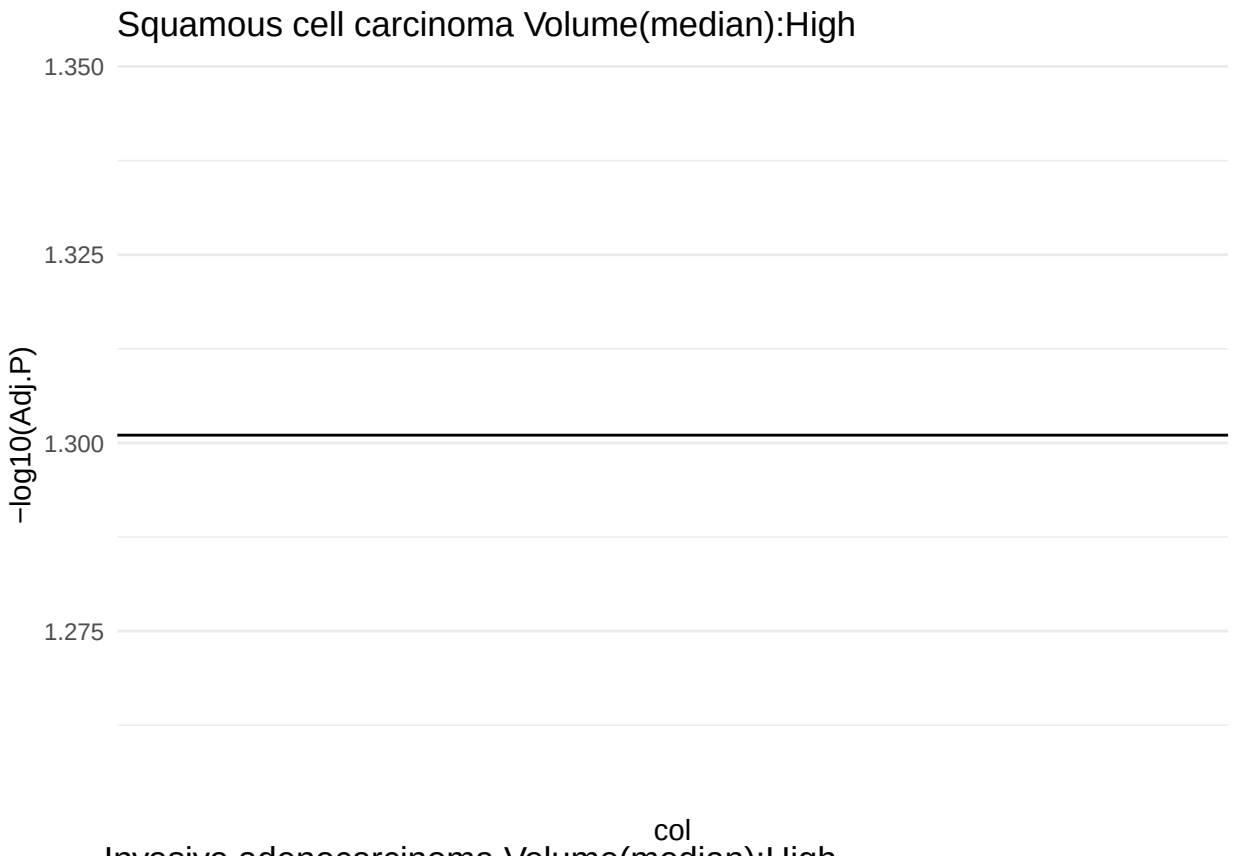
Let's group DRIVER mutations by Sanchez Vega def

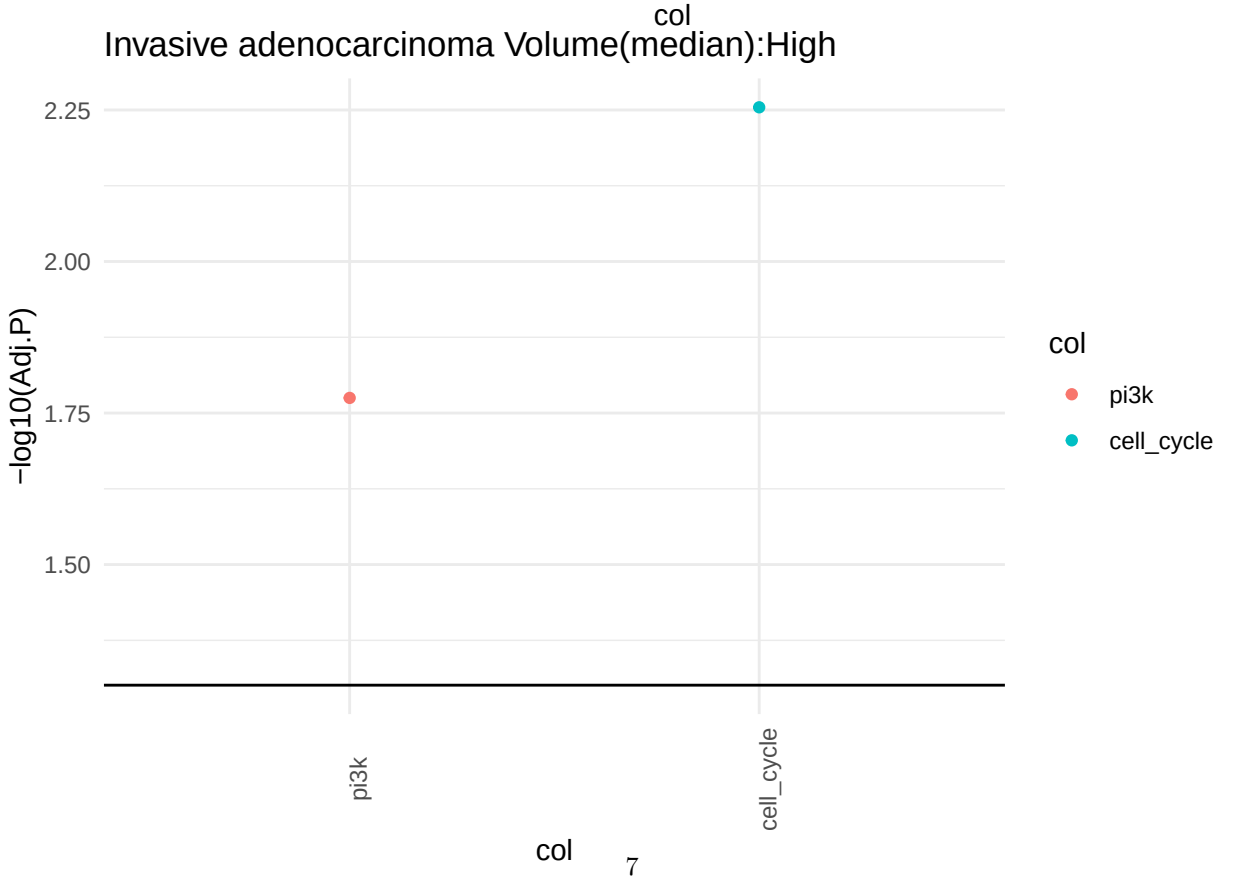
Let's test Sanchez Vega Muts vs radITH groups (q =3)

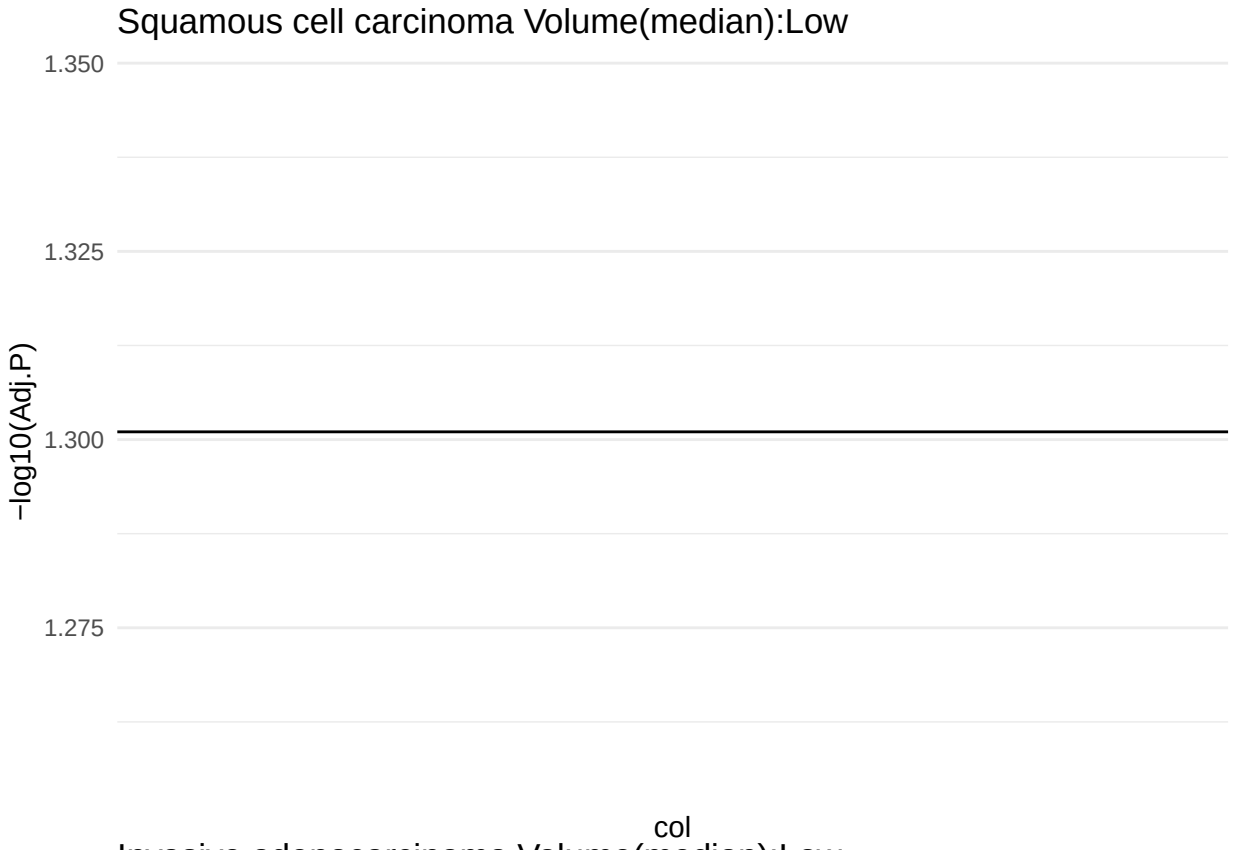
```
## [1] "Adeno fisher test results"

## [1] "pi3k"
##
##  Fisher's Exact Test for Count Data
##
## data:  table(tmp$radITH_group, tmp[, col])
## p-value = 0.0007203
## alternative hypothesis: two.sided
##
## [1] "cell_cycle"
##
##  Fisher's Exact Test for Count Data
##
## data:  table(tmp$radITH_group, tmp[, col])
## p-value = 0.004677
## alternative hypothesis: two.sided
```

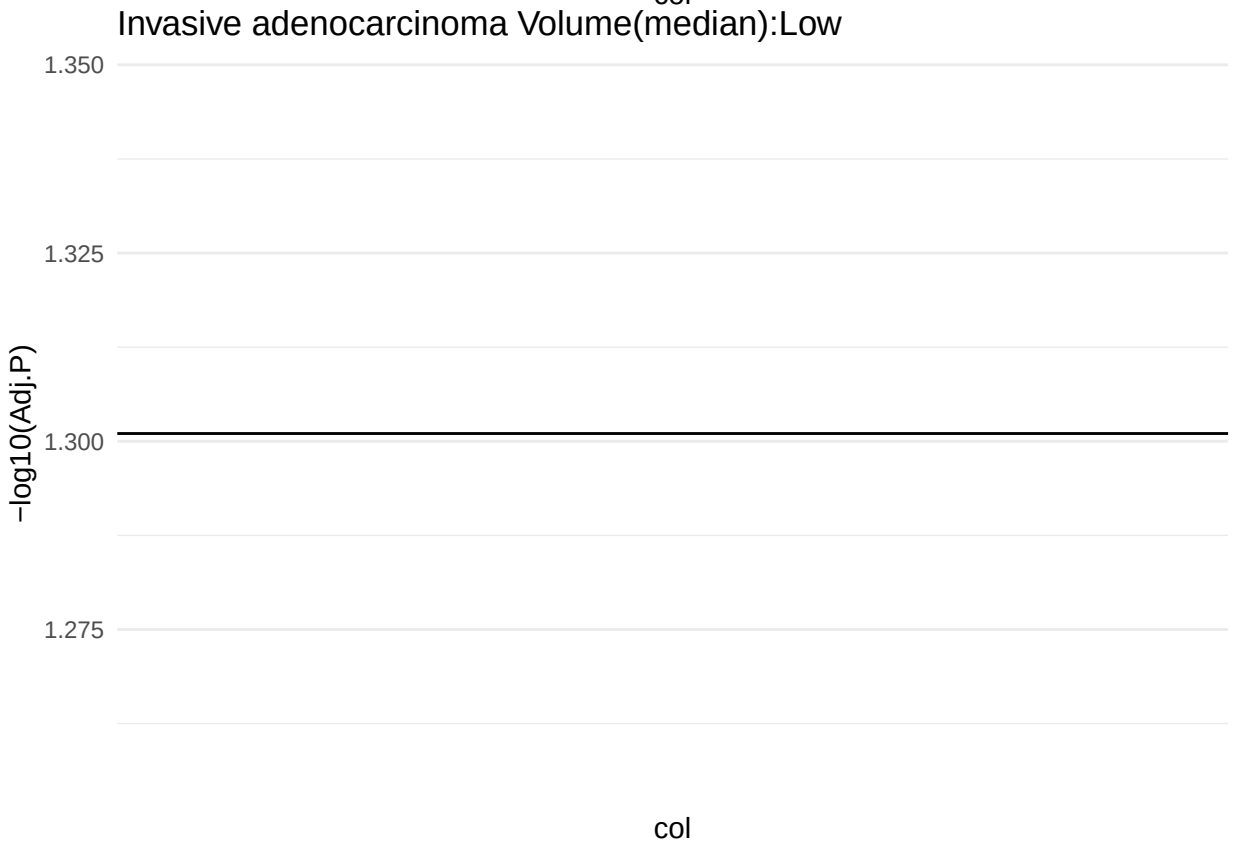
```
## [1] "Squamous fisher test results"
```

SanchezVega vs radITH groups (q = 3) vs pathology vs volume (Median)






Squamous cell carcinoma Volume(median):Low
1.350
1.325
1.300
1.275
−log10(Adj.P)
col


Invasive adenocarcinoma Volume(median):Low
1.350
1.325
1.300
1.275
−log10(Adj.P)
col

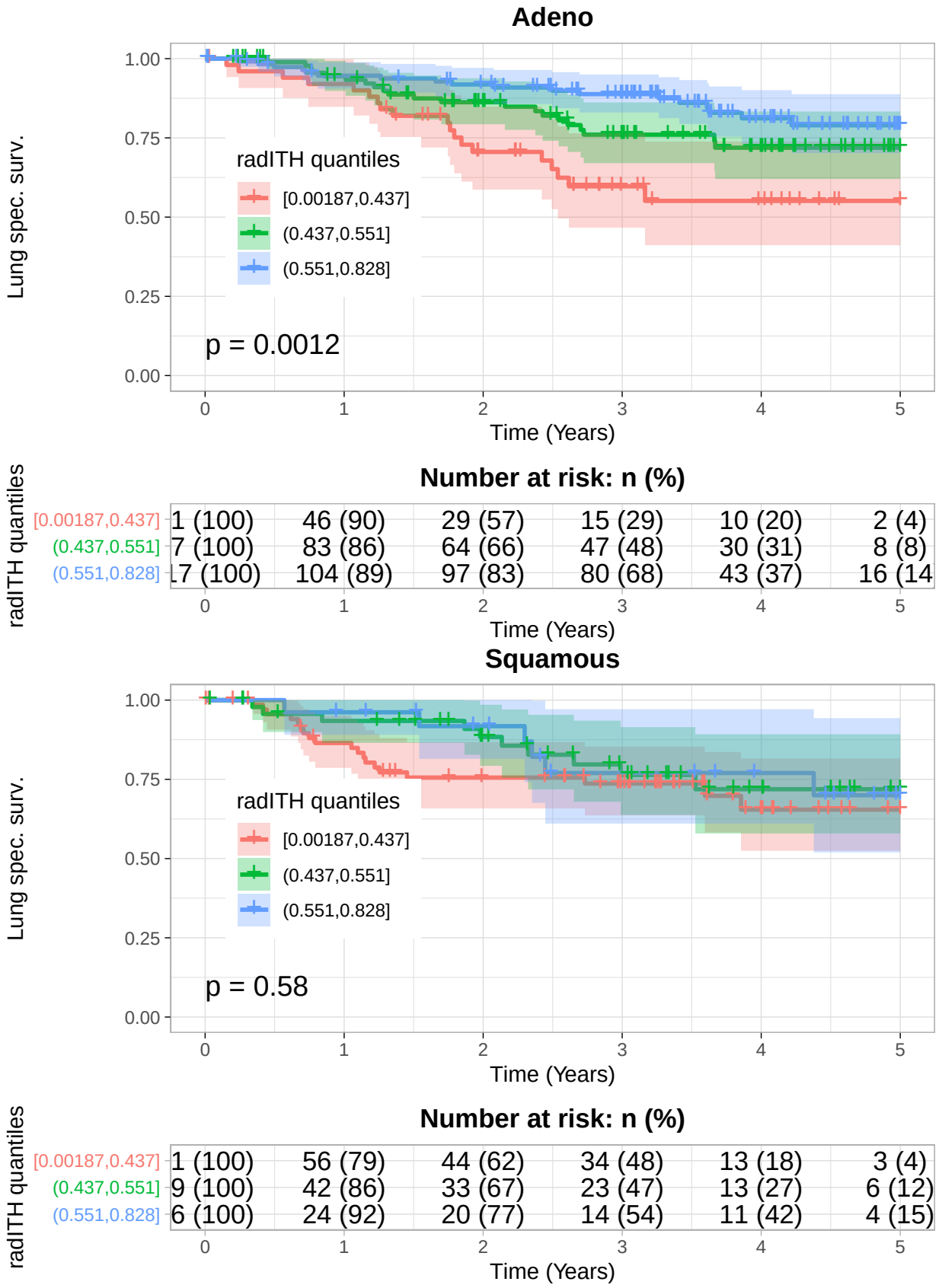
Does Volume or diameter predict biology?

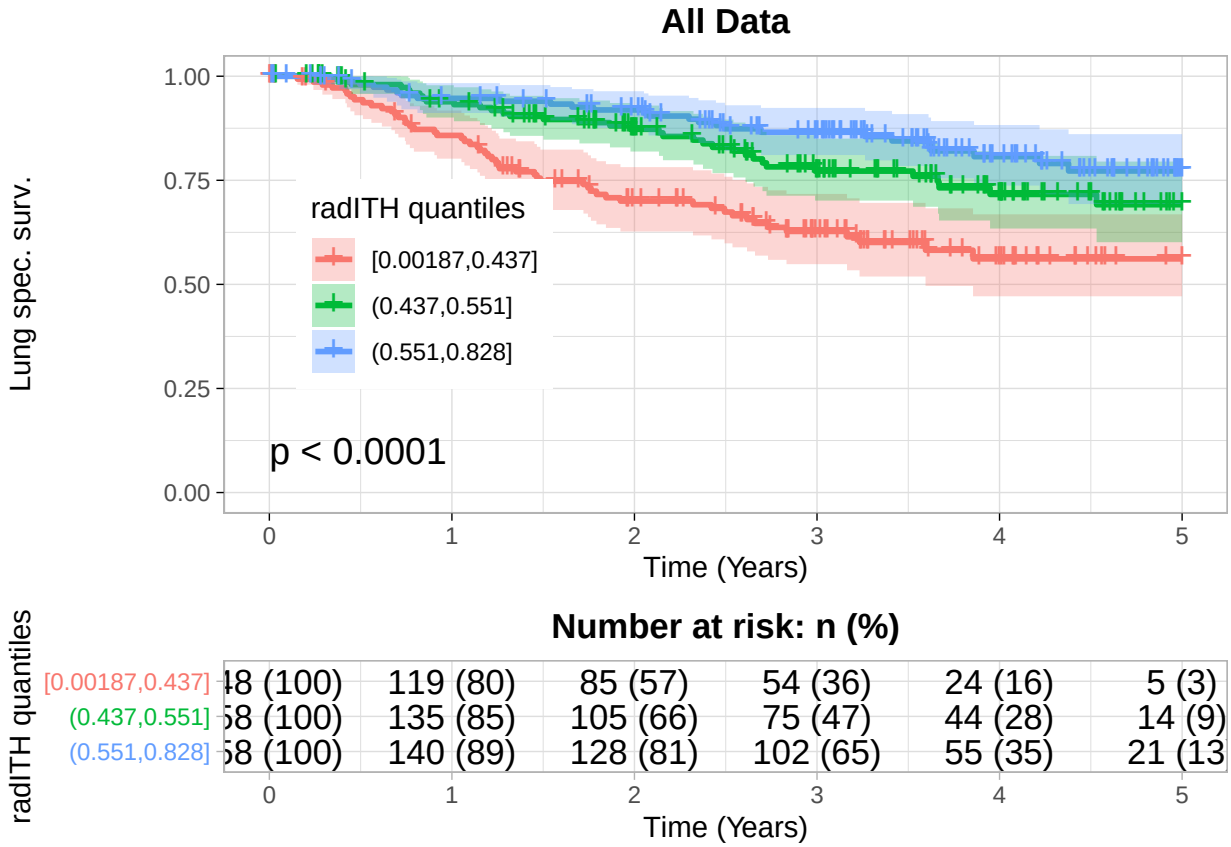
- Diameter is not associated at all
- Volume is associated with HIPPO

```
## [1] "Adeno fisher test results"
```

```
## [1] "Squamous fisher test results"
```

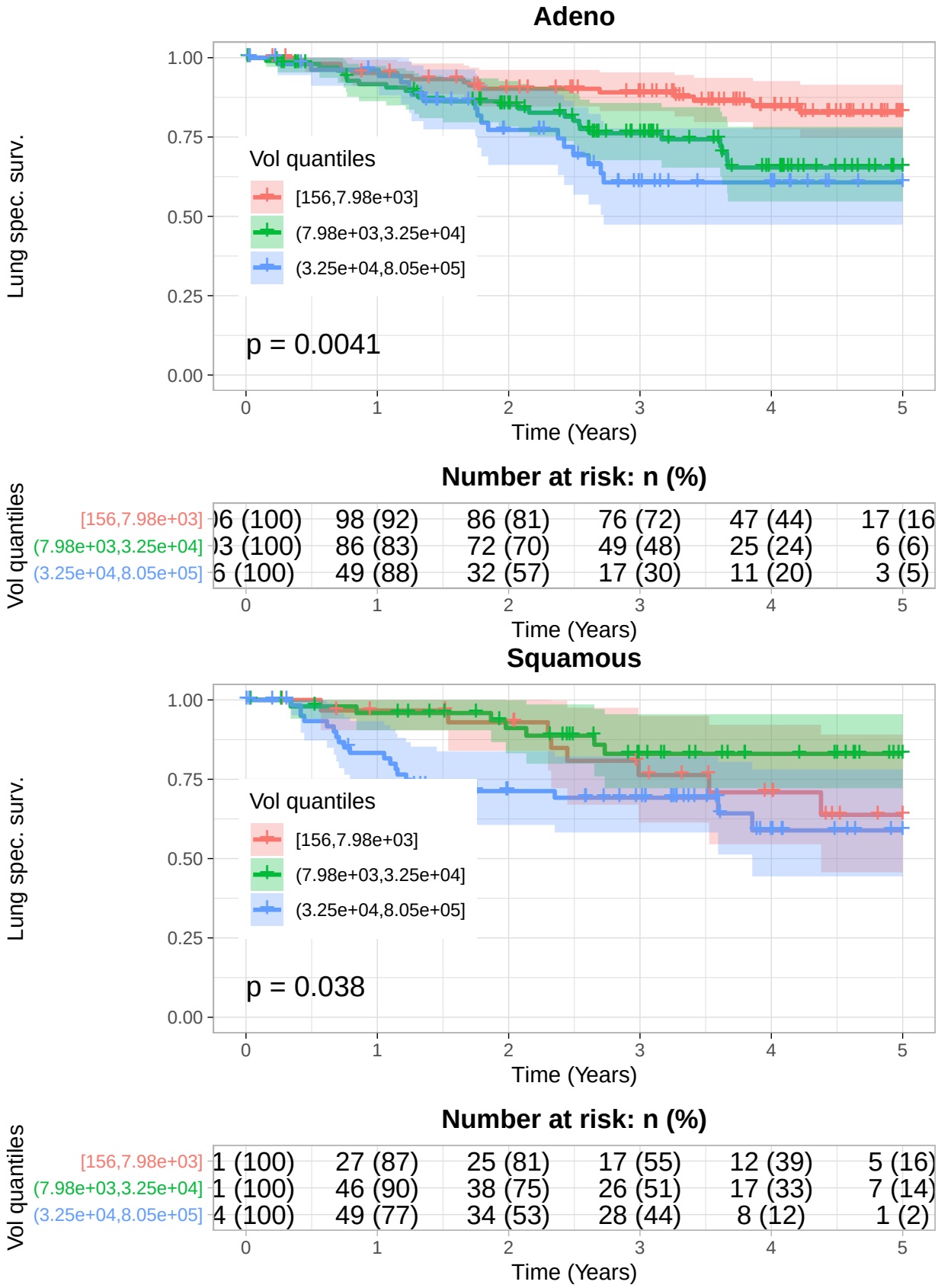
How does radITH associate with survival?

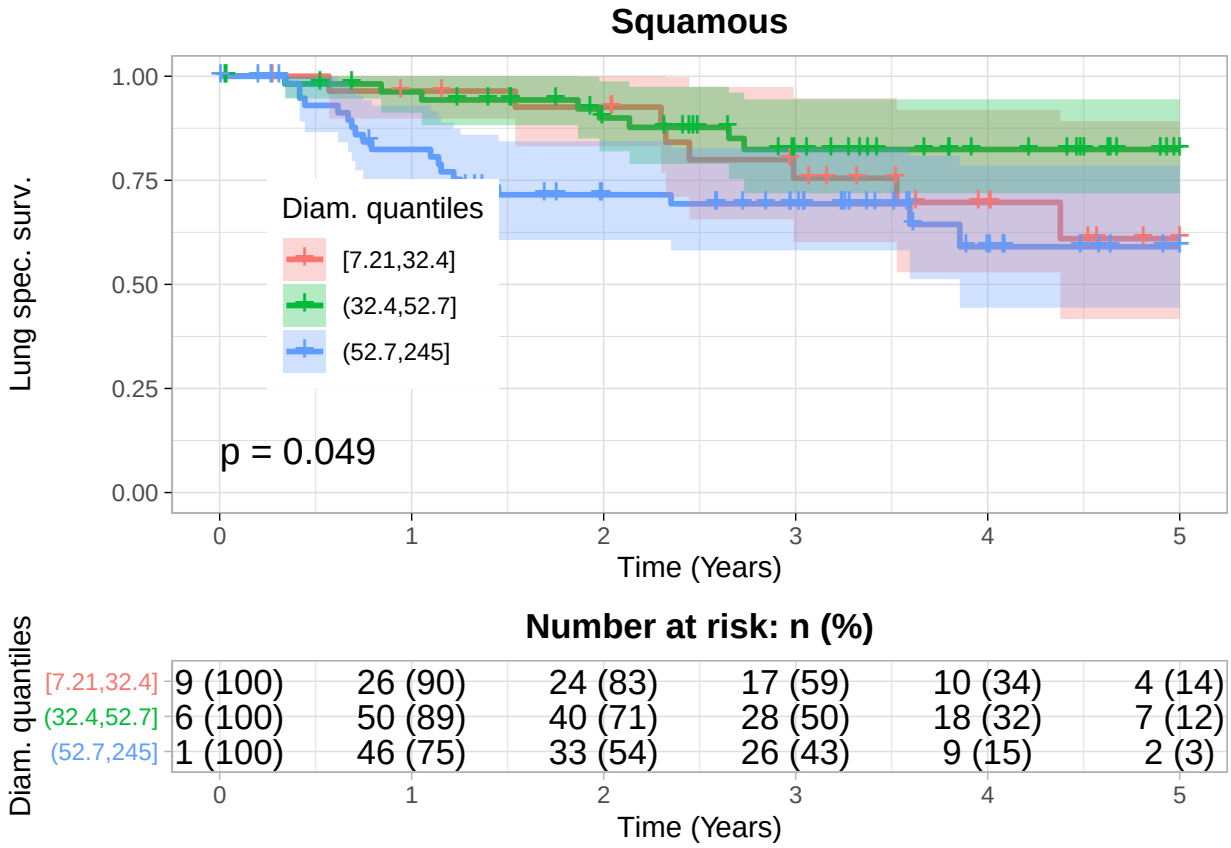



All Data
Lung spec. surv.
1.00
0.75
0.50
0.25
0.00
radITH quantiles
[0.00187,0.437]
(0.437,0.551]
(0.551,0.828]
p < 0.0001
0
1
2
3
4
5
Time (Years)

Number at risk: n (%)
radITH quantiles
[0.00187,0.437]
(0.437,0.551]
(0.551,0.828]
48 (100)
58 (100)
58 (100)
119 (80)
135 (85)
140 (89)
85 (57)
105 (66)
128 (81)
54 (36)
75 (47)
102 (65)
24 (16)
44 (28)
55 (35)
5 (3)
14 (9)
21 (13
0
1
2
3
4
5
Time (Years)

How does volume (diameter) associate to survival?

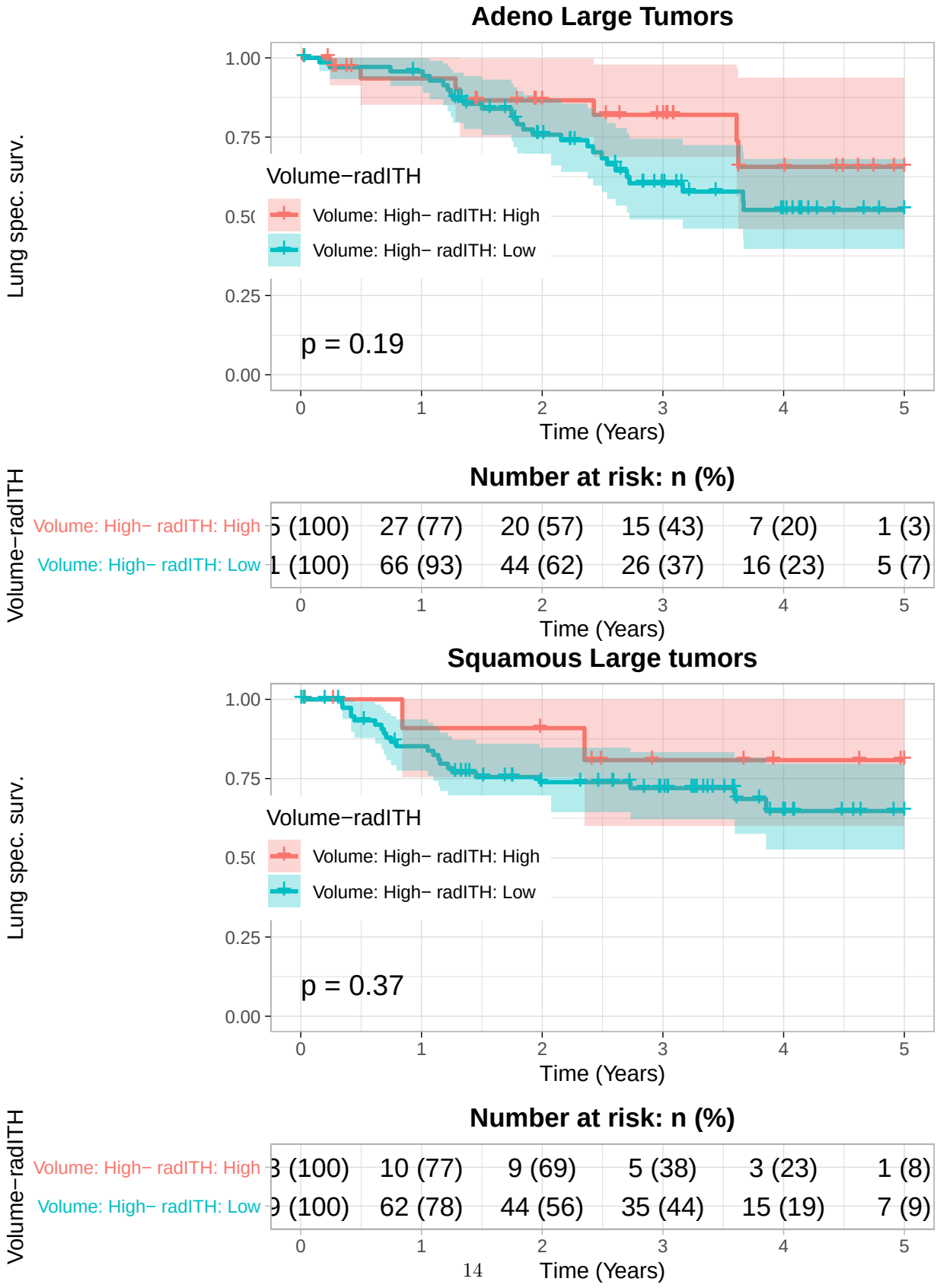


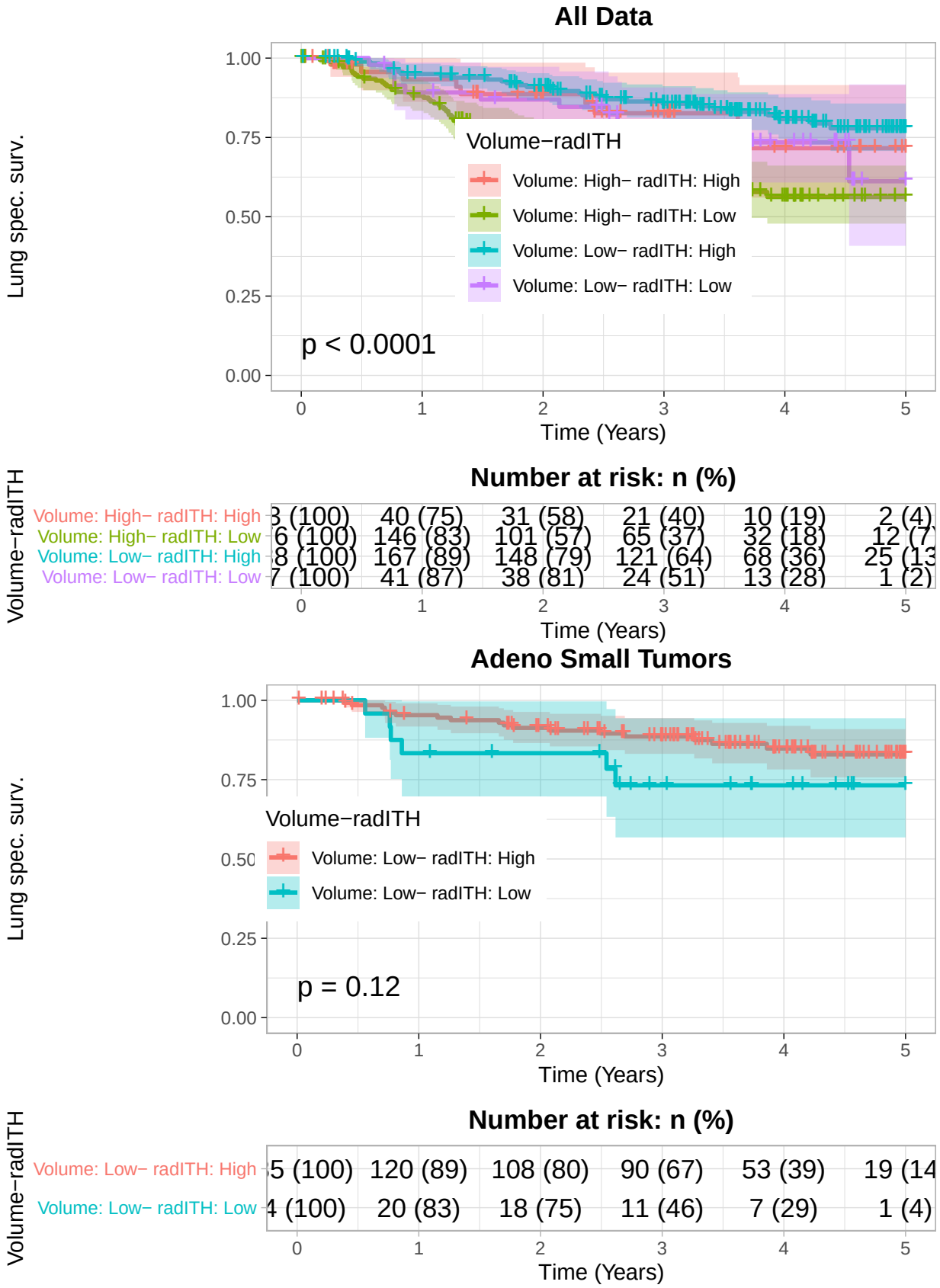

Squamous
1.00
0.75
0.50
0.25
0.00
Lung spec. surv.
Diam. quantiles
[7.21,32.4]
(32.4,52.7]
(52.7,245]
p = 0.049
0
1
2
3
4
5
Time (Years)

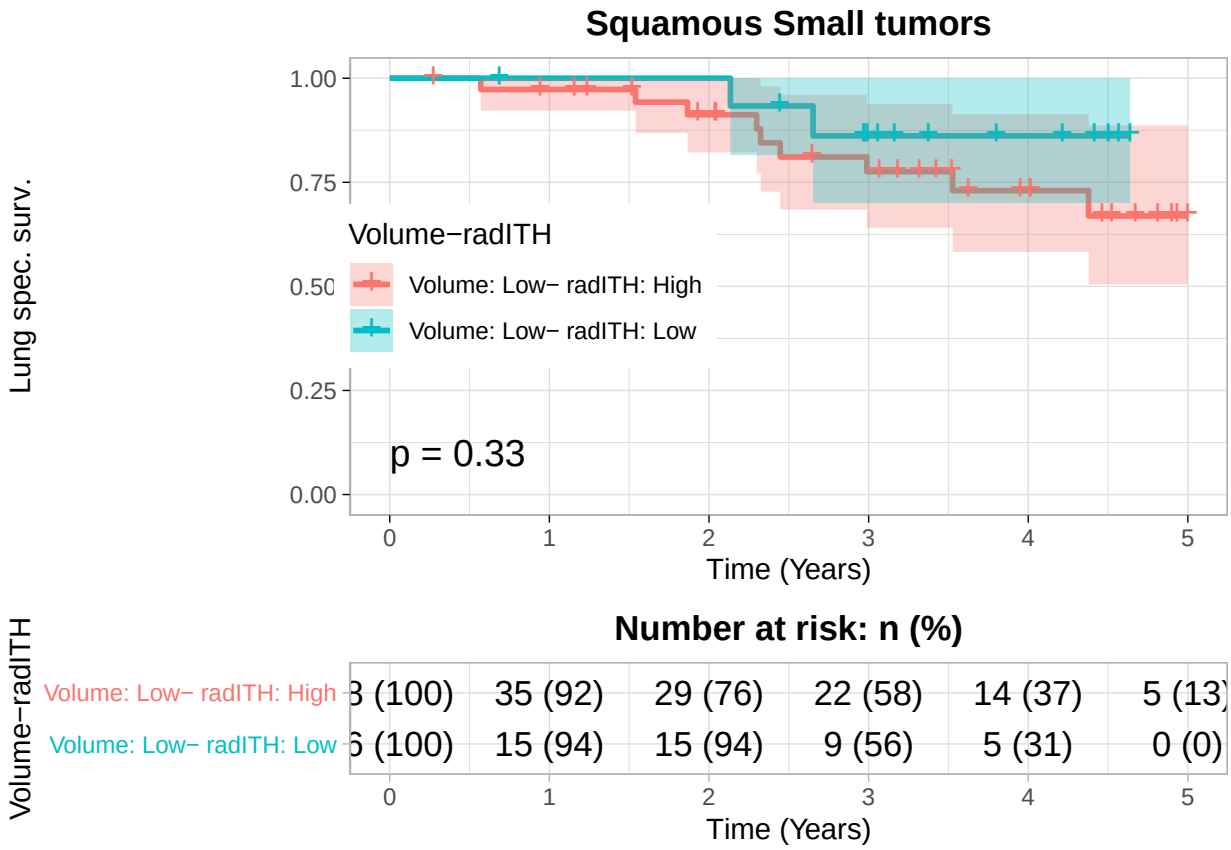
Number at risk: n (%)
Diam. quantiles
[7.21,32.4] 9 (100) 26 (90) 24 (83) 17 (59) 10 (34) 4 (14)
(32.4,52.7] 6 (100) 50 (89) 40 (71) 28 (50) 18 (32) 7 (12)
(52.7,245] 1 (100) 46 (75) 33 (54) 26 (43) 9 (15) 2 (3)
0
1
2
3
4
5
Time (Years)

Can we overlap radITH and Volume groups and check survival?

In order to increase group sizes, all measures will be split by median (Q=2)

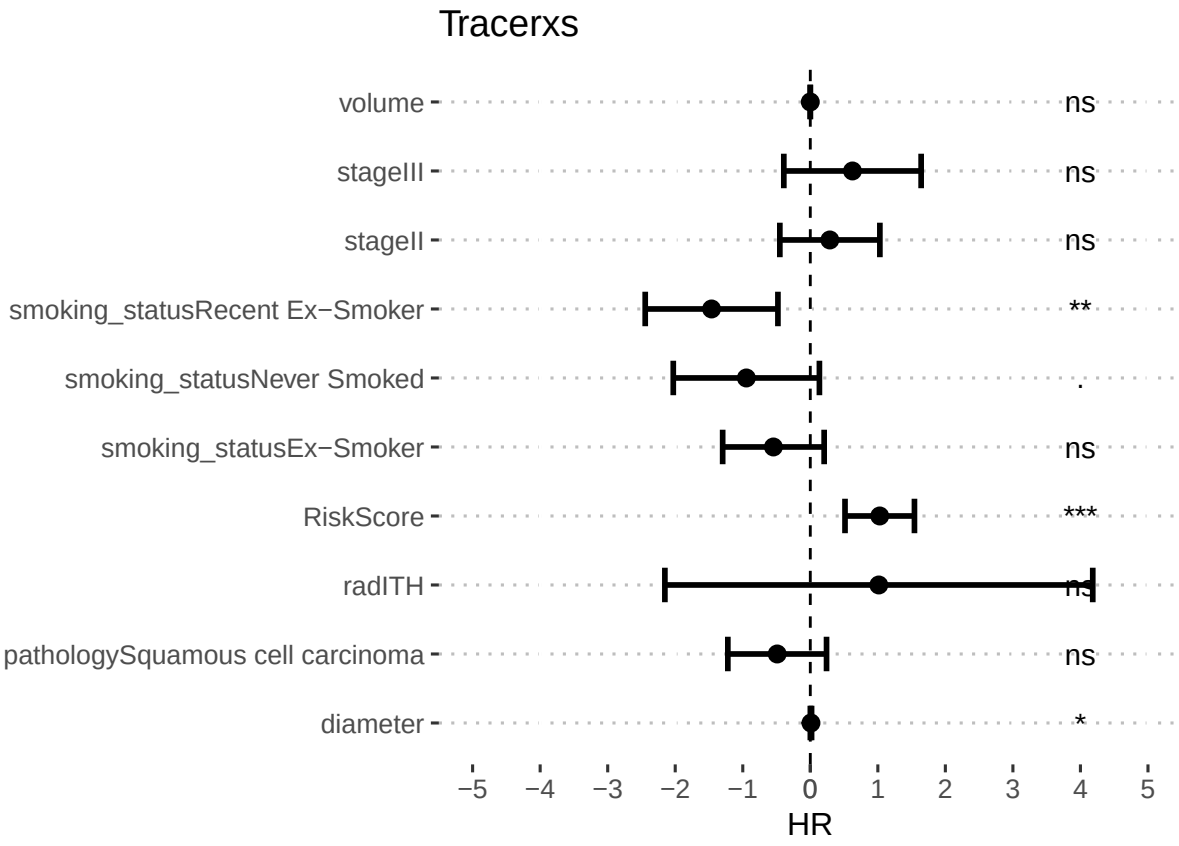



All Data
Lung spec. surv.
1.00
0.75
0.50
0.25
0.00
Volume−radITH
Volume: High− radITH: High
Volume: High− radITH: Low
Volume: Low− radITH: High
Volume: Low− radITH: Low
p < 0.0001
0
1
2
3
4
5
Time (Years)
Number at risk: n (%)
Volume−radITH
Volume: High− radITH: High
Volume: High− radITH: Low
Volume: Low− radITH: High
Volume: Low− radITH: Low
3 (100)
40 (75)
31 (58)
21 (40)
10 (19)
2 (4)
6 (100)
146 (83)
101 (57)
65 (37)
32 (18)
12 (7
8 (100)
167 (89)
148 (79)
121 (64)
68 (36)
25 (13
7 (100)
41 (87)
38 (81)
24 (51)
13 (28)
1 (2)
0
1
2
3
4
5
Time (Years)
Adeno Small Tumors
Lung spec. surv.
1.00
0.75
0.50
0.25
0.00
Volume−radITH
Volume: Low− radITH: High
Volume: Low− radITH: Low
p = 0.12
0
1
2
3
4
5
Time (Years)
Number at risk: n (%)
Volume−radITH
Volume: Low− radITH: High
Volume: Low− radITH: Low
5 (100)
120 (89)
108 (80)
90 (67)
53 (39)
19 (14
4 (100)
20 (83)
18 (75)
11 (46)
7 (29)
1 (4)
0
1
2
3
4
5
Time (Years)



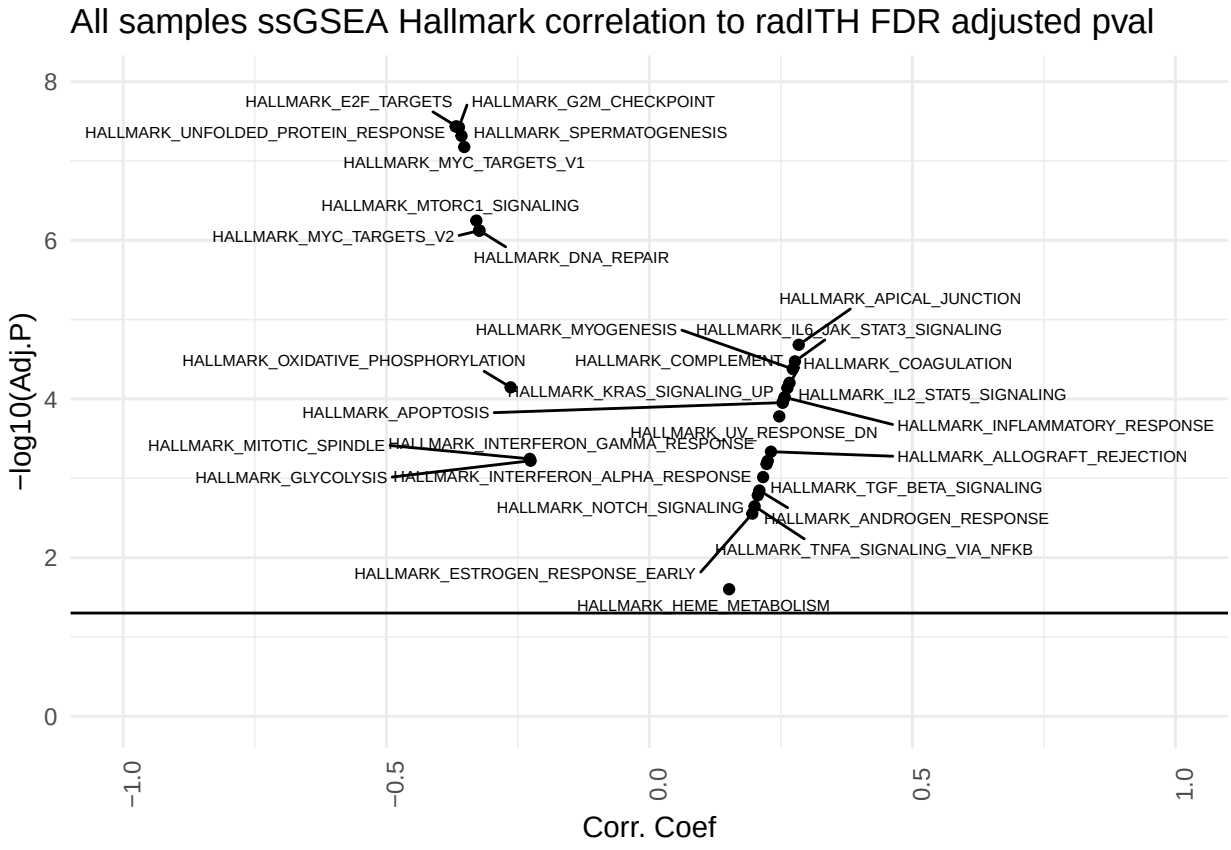
Coxph Model

- radITH does not help improve cox ph model

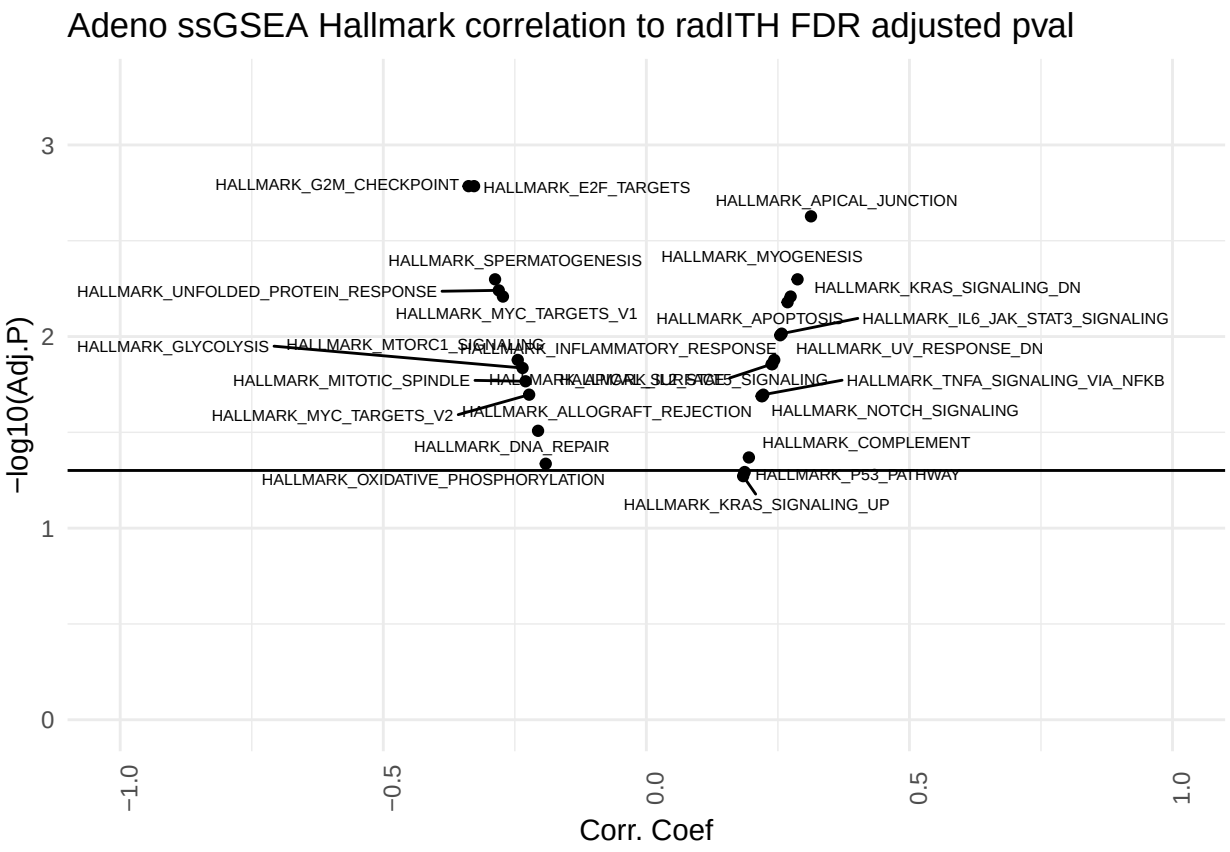


Hallmarks all samples

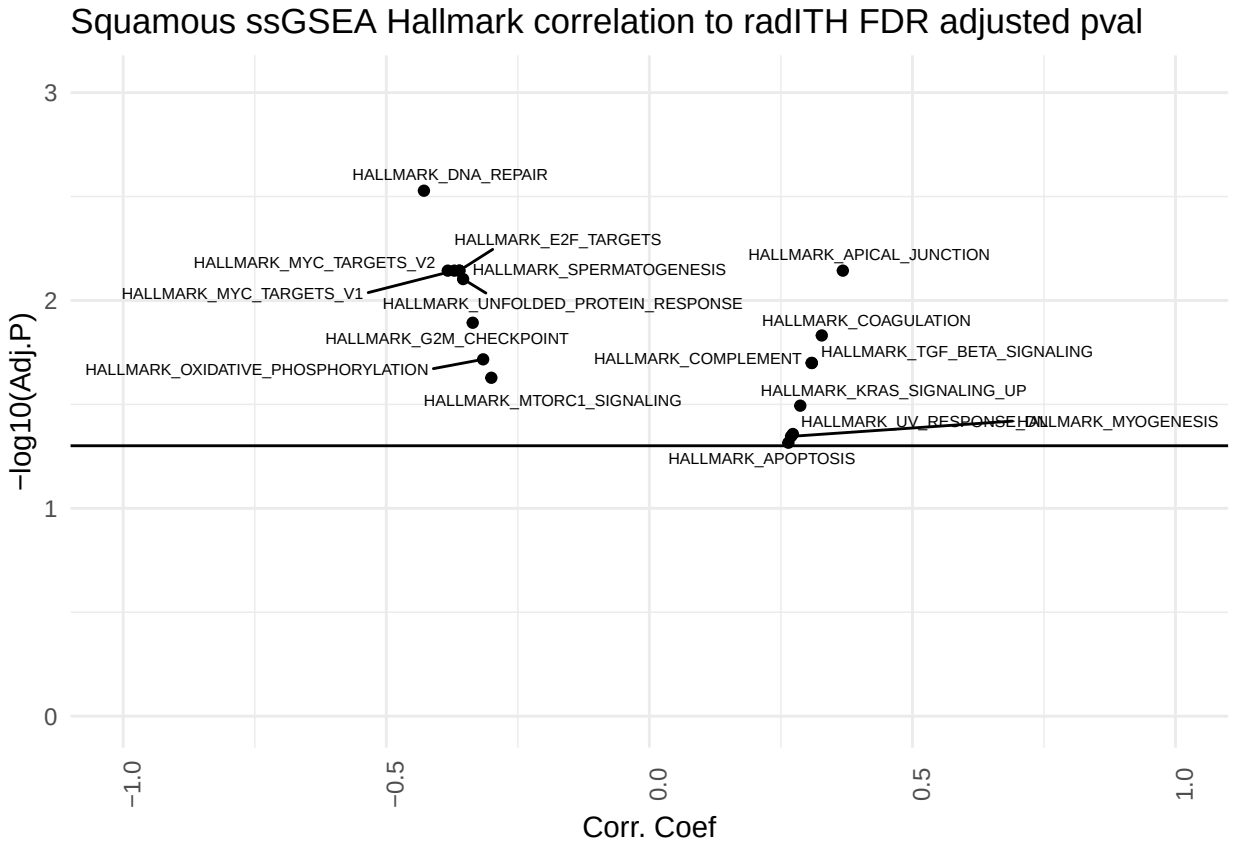
- Association (cor) of radITH with hallmarks
- Hallmarks computed with SS-GSEA
- P values are adjusted using FDR


All samples ssGSEA Hallmark correlation to radITH FDR adjusted pval
−log10(Adj.P)
Corr. Coef
8
6
4
2
0
−1.0
−0.5
0.0
0.5
1.0
HALLMARK_E2F_TARGETS
HALLMARK_G2M_CHECKPOINT
HALLMARK_UNFOLDED_PROTEIN_RESPONSE
HALLMARK_SPERMATOGENESIS
HALLMARK_MYC_TARGETS_V1
HALLMARK_MTORC1_SIGNALING
HALLMARK_MYC_TARGETS_V2
HALLMARK_DNA_REPAIR
HALLMARK_APICAL_JUNCTION
HALLMARK_MYOGENESIS
HALLMARK_IL6_JAK_STAT3_SIGNALING
HALLMARK_OXIDATIVE_PHOSPHORYLATION
HALLMARK_COMPLEMENT
HALLMARK_COAGULATION
HALLMARK_KRAS_SIGNALING_UP
HALLMARK_IL2_STAT5_SIGNALING
HALLMARK_APOPTOSIS
HALLMARK_INFLAMMATORY_RESPONSE
HALLMARK_UV_RESPONSE_DN
HALLMARK_MITOTIC_SPINDLE
HALLMARK_INTERFERON_GAMMA_RESPONSE
HALLMARK_ALLOGRAFT_REJECTION
HALLMARK_GLYCOLYSIS
HALLMARK_INTERFERON_ALPHA_RESPONSE
HALLMARK_TGF_BETA_SIGNALING
HALLMARK_NOTCH_SIGNALING
HALLMARK_ANDROGEN_RESPONSE
HALLMARK_TNFA_SIGNALING_VIA_NFKB
HALLMARK_ESTROGEN_RESPONSE_EARLY
HALLMARK_HEME_METABOLISM

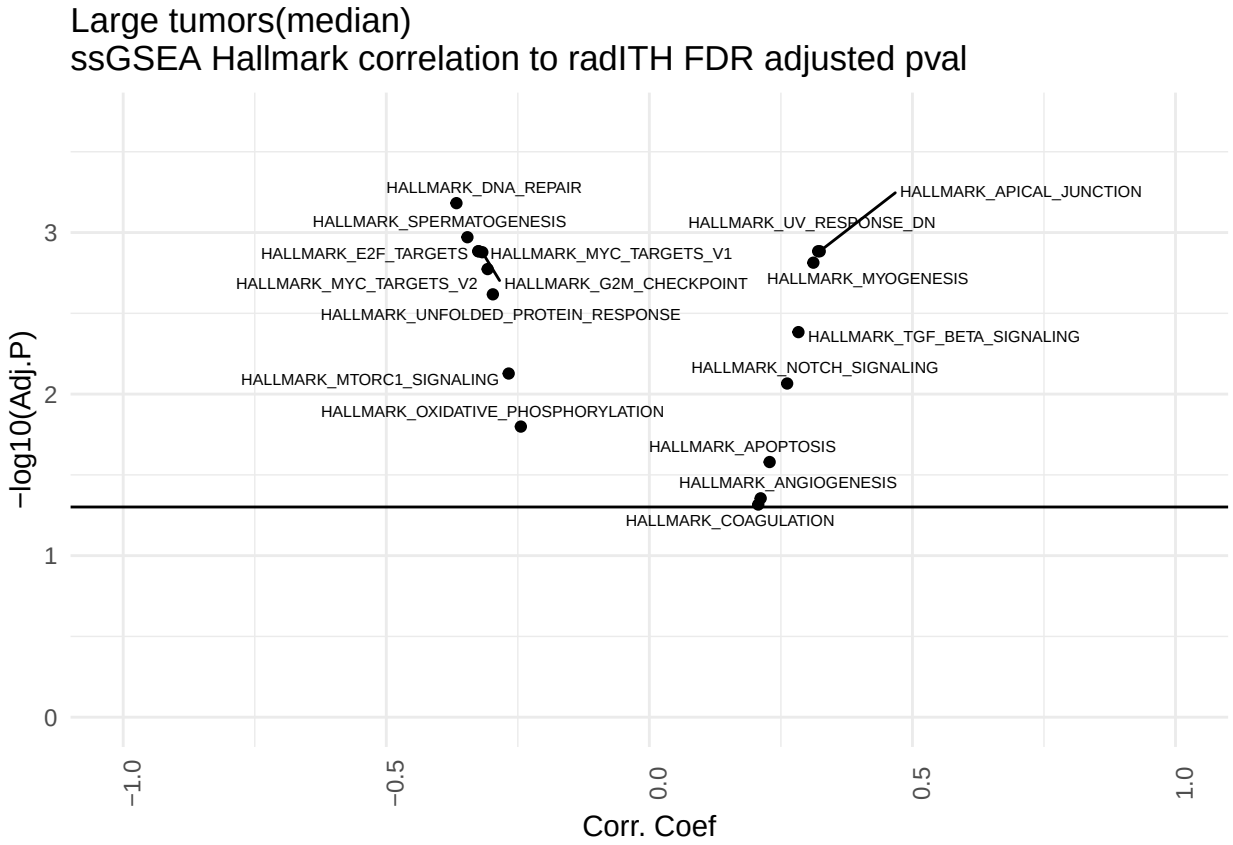
Hallmarks Adeno

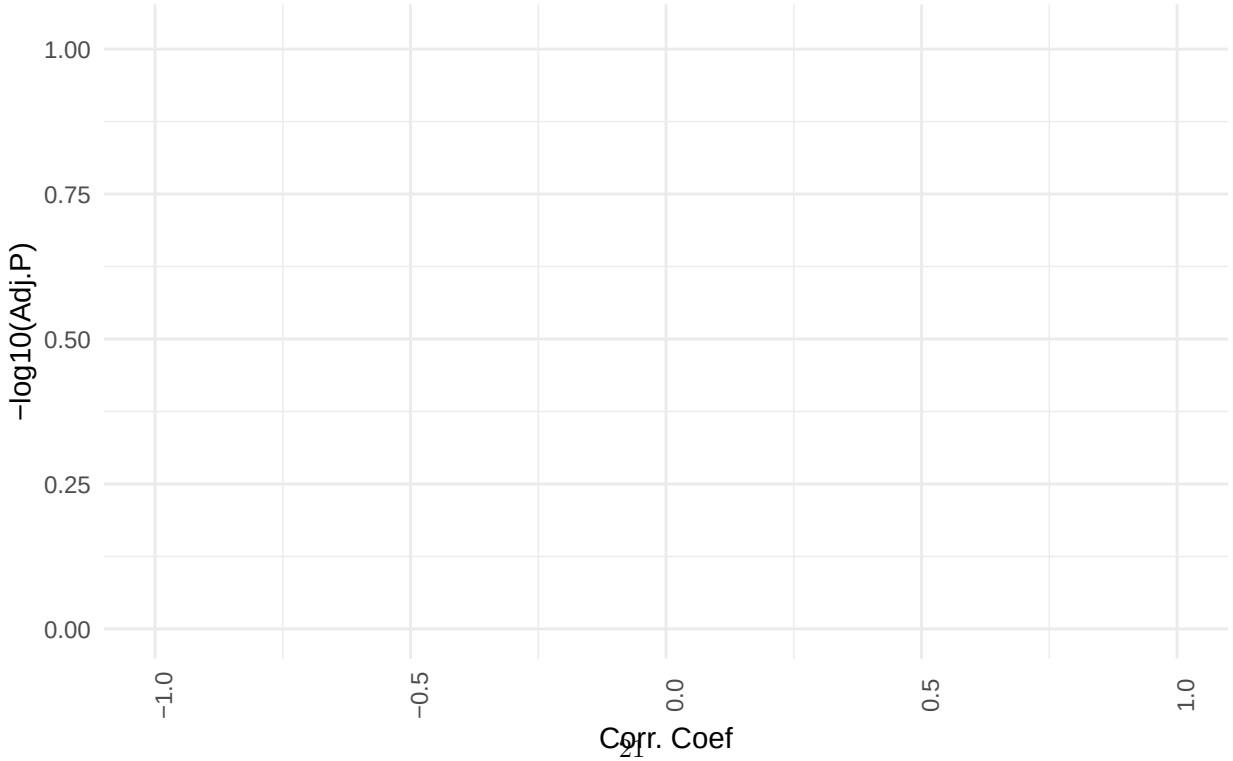


Hallmarks Squamous

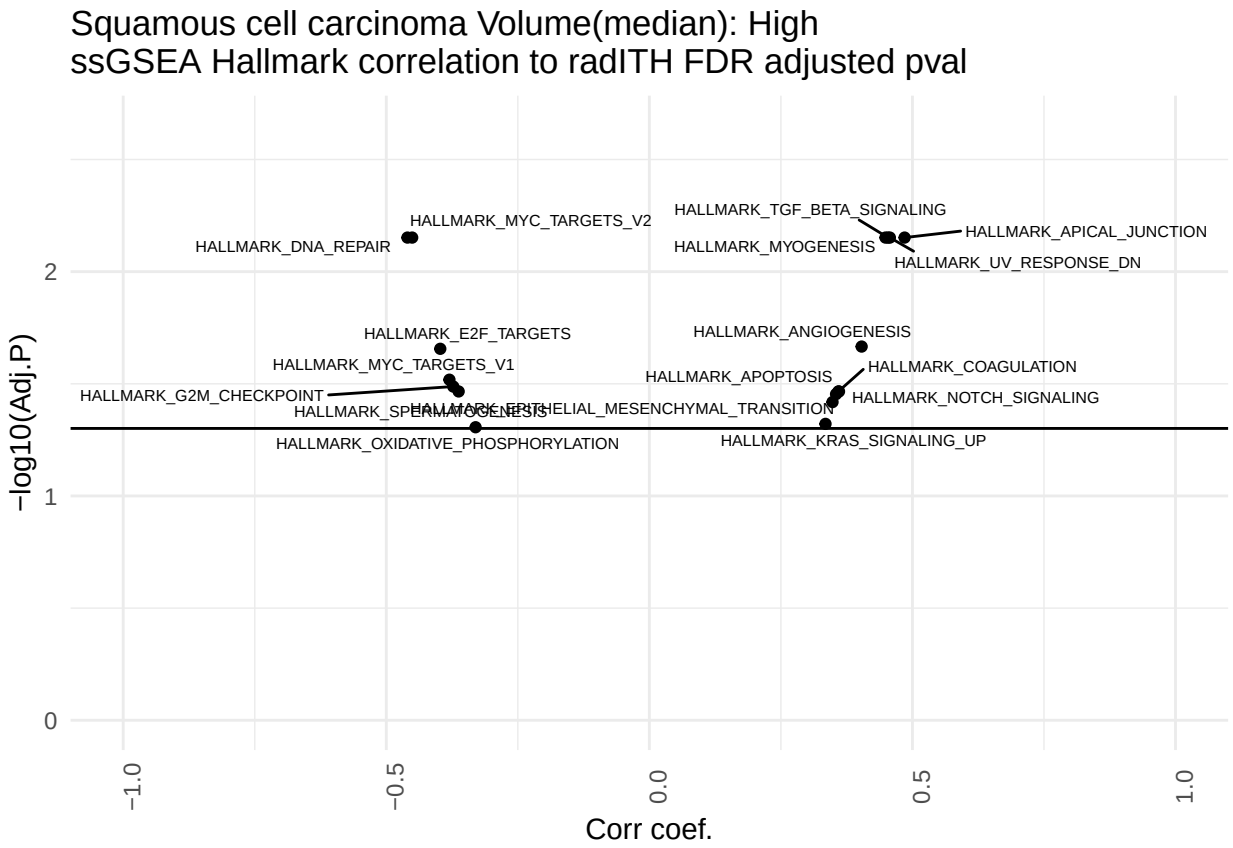


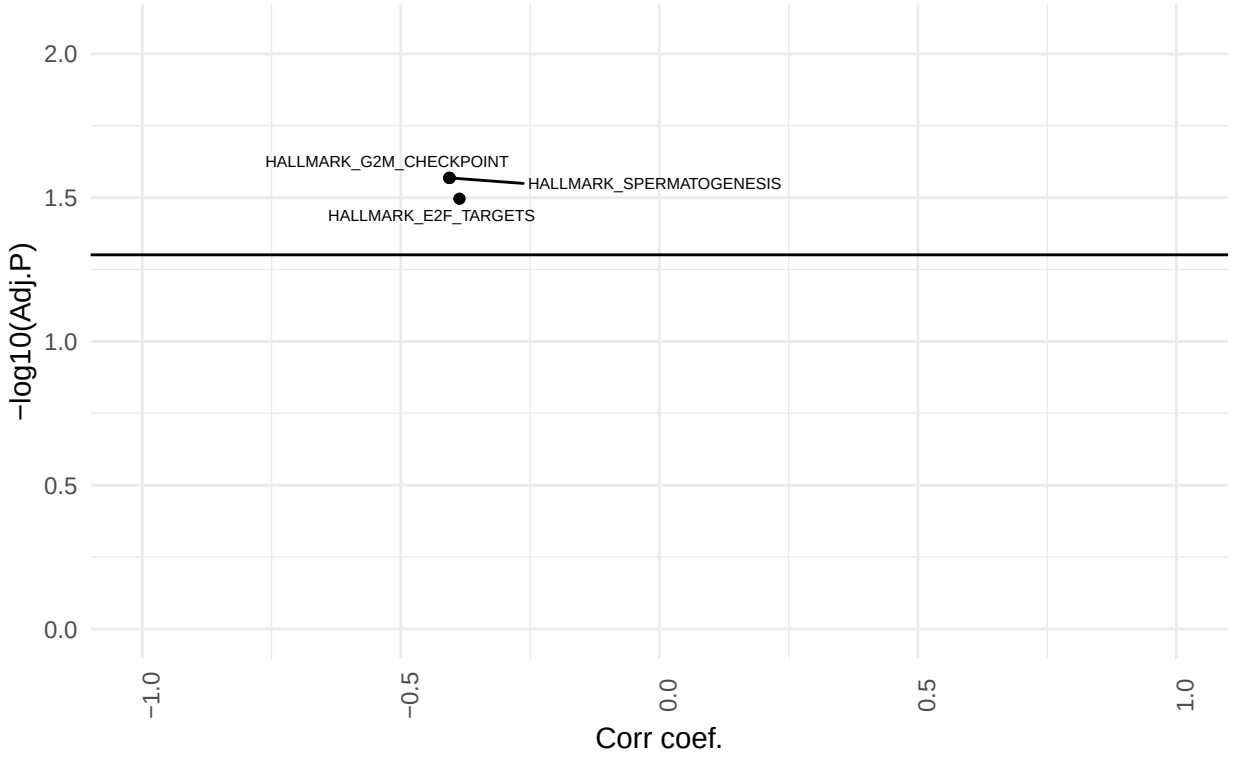
Hallmark expression-radITH Correlation in Large vs Small tumors (all samples)

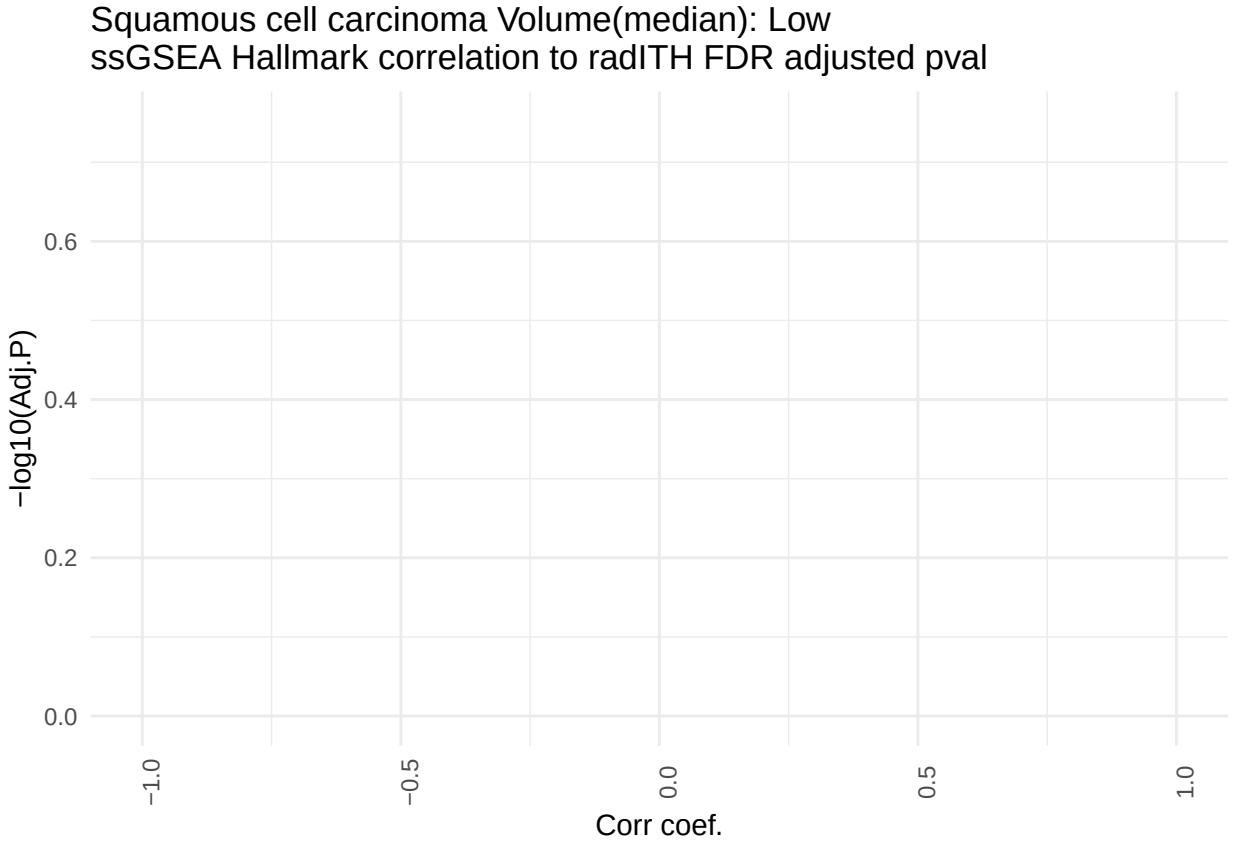


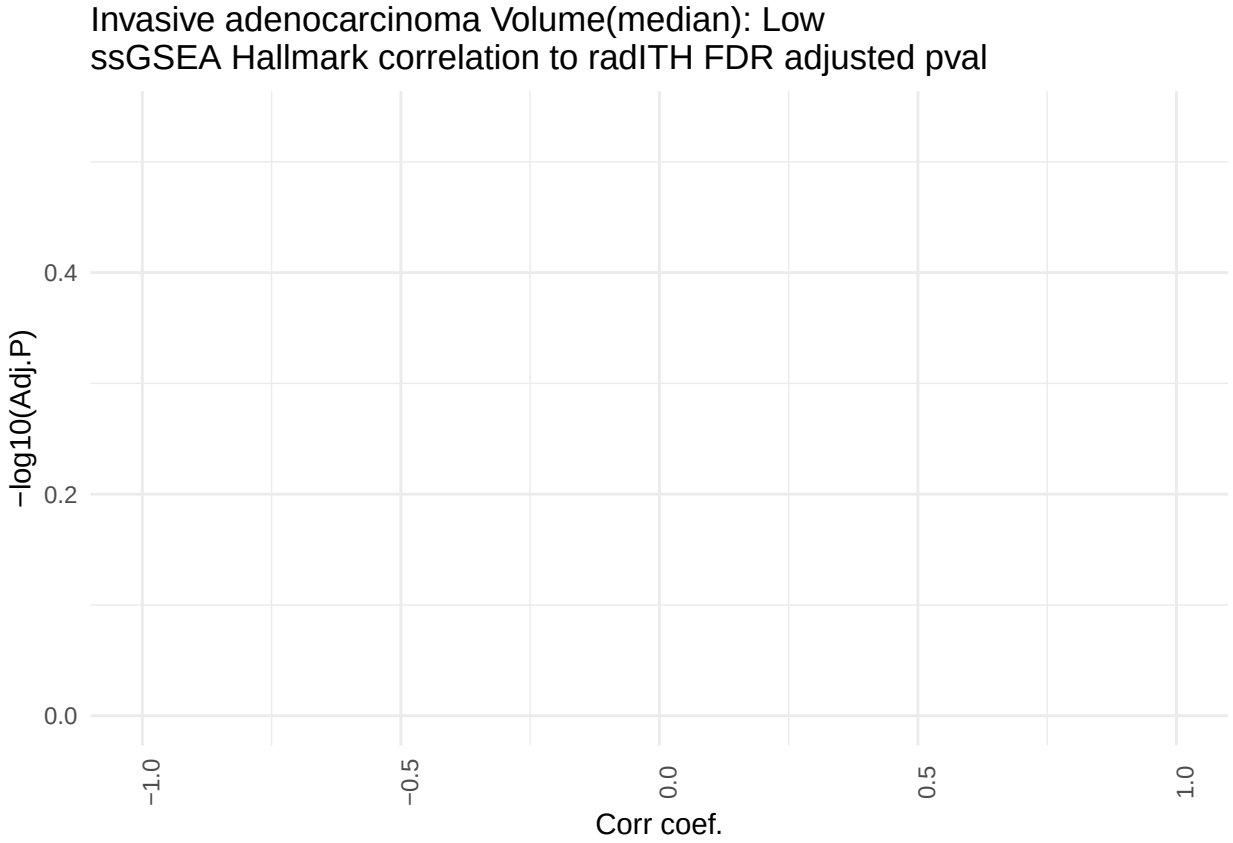


Let's split by Size and Pathology and repeat



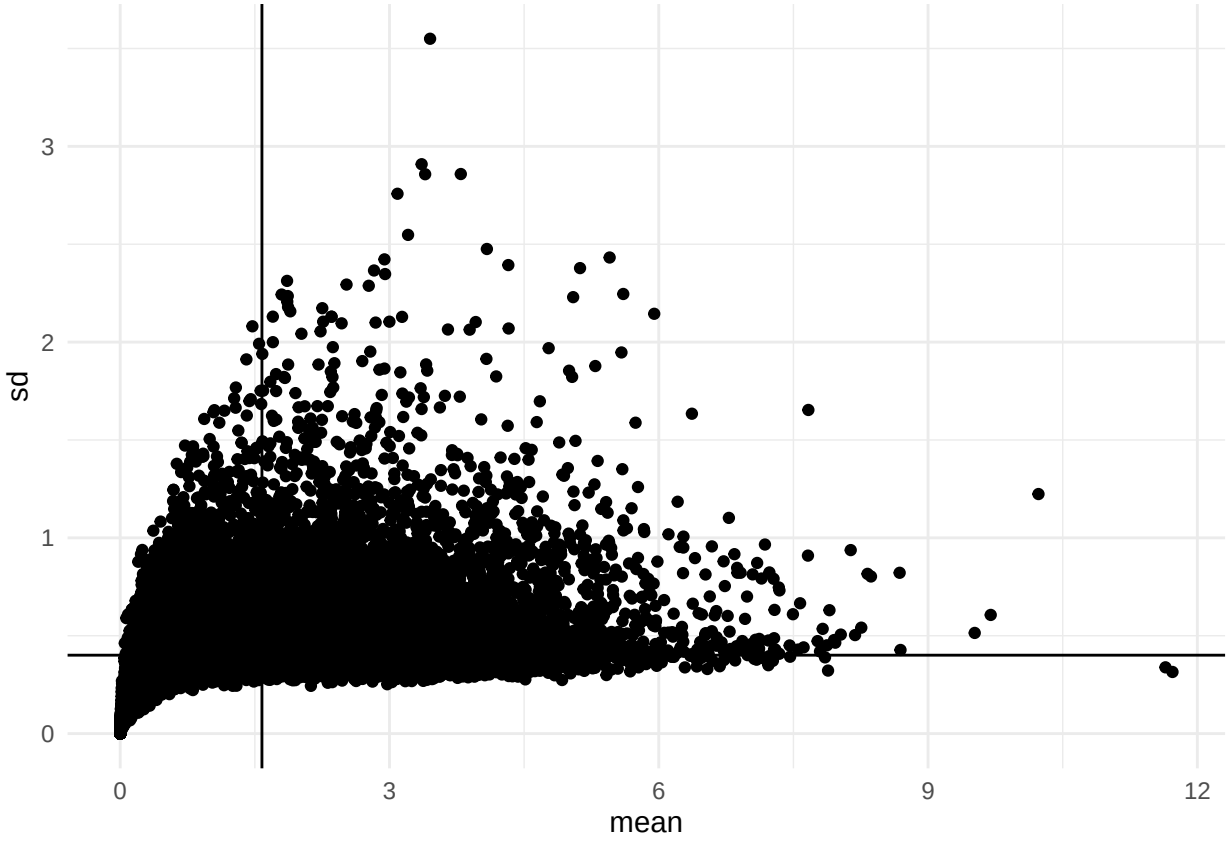



Squamous cell carcinoma Volume(median): Low
ssGSEA Hallmark correlation to radITH FDR adjusted pval
−log10(Adj.P)
0.6
0.4
0.2
0.0
−1.0
−0.5
0.0
0.5
1.0
Corr coef.


Invasive adenocarcinoma Volume(median): Low
ssGSEA Hallmark correlation to radITH FDR adjusted pval
−log10(Adj.P)
0.4
0.2
0.0
−1.0
−0.5
0.0
0.5
1.0
Corr coef.

Picking genes for Gene Expression Analysis

- Mean and SD value based on entire cohort

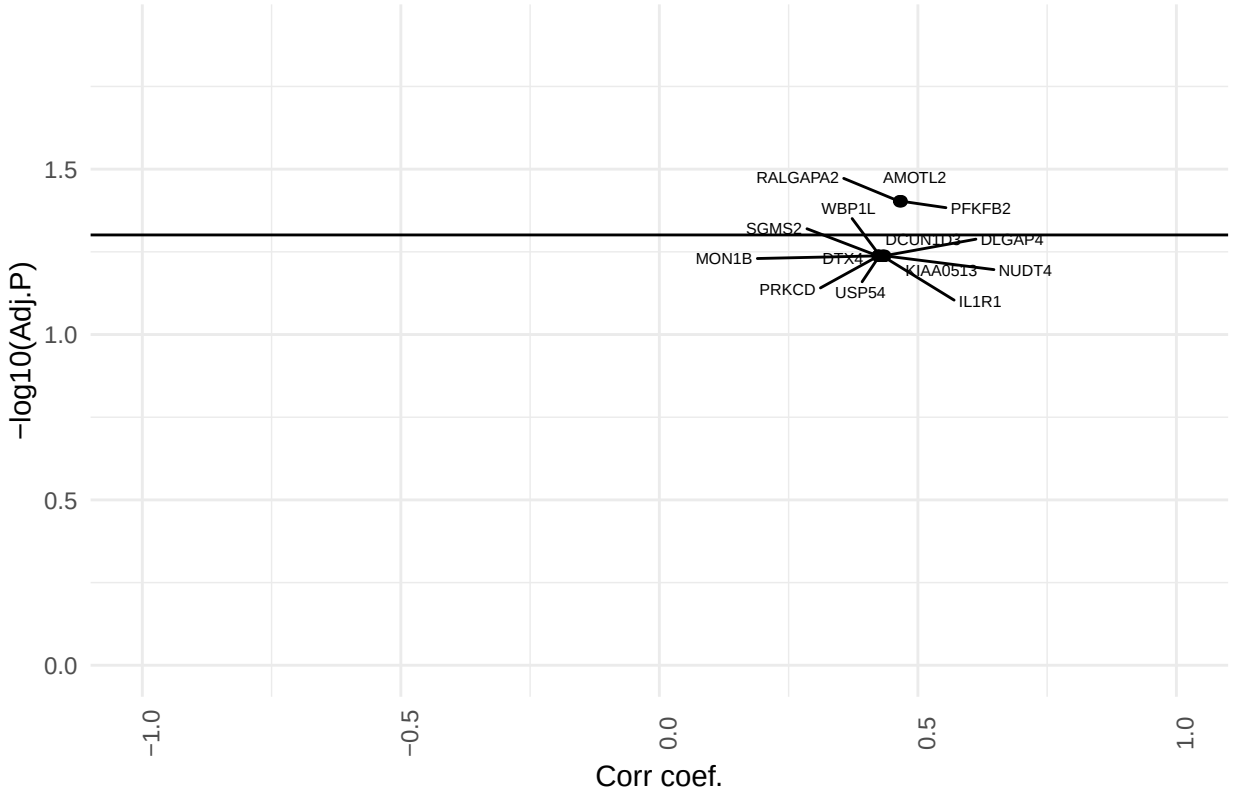



```
## Number of Genes after cutoff:  10332
```

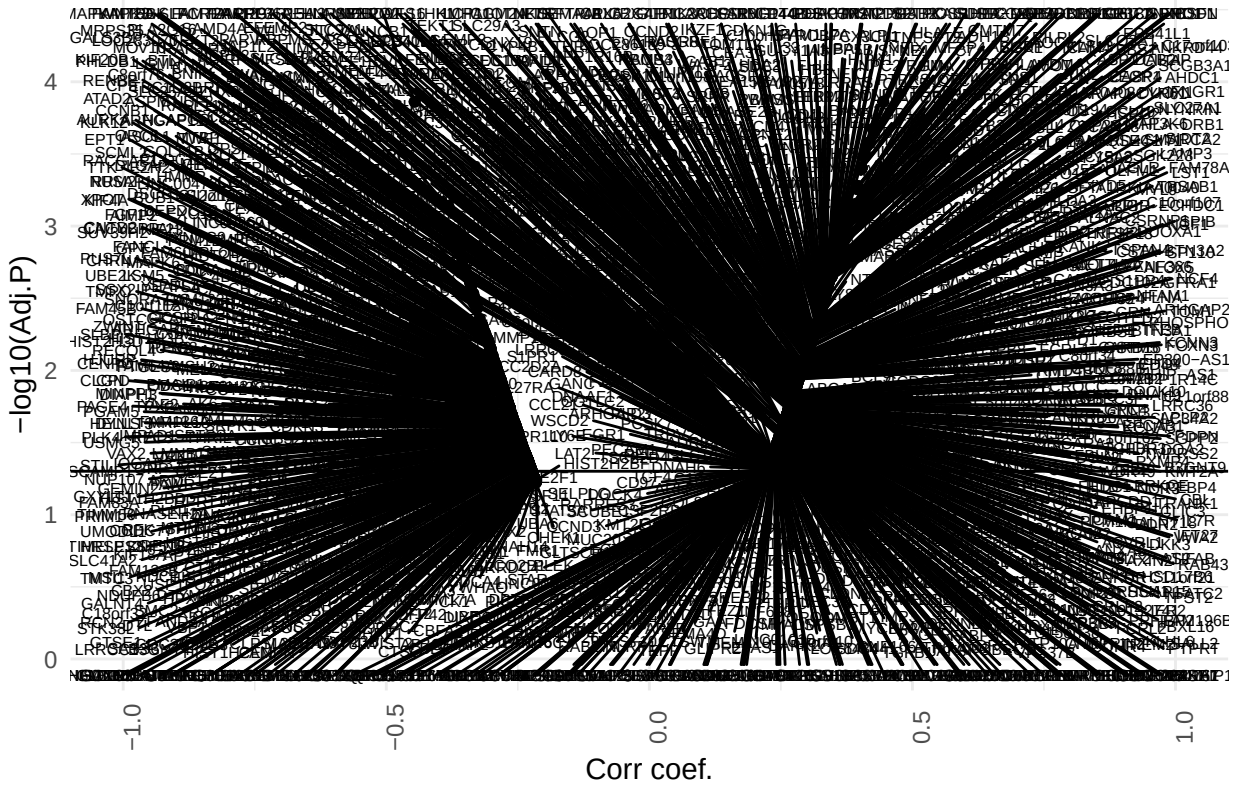
Gene Expression Analysis without volume

- Genes that were picked for analysis were based on mean and SD (entire cohort)

Squamous all samples

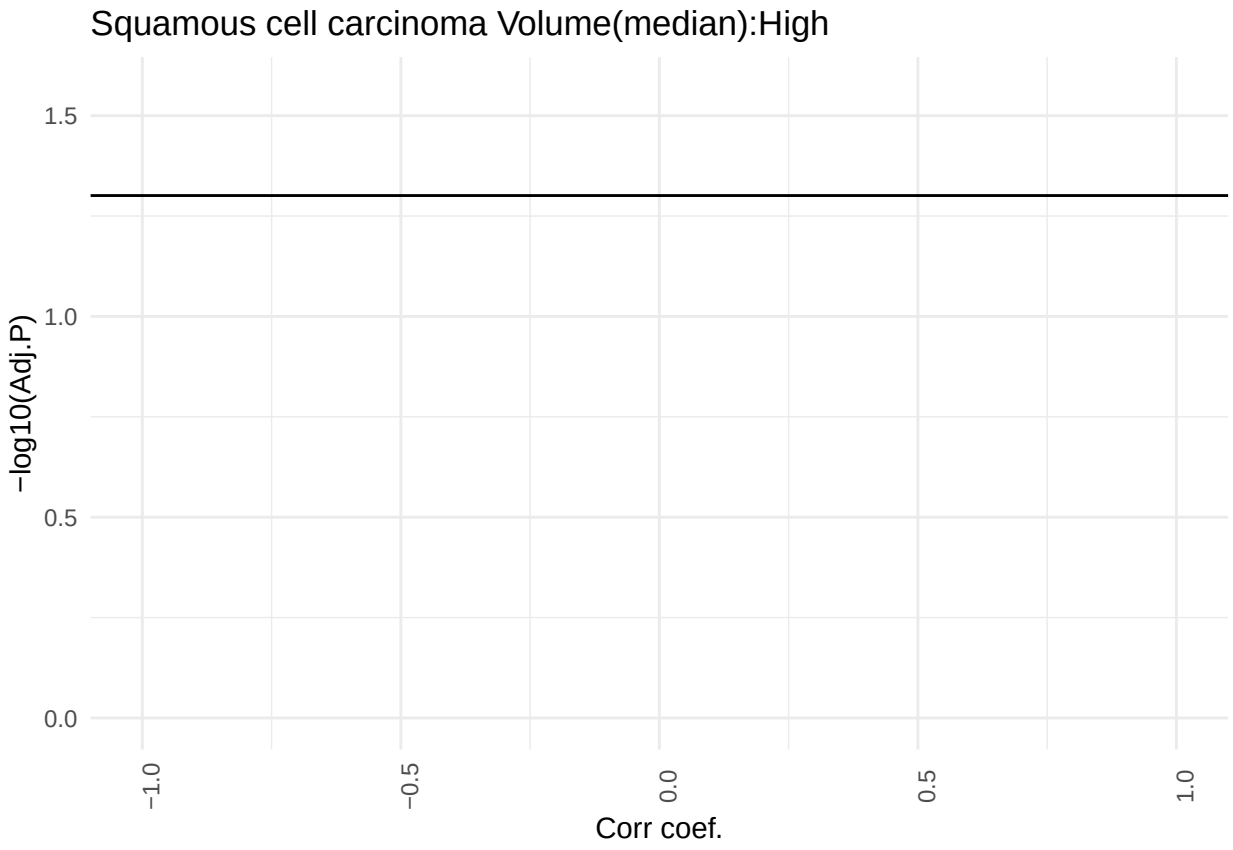


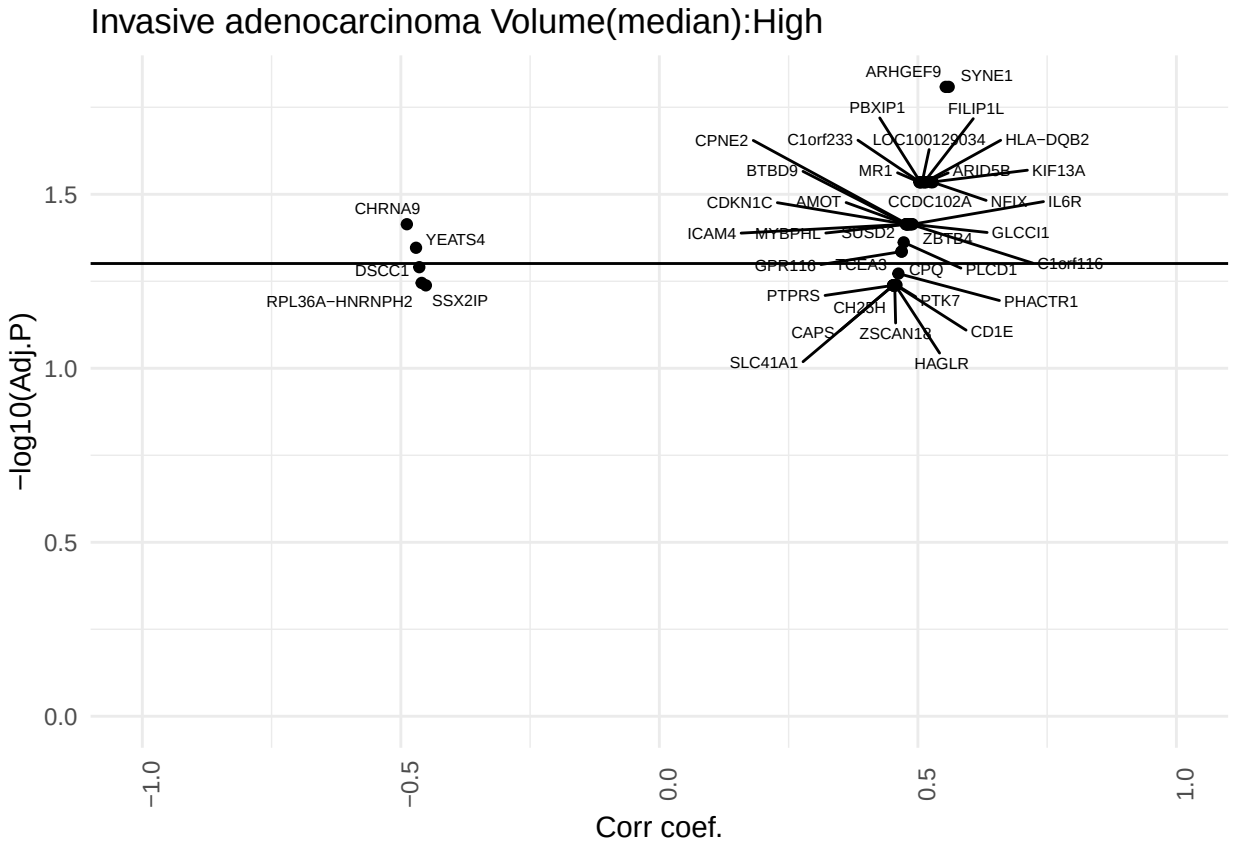
Adeno all samples

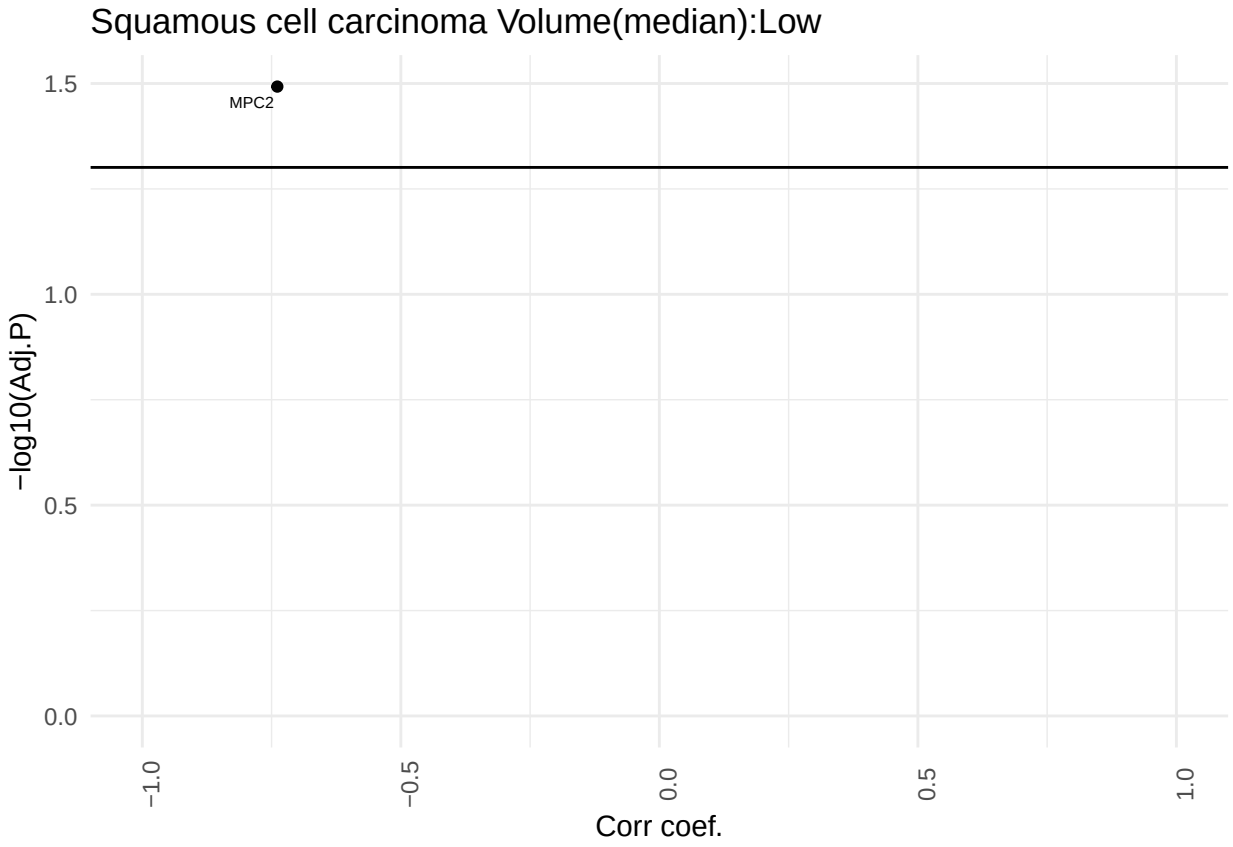


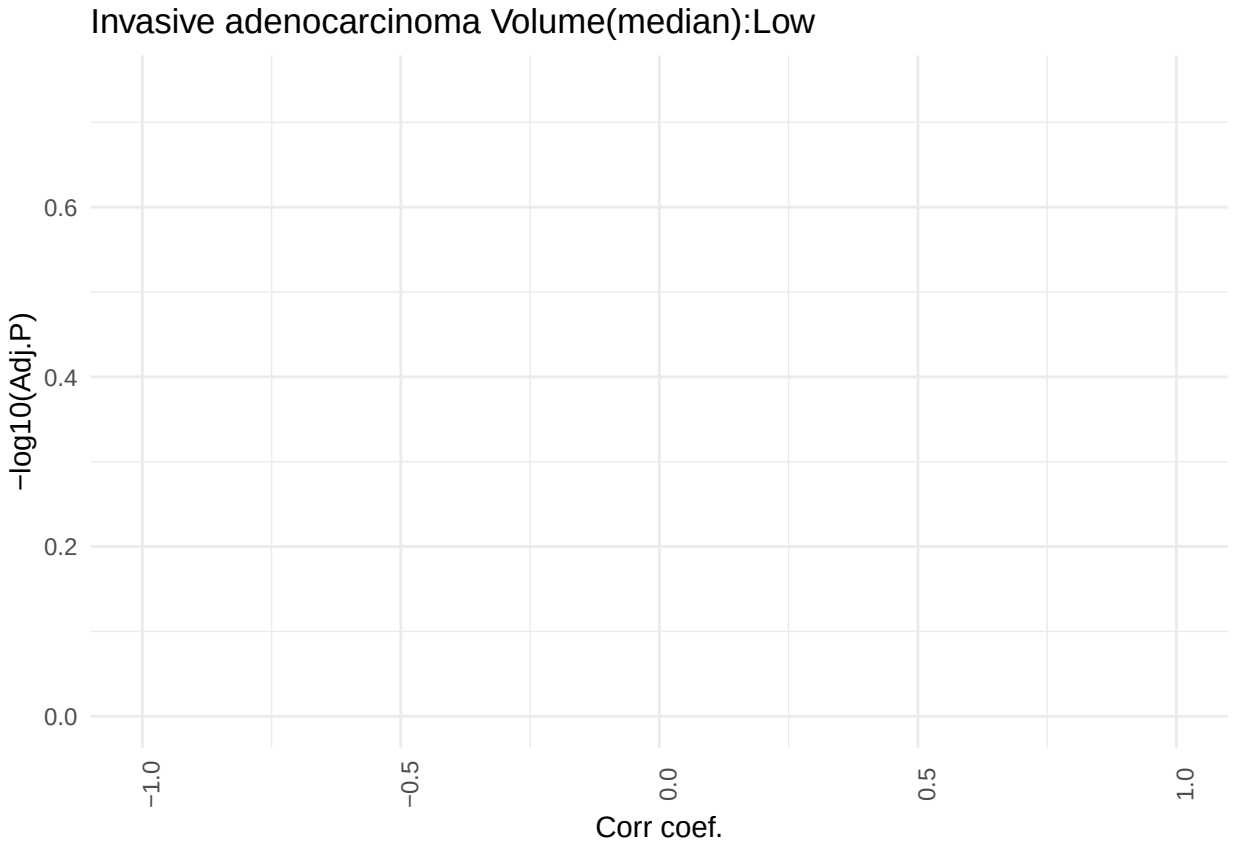
Gene Expression Analysis by volume group and cancer type (CORRELATION)

- Genes that were picked for analysis were based on mean and SD (entire cohort)



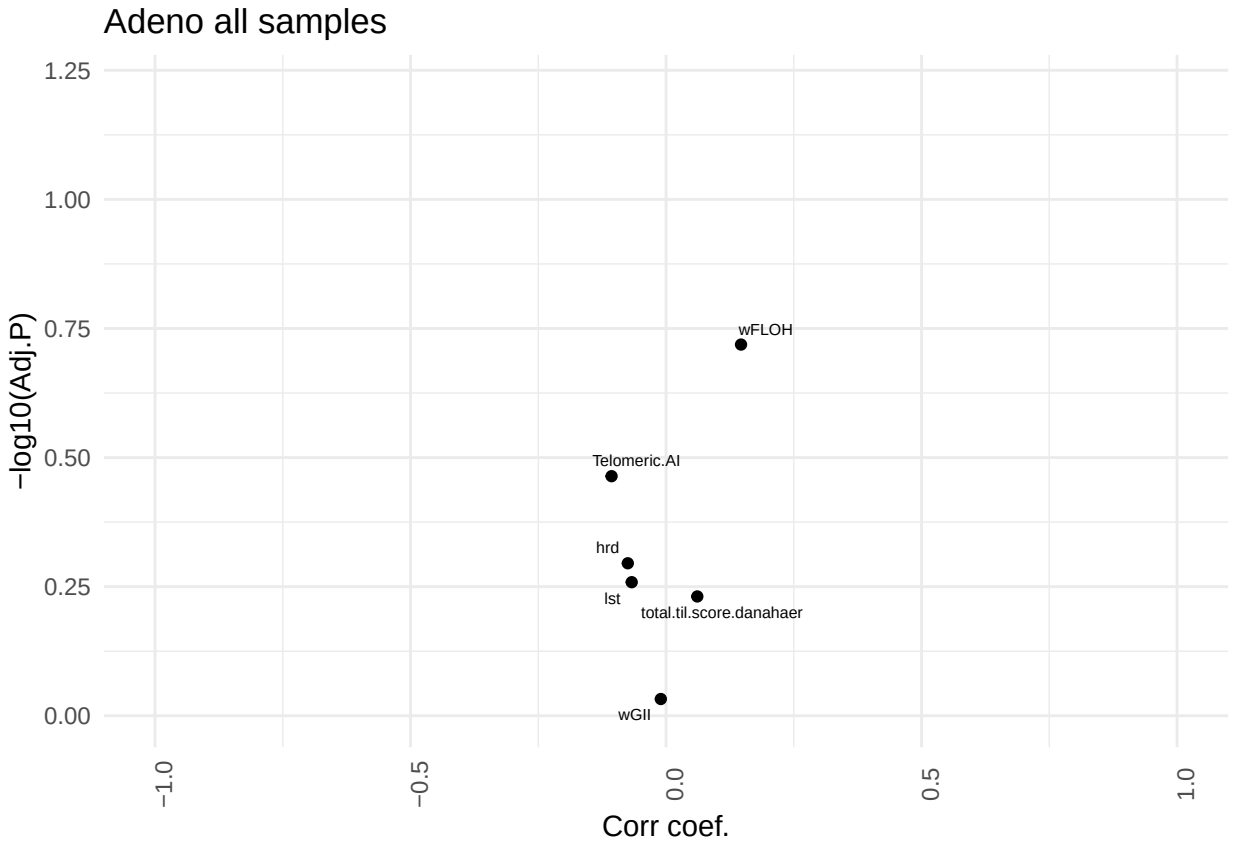

Invasive adenocarcinoma Volume(median):High
ARHGEF9
SYNE1
PBXIP1
FILIP1L
CPNE2
C1orf233
LOC100129034
HLA−DQB2
BTBD9
MR1
ARID5B
KIF13A
CHRNA9
CDKN1C
AMOT
CCDC102A
NFIX
IL6R
YEATS4
ICAM4
MYBPHL
SUSD2
GLCCI1
ZBTB4
DSCC1
GPR116
TCEA3
CPQ
PLCD1
C1orf116
RPL36A−HNRNPH2
SSX2IP
PTPRS
CH25H
PTK7
PHACTR1
CAPS
ZSCAN18
CD1E
SLC41A1
HAGLR
1.5
1.0
0.5
0.0
−log10(Adj.P)
−1.0
−0.5
0.0
0.5
1.0
Corr coef.

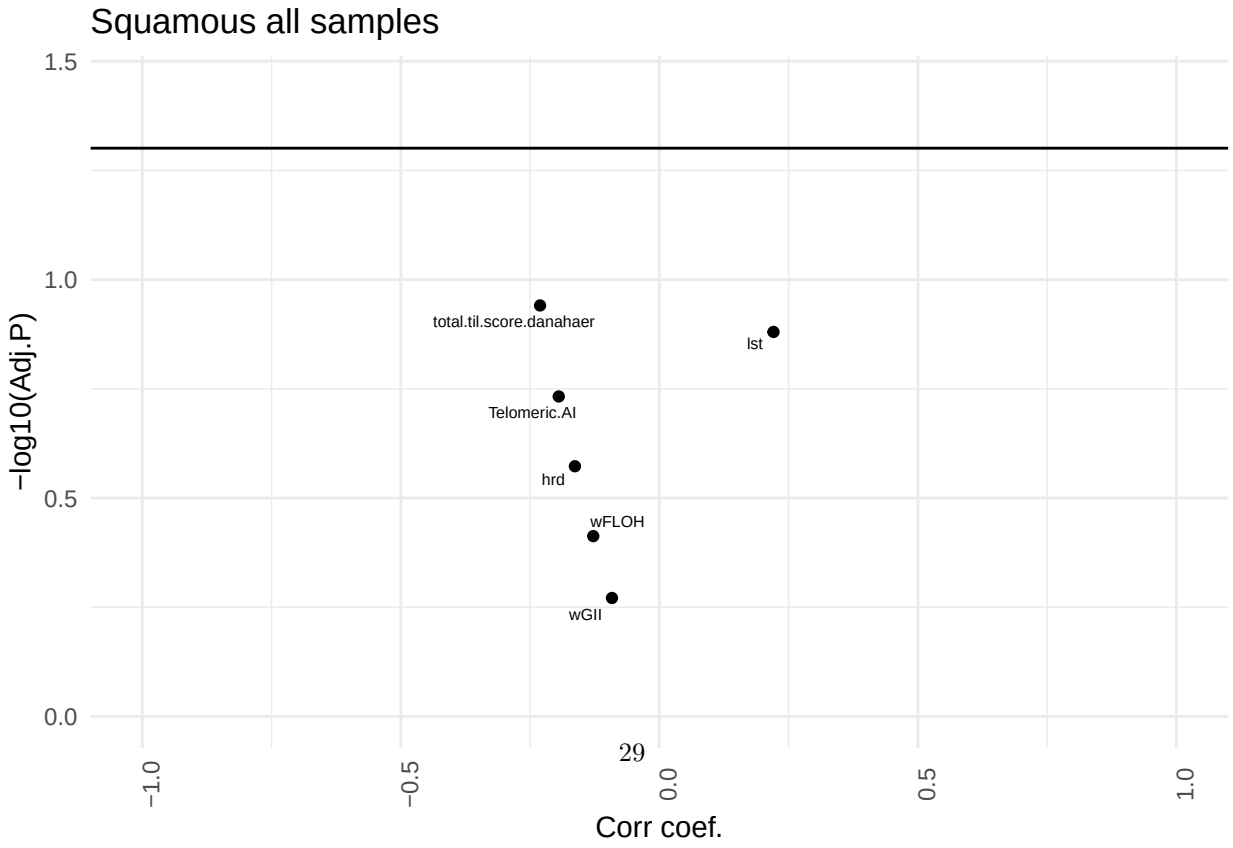

Squamous cell carcinoma Volume(median):Low
MPC2
1.5
1.0
0.5
0.0
−log10(Adj.P)
−1.0
−0.5
0.0
0.5
1.0
Corr coef.

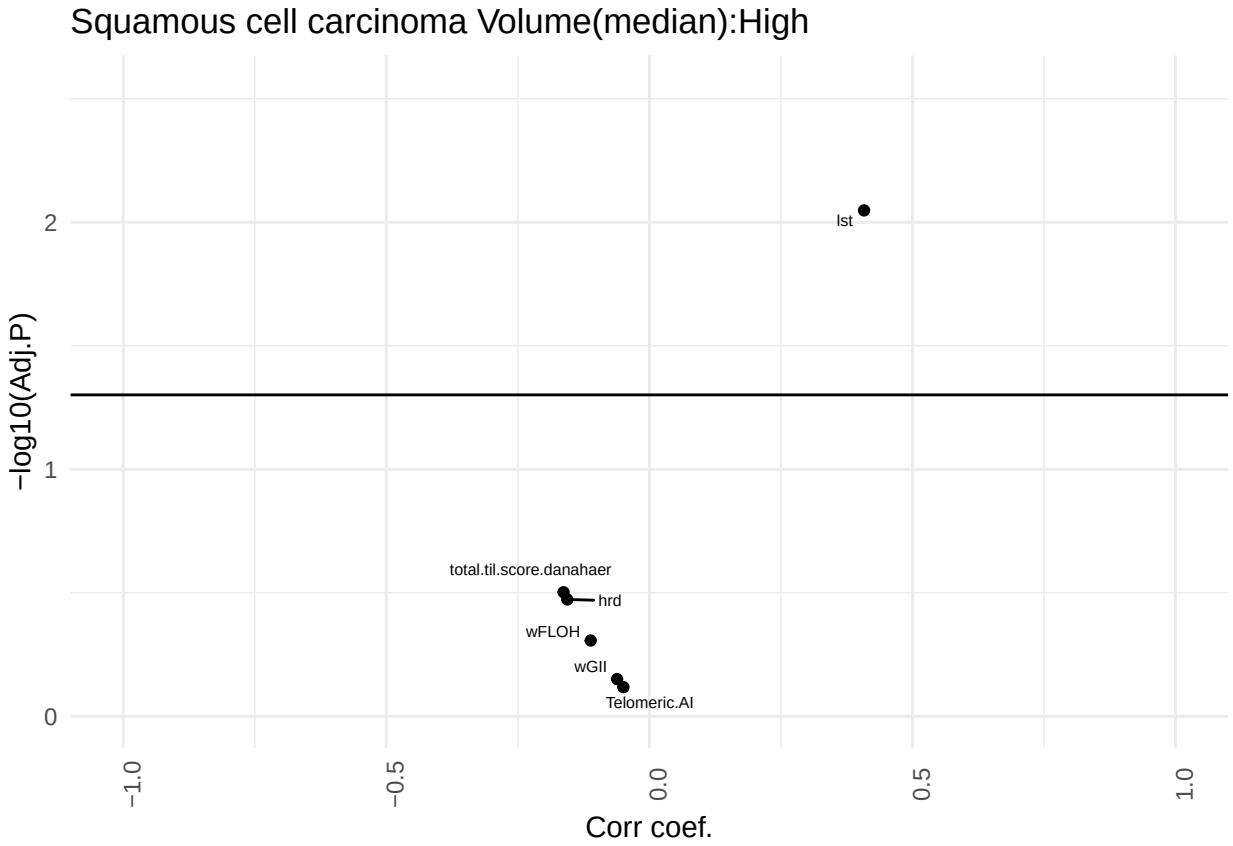

Invasive adenocarcinoma Volume(median):Low
0.6
0.4
0.2
0.0
−log10(Adj.P)
−1.0
−0.5
0.0
0.5
1.0
Corr coef.

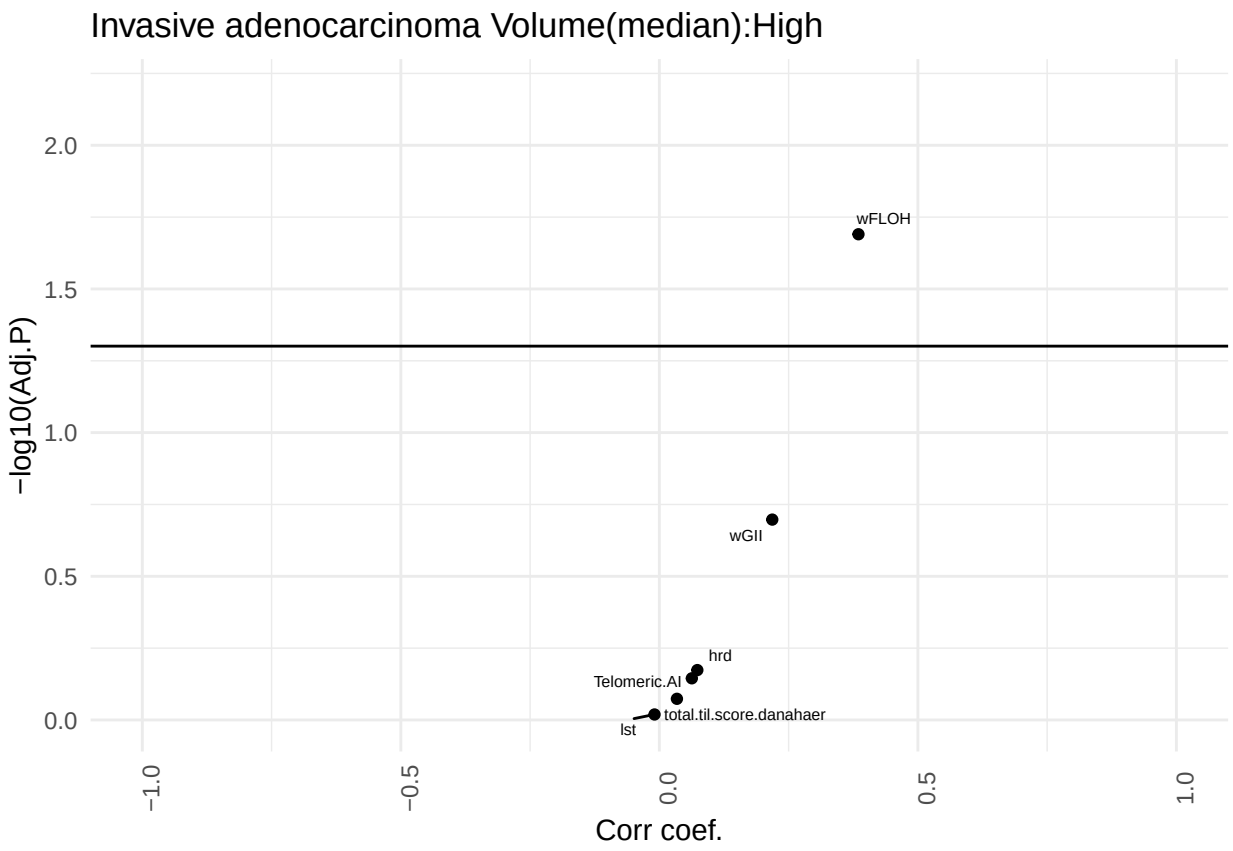
Gene Expression Analysis by volume group and cancer type and radITH group (T-TEST)

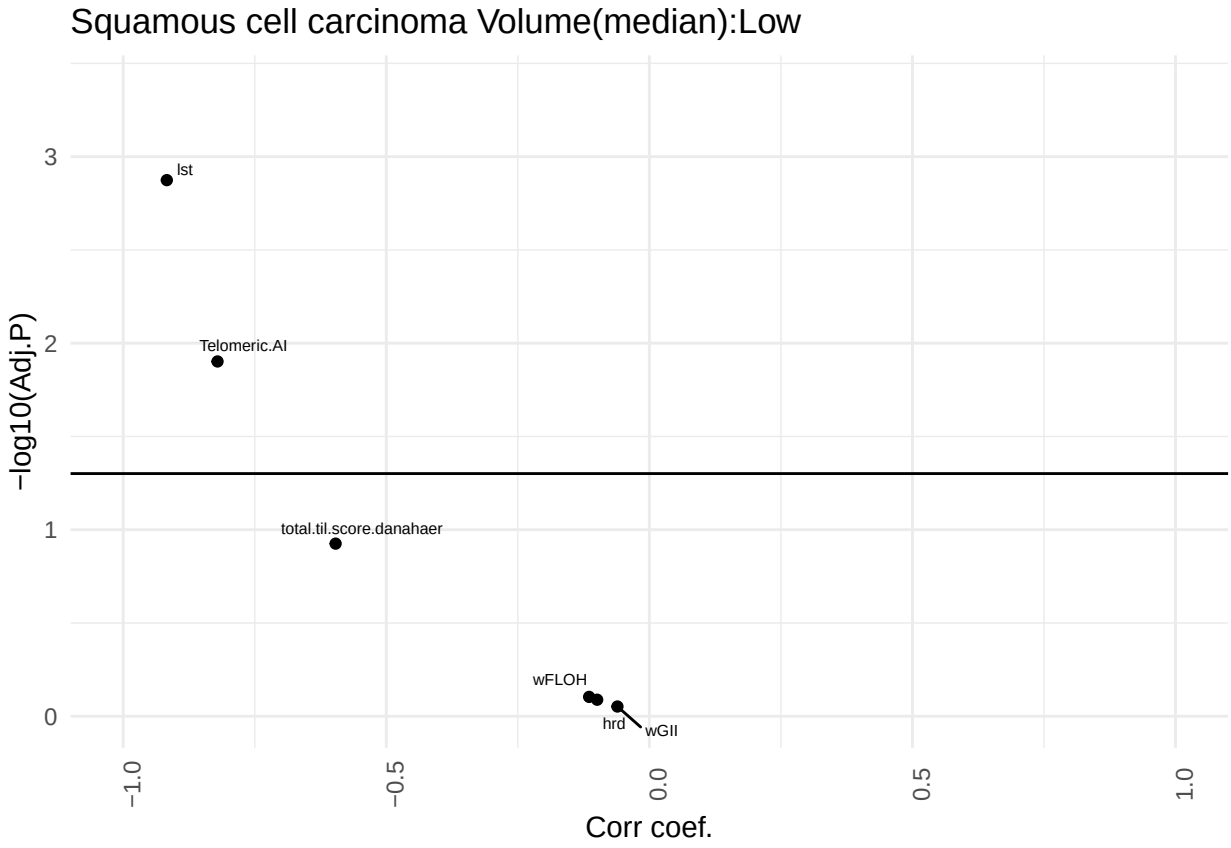
Chr Instability and TIL

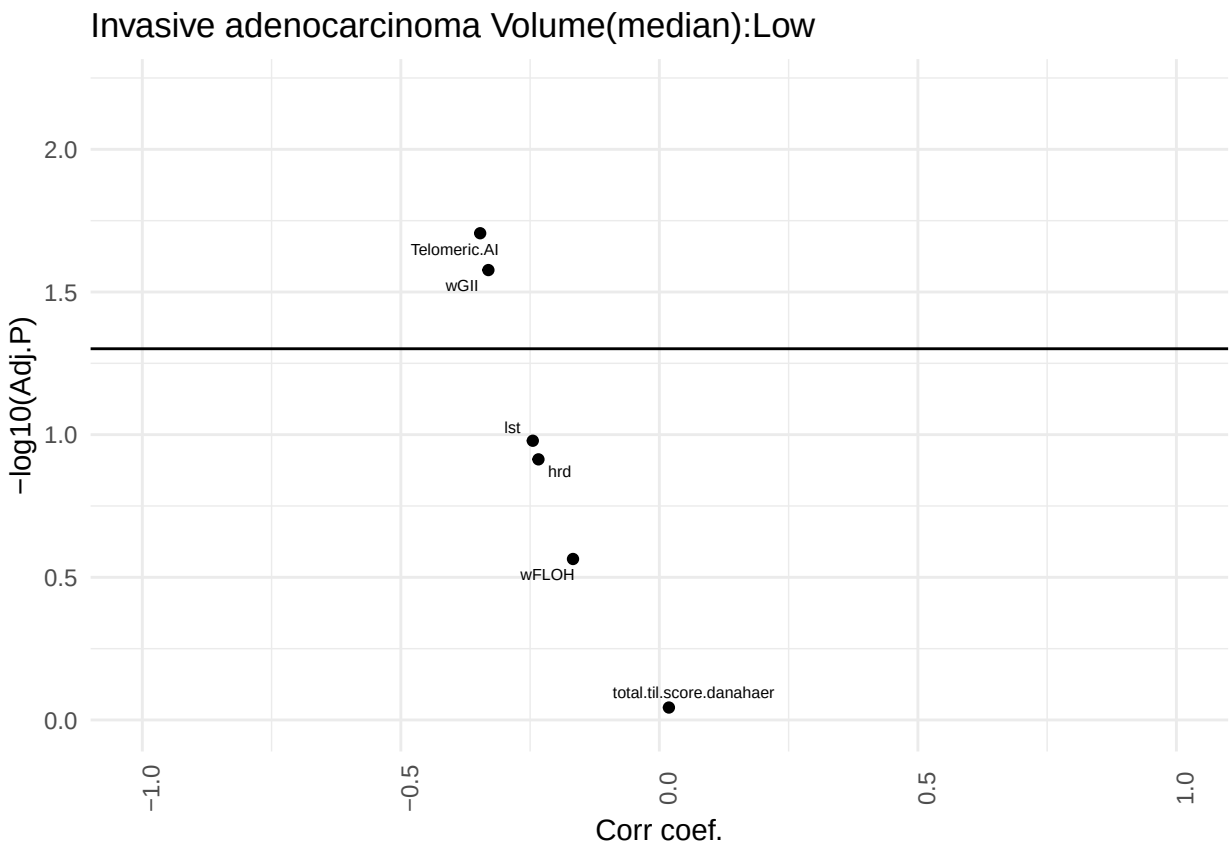





Squamous cell carcinoma Volume(median):High
−log10(Adj.P)
2
1
0
lst
total.til.score.danahaer
hrd
wFLOH
wGII
Telomeric.AI
−1.0
−0.5
0.0
0.5
1.0
Corr coef.


Invasive adenocarcinoma Volume(median):High
−log10(Adj.P)
2.0
1.5
1.0
0.5
0.0
wFLOH
wGII
hrd
Telomeric.AI
lst
total.til.score.danahaer
−1.0
−0.5
0.0
0.5
1.0
Corr coef.

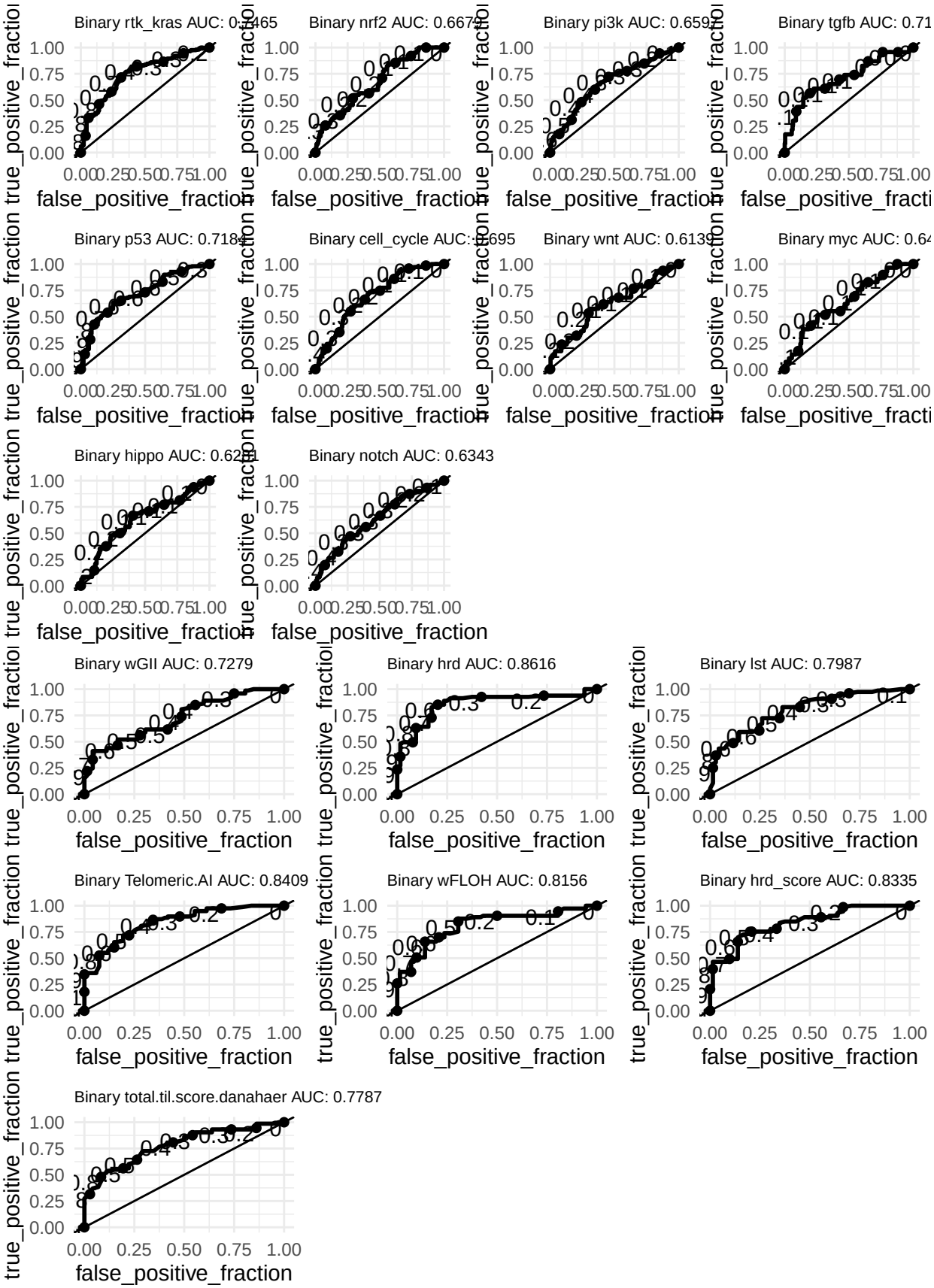

Squamous cell carcinoma Volume(median):Low
lst
Telomeric.AI
total.til.score.danahaer
wFLOH
hrd
wGII
−log10(Adj.P)
Corr coef.


Invasive adenocarcinoma Volume(median):Low
Telomeric.AI
wGII
lst
hrd
wFLOH
total.til.score.danahaer
−log10(Adj.P)
Corr coef.

Can we use pyrad features to predict biological features?

- As it does not make sense to split small amount of observations by pathology I will include pathology as parameter in the model
- IMPORTANT! Params in model: chosen pyRad features, pathology, radITH and volume
- The results represent Logistic regression with 3-fold (5 times repeated) CV using R-Caret package

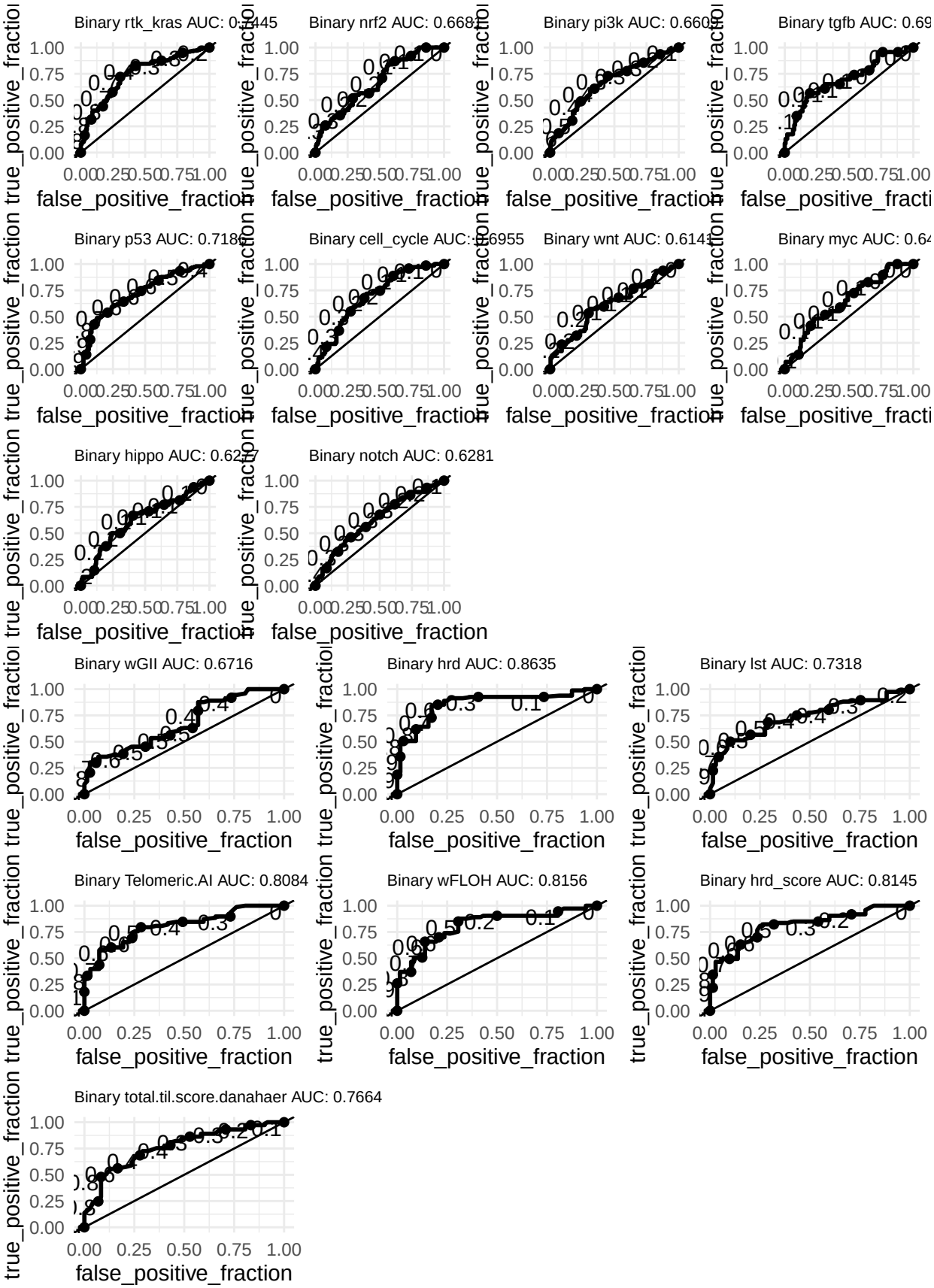
```
## [1] "rtk_kras"
## [1] "nrf2"
## [1] "pi3k"
## [1] "tgfb"
## [1] "p53"
## [1] "cell_cycle"
## [1] "wnt"
## [1] "myc"
## [1] "hippo"
## [1] "notch"
## [1] "wGII"
## [1] "hrd"
## [1] "lst"
## [1] "Telomeric.AI"
## [1] "wFLOH"
## [1] "hrd_score"
## [1] "total.til.score.danahaer"
```


Binary rtk_kras AUC: 0.7465
Binary nrf2 AUC: 0.6675
Binary pi3k AUC: 0.6597
Binary tgfb AUC: 0.71
Binary p53 AUC: 0.7184
Binary cell_cycle AUC: 0.7695
Binary wnt AUC: 0.6139
Binary myc AUC: 0.64
Binary hippo AUC: 0.6281
Binary notch AUC: 0.6343
Binary wGII AUC: 0.7279
Binary hrd AUC: 0.8616
Binary lst AUC: 0.7987
Binary Telomeric.AI AUC: 0.8409
Binary wFLOH AUC: 0.8156
Binary hrd_score AUC: 0.8335
Binary total.til.score.danahaer AUC: 0.7787
true_positive_fraction
false_positive_fraction
0.00
0.25
0.50
0.75
1.00

Can we use radITH and pyRad Features for the same results?

- As it does not make sense to split small amount of observations by pathology I will include pathology as parameter in the model
- IMPORTANT! Params in model: chosen pyRad features, pathology, radITH
- The results represent Logistic regression with 3-fold (5 time repeated) CV using R-Caret package

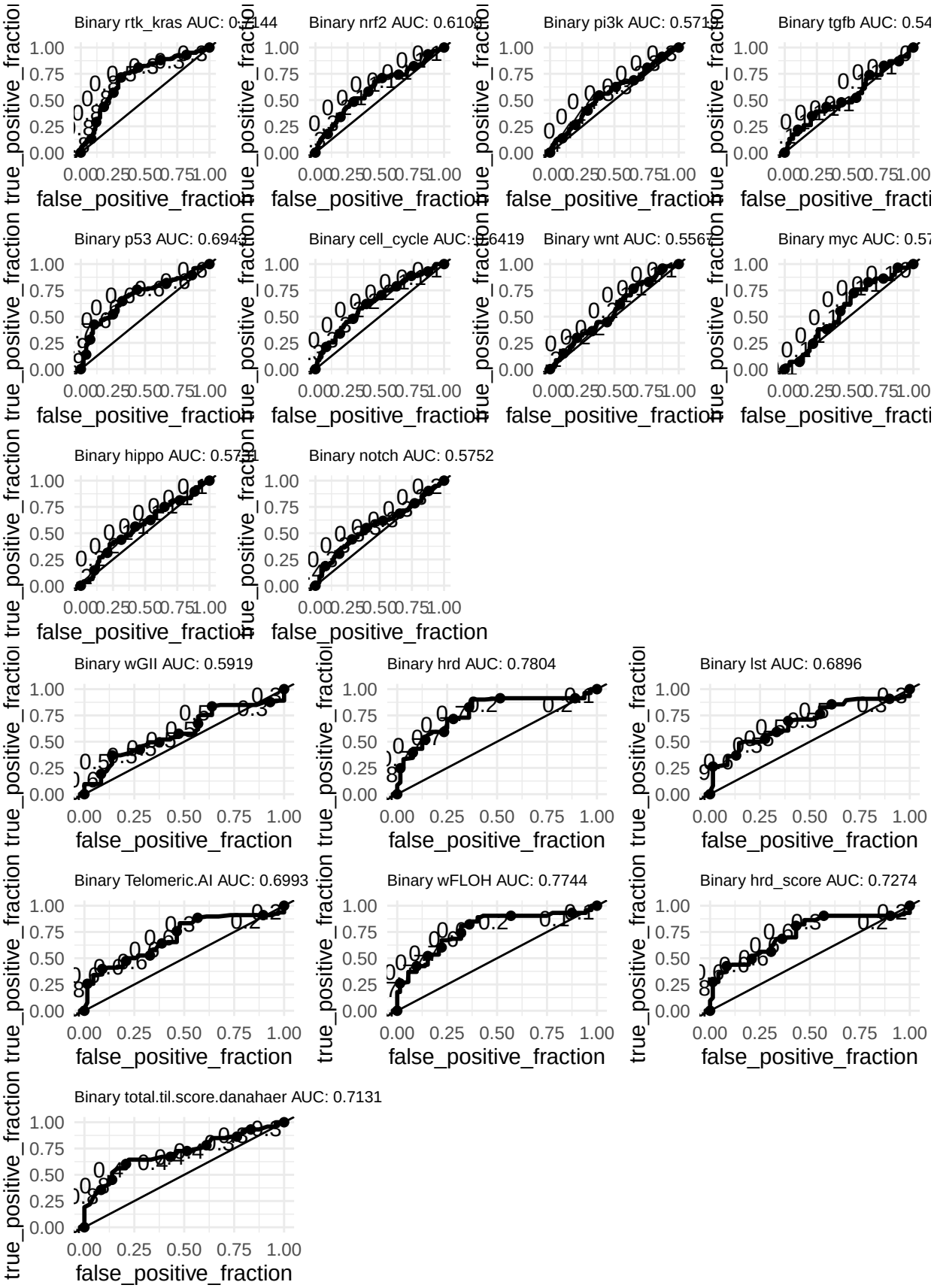
```
## [1] "rtk_kras"
## [1] "nrf2"
## [1] "pi3k"
## [1] "tgfb"
## [1] "p53"
## [1] "cell_cycle"
## [1] "wnt"
## [1] "myc"
## [1] "hippo"
## [1] "notch"
## [1] "wGII"
## [1] "hrd"
## [1] "lst"
## [1] "Telomeric.AI"
## [1] "wFLOH"
## [1] "hrd_score"
## [1] "total.til.score.danahaer"
```


Binary rtk_kras AUC: 0.7445
Binary nrf2 AUC: 0.6683
Binary pi3k AUC: 0.6605
Binary tgfb AUC: 0.69
Binary p53 AUC: 0.7186
Binary cell_cycle AUC: 0.6955
Binary wnt AUC: 0.6143
Binary myc AUC: 0.64
Binary hippo AUC: 0.627
Binary notch AUC: 0.6281
Binary wGII AUC: 0.6716
Binary hrd AUC: 0.8635
Binary lst AUC: 0.7318
Binary Telomeric.AI AUC: 0.8084
Binary wFLOH AUC: 0.8156
Binary hrd_score AUC: 0.8145
Binary total.til.score.danahaer AUC: 0.7664
true_positive_fraction
false_positive_fraction
0.00
0.25
0.50
0.75
1.00

Can we use only volume and Patholgy for the same results?

- As it does not make sense to split small amount of observations by pathology I will include pathology as parameter in the model
- IMPORTANT! Params in model: pathology, volume
- The results represent Logistic regression with 3-fold (5 time repeated) CV using R-Caret package

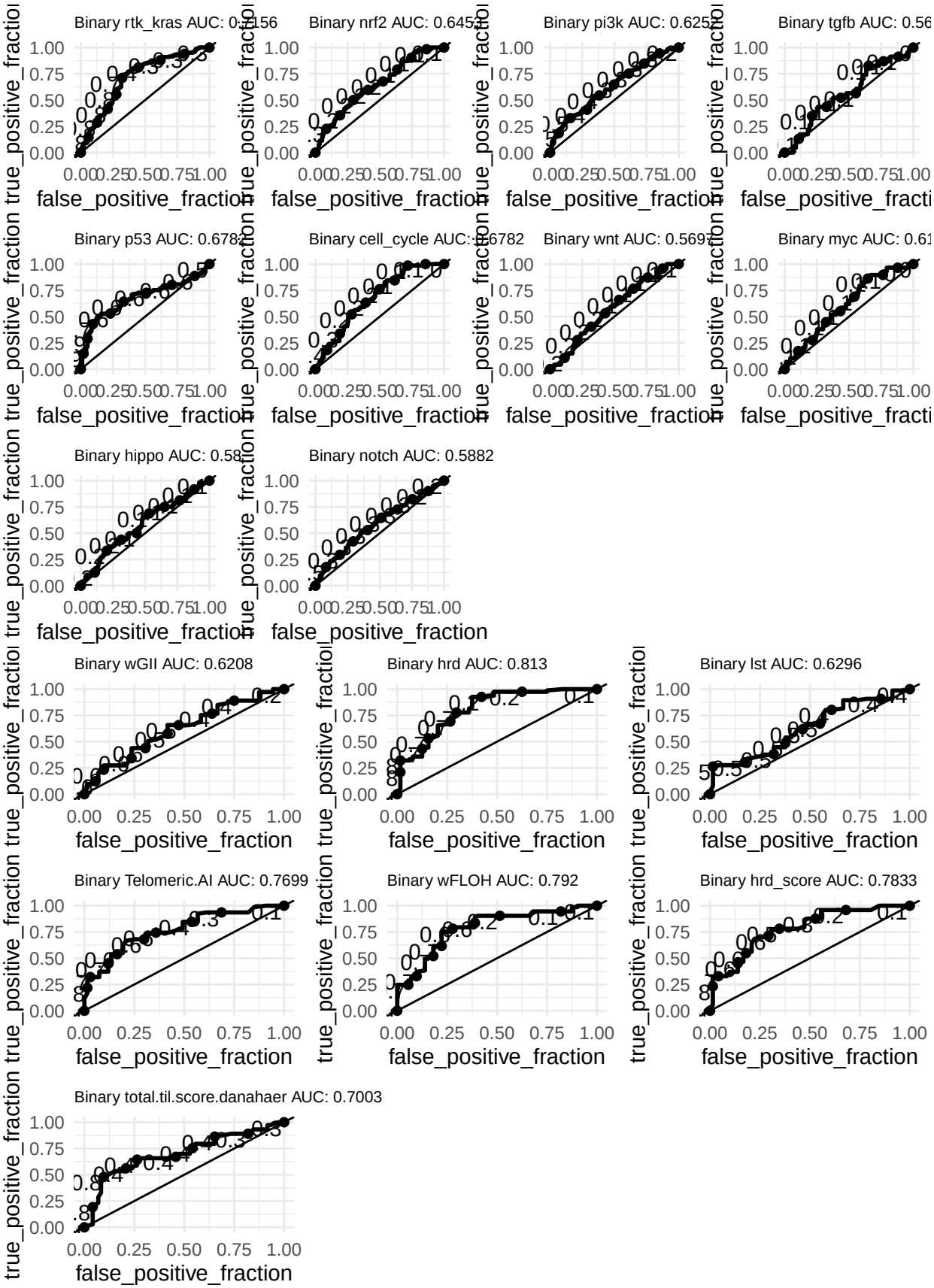
```
## [1] "rtk_kras"
## [1] "nrf2"
## [1] "pi3k"
## [1] "tgfb"
## [1] "p53"
## [1] "cell_cycle"
## [1] "wnt"
## [1] "myc"
## [1] "hippo"
## [1] "notch"
## [1] "wGII"
## [1] "hrd"
## [1] "lst"
## [1] "Telomeric.AI"
## [1] "wFLOH"
## [1] "hrd_score"
## [1] "total.til.score.danahaer"
```


Binary rtk_kras AUC: 0.5144
Binary nrf2 AUC: 0.6108
Binary pi3k AUC: 0.5715
Binary tgfb AUC: 0.54
Binary p53 AUC: 0.6943
Binary cell_cycle AUC: 0.6419
Binary wnt AUC: 0.5567
Binary myc AUC: 0.57
Binary hippo AUC: 0.5731
Binary notch AUC: 0.5752
Binary wGII AUC: 0.5919
Binary hrd AUC: 0.7804
Binary lst AUC: 0.6896
Binary Telomeric.AI AUC: 0.6993
Binary wFLOH AUC: 0.7744
Binary hrd_score AUC: 0.7274
Binary total.til.score.danahaer AUC: 0.7131
true_positive_fraction
false_positive_fraction
0.00
0.25
0.50
0.75
1.00

Can we use only radITH and Patholgy for the same results?

- As it does not make sense to split small amount of observations by pathology I will include pathology as parameter in the model
- IMPORTANT! Params in model: pathology, radITH
- The results represent Logistic regression with 3-fold (5 time repeated) CV using R-Caret package

```
## [1] "rtk_kras"
## [1] "nrf2"
## [1] "pi3k"
## [1] "tgfb"
## [1] "p53"
## [1] "cell_cycle"
## [1] "wnt"
## [1] "myc"
## [1] "hippo"
## [1] "notch"
## [1] "wGII"
## [1] "hrd"
## [1] "lst"
## [1] "Telomeric.AI"
## [1] "wFLOH"
## [1] "hrd_score"
## [1] "total.til.score.danahaer"
```


Binary rtk_kras AUC: 0.6156
Binary nrf2 AUC: 0.6455
Binary pi3k AUC: 0.6255
Binary tgfb AUC: 0.56
Binary p53 AUC: 0.6782
Binary cell_cycle AUC: 0.6782
Binary wnt AUC: 0.5697
Binary myc AUC: 0.61
Binary hippo AUC: 0.58
Binary notch AUC: 0.5882
Binary wGII AUC: 0.6208
Binary hrd AUC: 0.813
Binary lst AUC: 0.6296
Binary Telomeric.AI AUC: 0.7699
Binary wFLOH AUC: 0.792
Binary hrd_score AUC: 0.7833
Binary total.til.score.danahaer AUC: 0.7003
true_positive_fraction
false_positive_fraction
0.00
0.25
0.50
0.75
1.00