Identifying and characterizing clinically relevant lncRNA candidates

Summary

Karin Isaev

February 12th, 2018

Workflow

Data	1200 lncRNAs (FANTOM labelled as intergenic) Four cancer types (TCGA & PCAWG): OV, LIHC, PAAD, KIRC

$\downarrow$

Filtering	Identify lncRNAs whose expression is associated with survival (LASSO for feature selection + 1000 Cross Validations) X10 *Cox Proportional-Hazards Model with stage and grade included as cofactors*

$\downarrow$
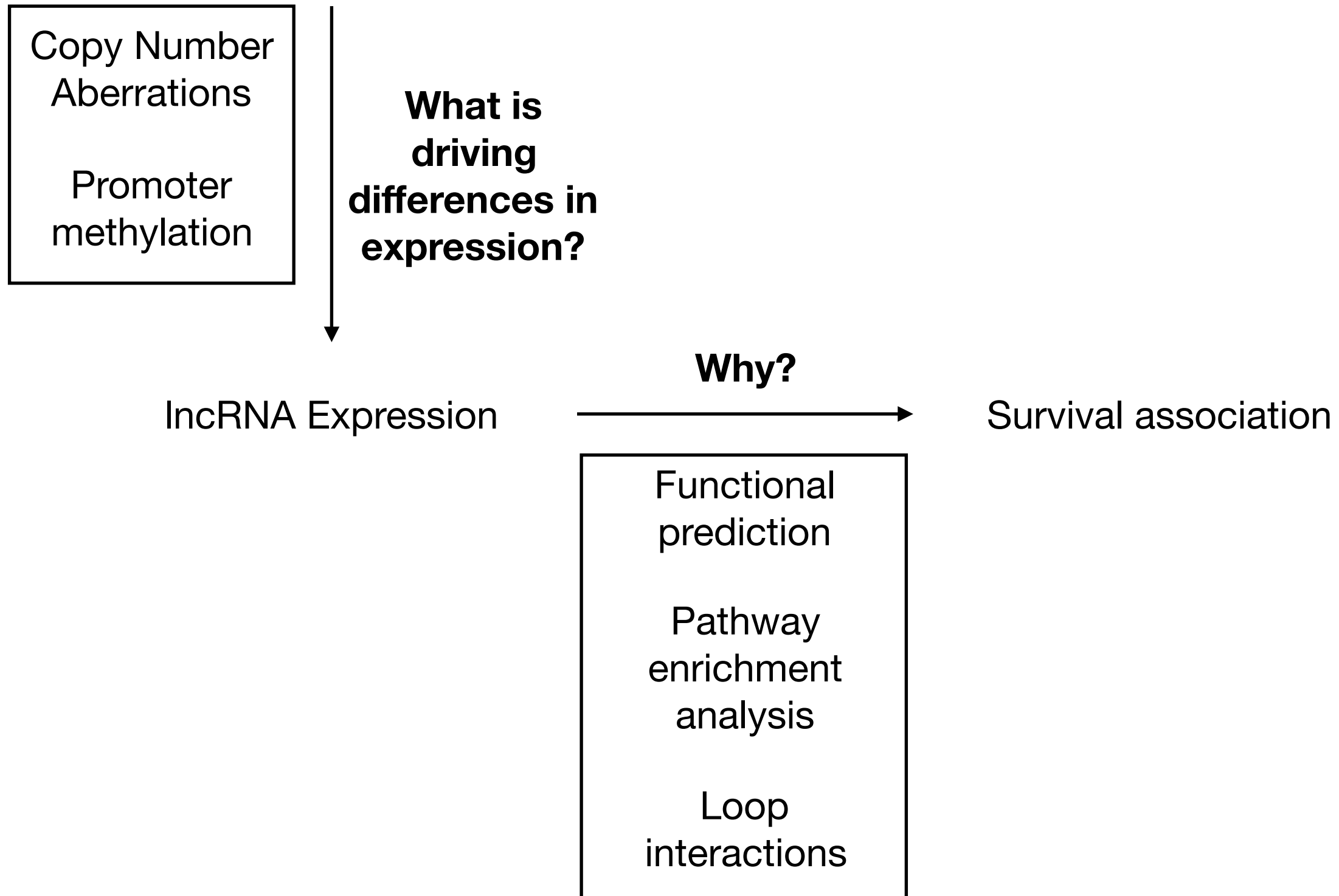
External Validation	Use PCAWG as validation dataset to confirm lncRNA-survival association

$\downarrow$

Functional Prediction	– Identify aberrations (SCNA, methylation) associated with expression – Associate lncRNA with PCGS, pathways, miRNAs and predict localization/function

Workflow

Copy Number
Aberrations

Promoter
methylation

**What is
driving
differences in
expression?**

lncRNA Expression **Why?** → Survival association

Functional
prediction

Pathway
enrichment
analysis

Loop
interactions

Results: Overview of Candidates

1. Significant p-value < 0.05 in Cox PH model even when clinical covariates included
2. Within each batch of cross validation, was selected by LASSO at least 400 times (2/5)
3. Met filter #1 in at least 5/10 batches

n = 38 Candidates (TCGA)
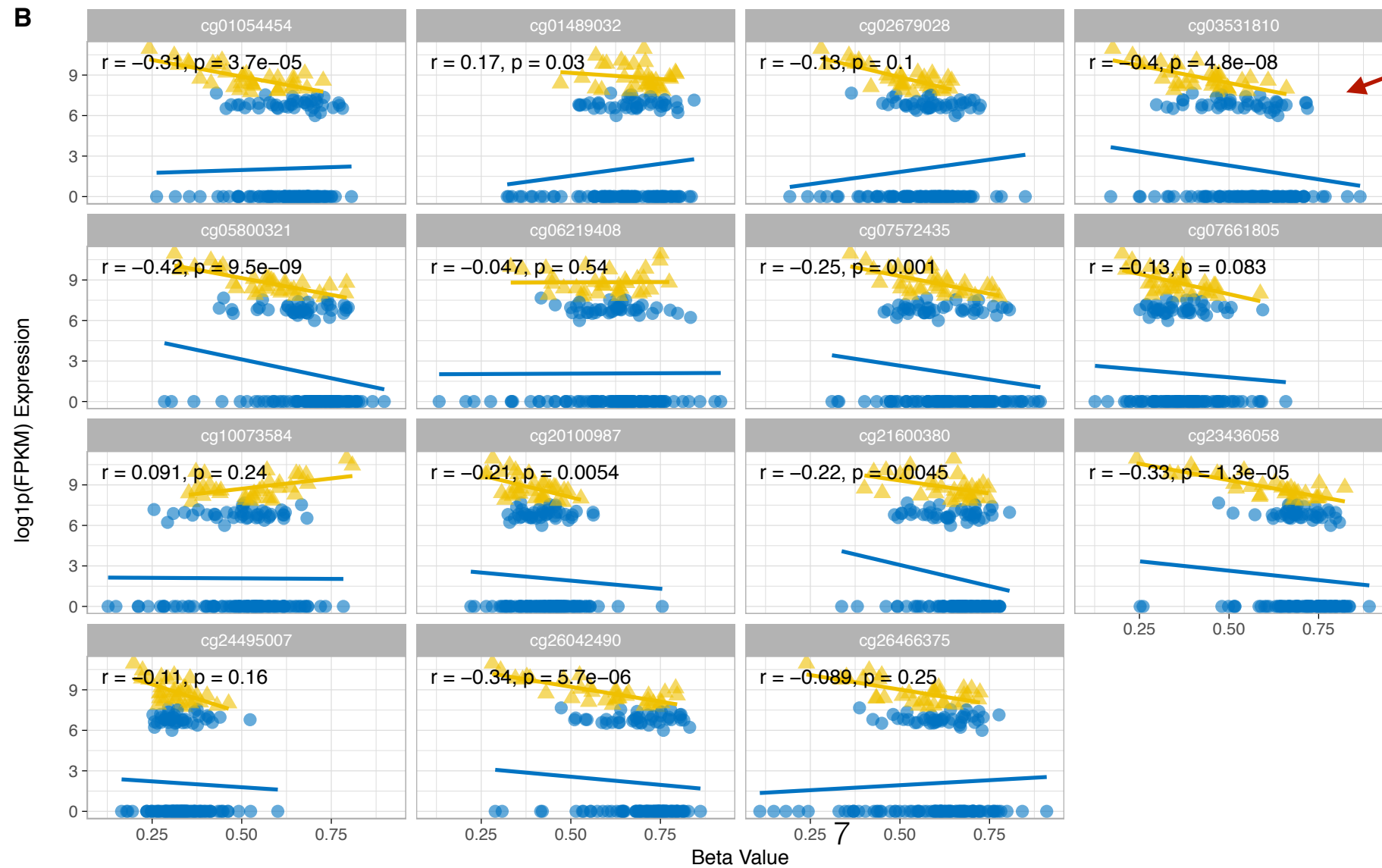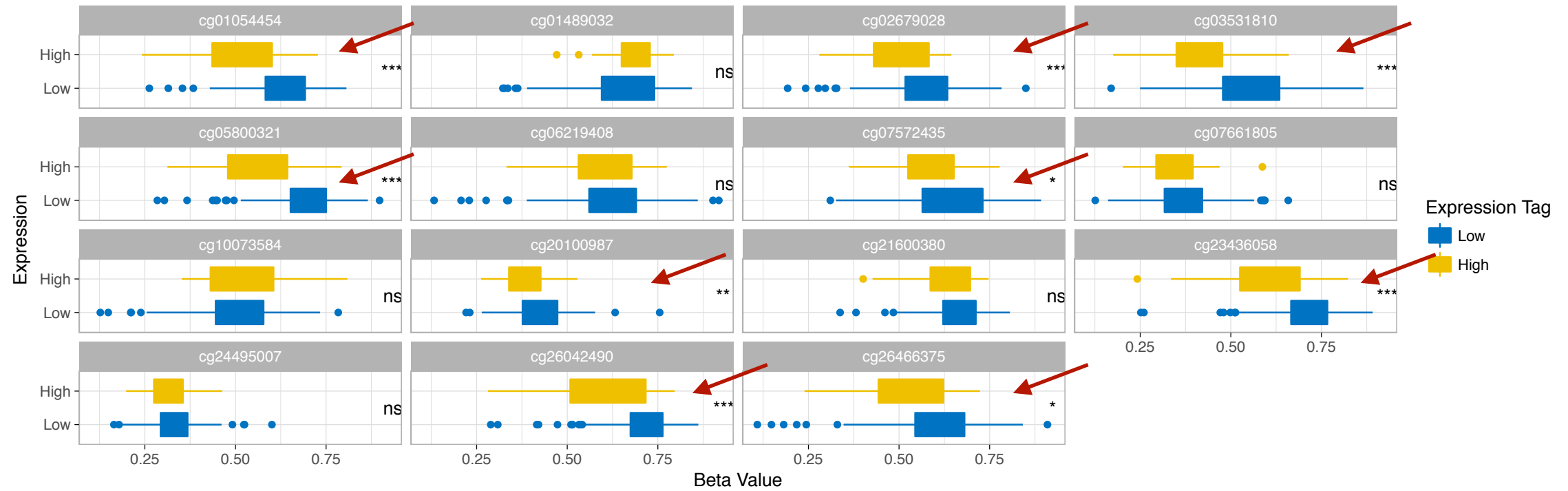- 14 in KIRC
- 13 in LIHC
- 9 in OV
- 2 in PAAD

$\longrightarrow$

n = 7 Candidates (PCAWG), p-val < 0.15
- 1 in KIRC
- 3 in LIHC
- 2 in OV
- 1 in PAAD

Pancreatic Adenocarcinoma
n = 177 patients in TCGA
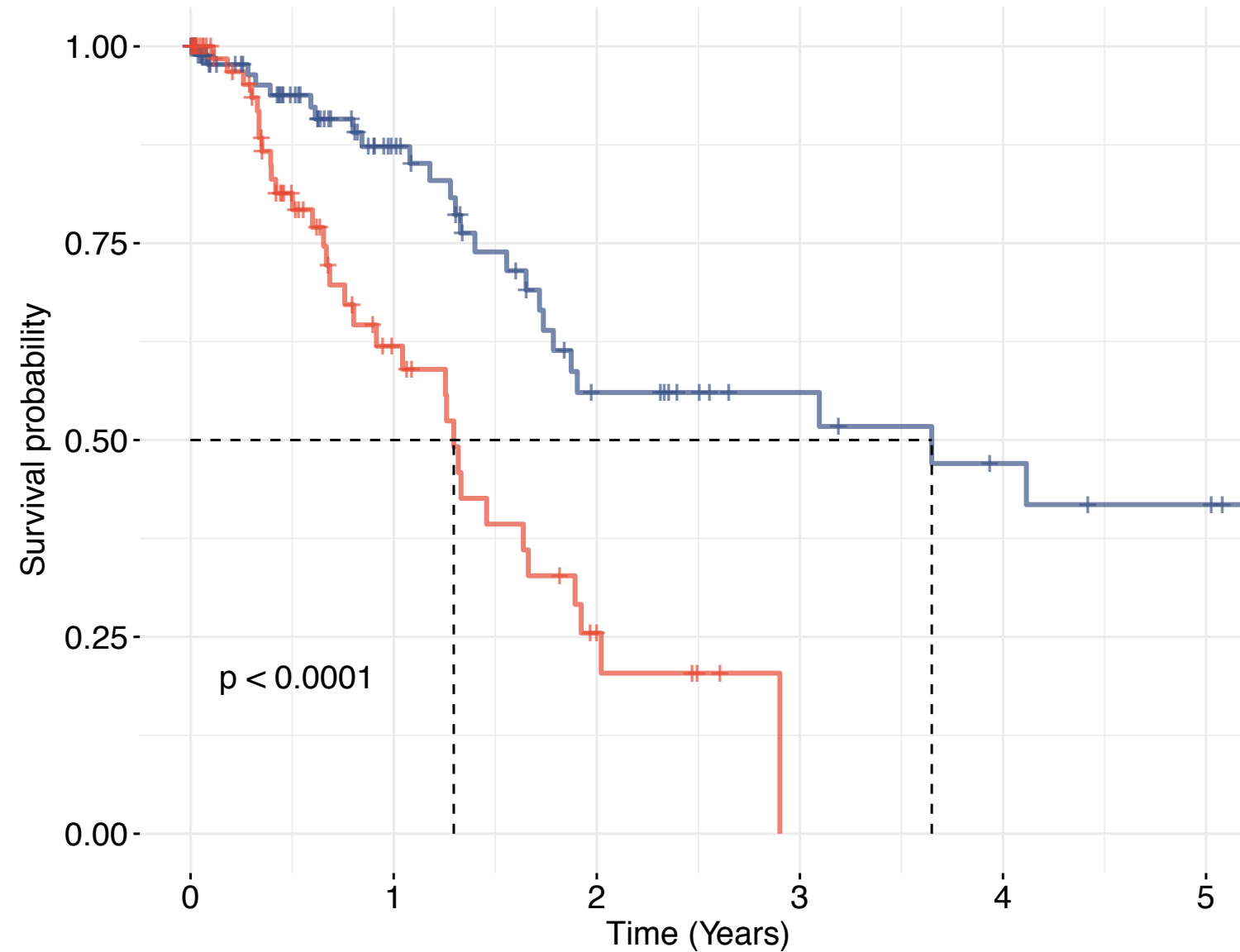# n = 68 patients in PCAWG

1. Validated in PCAWG
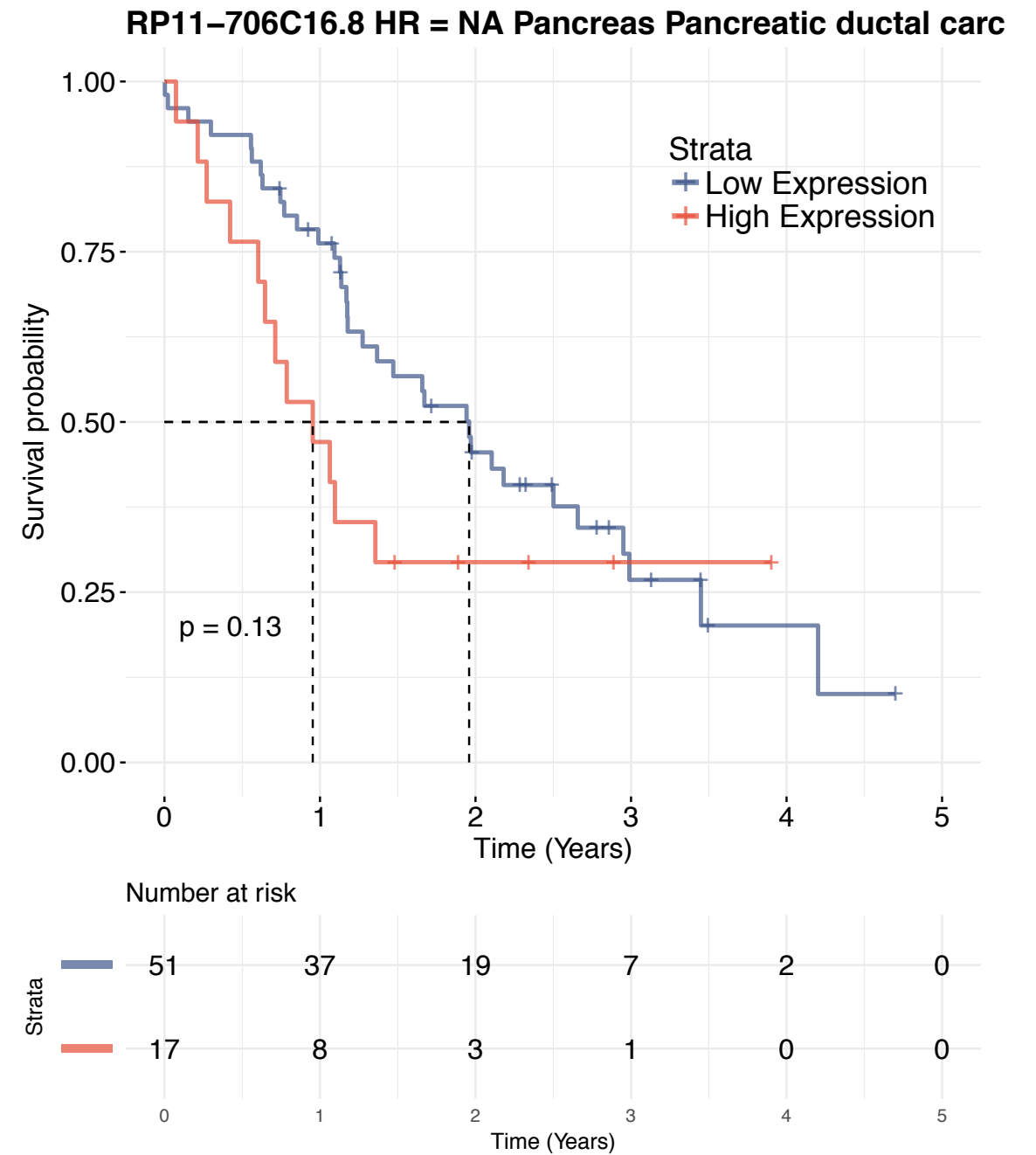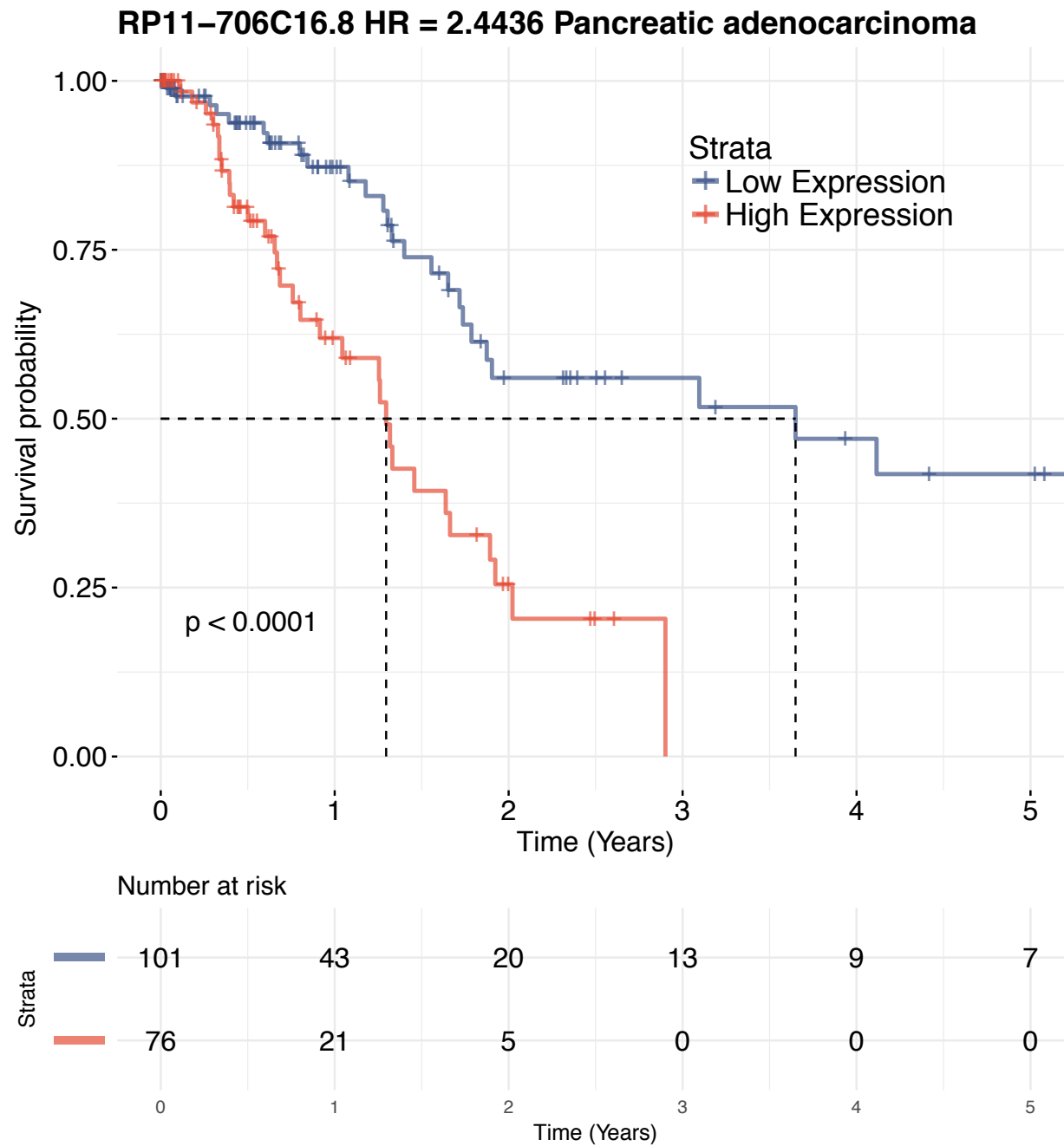
2. Methylation - CpG probes that overlap lncRNA
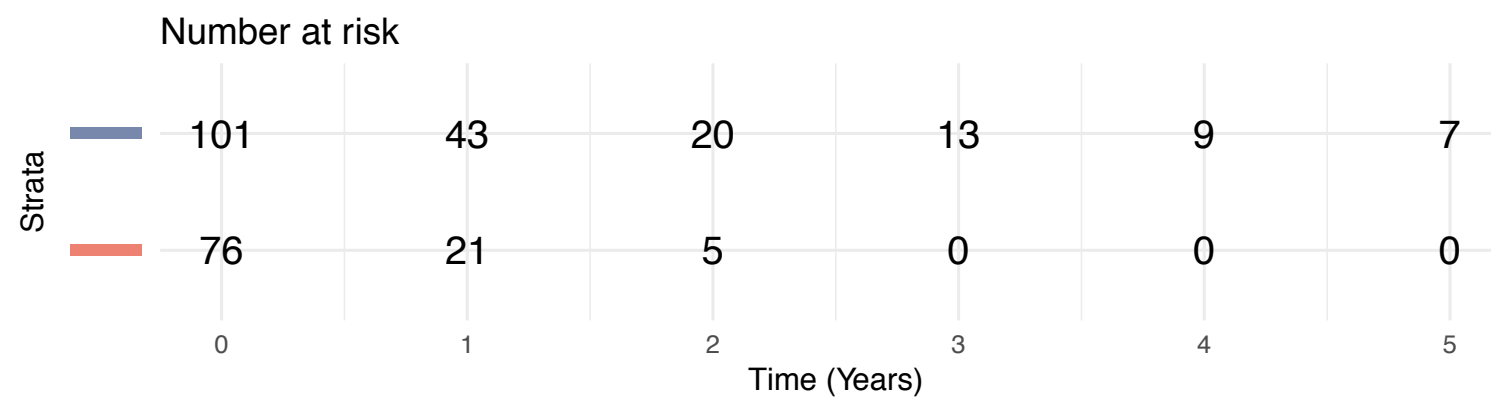

A RP11−706C16.8 pancreas Methylation vs Exp n= 169
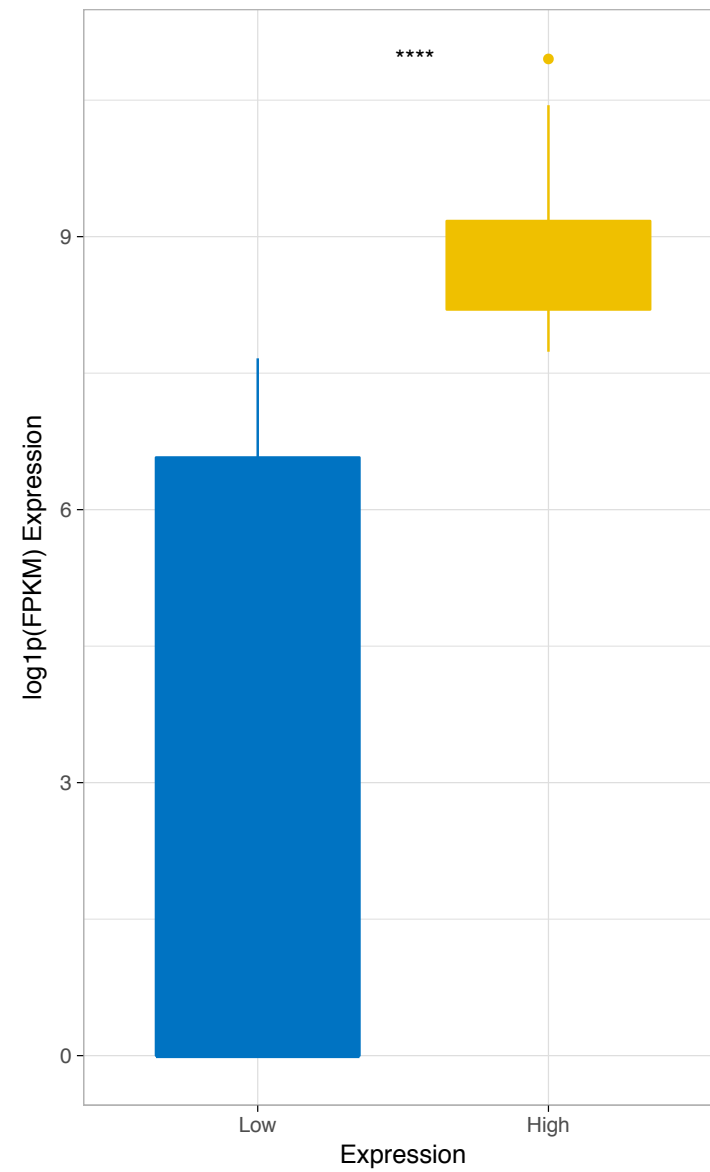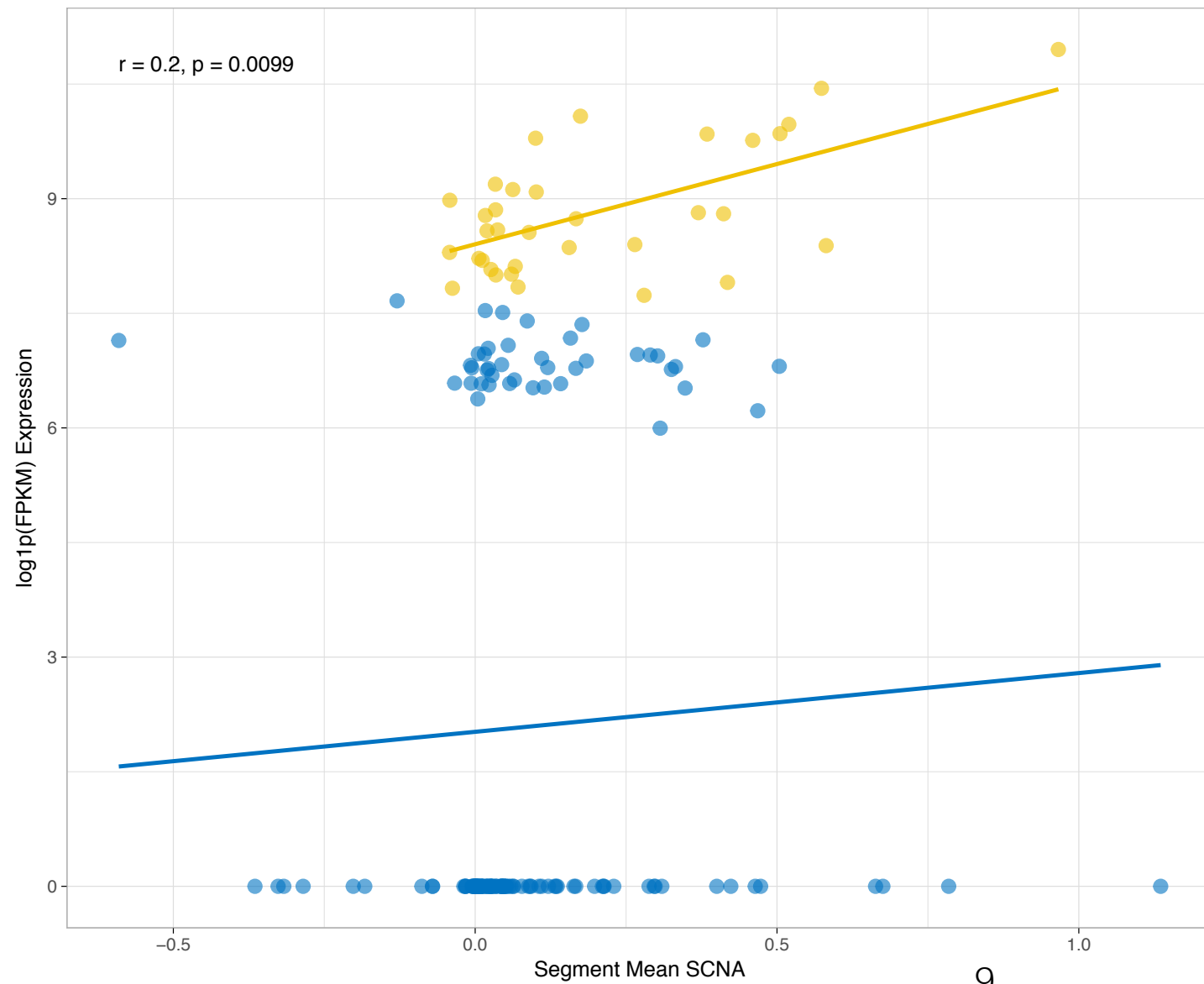
RP11−706C16.8 HR = 2.4436 Pancreatic adenocarcinoma

p < 0.0001
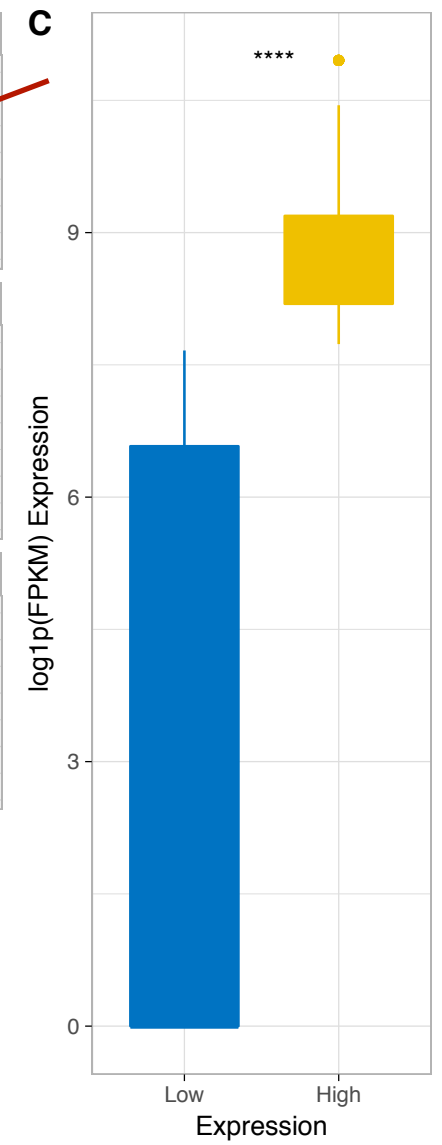
methylation protective?

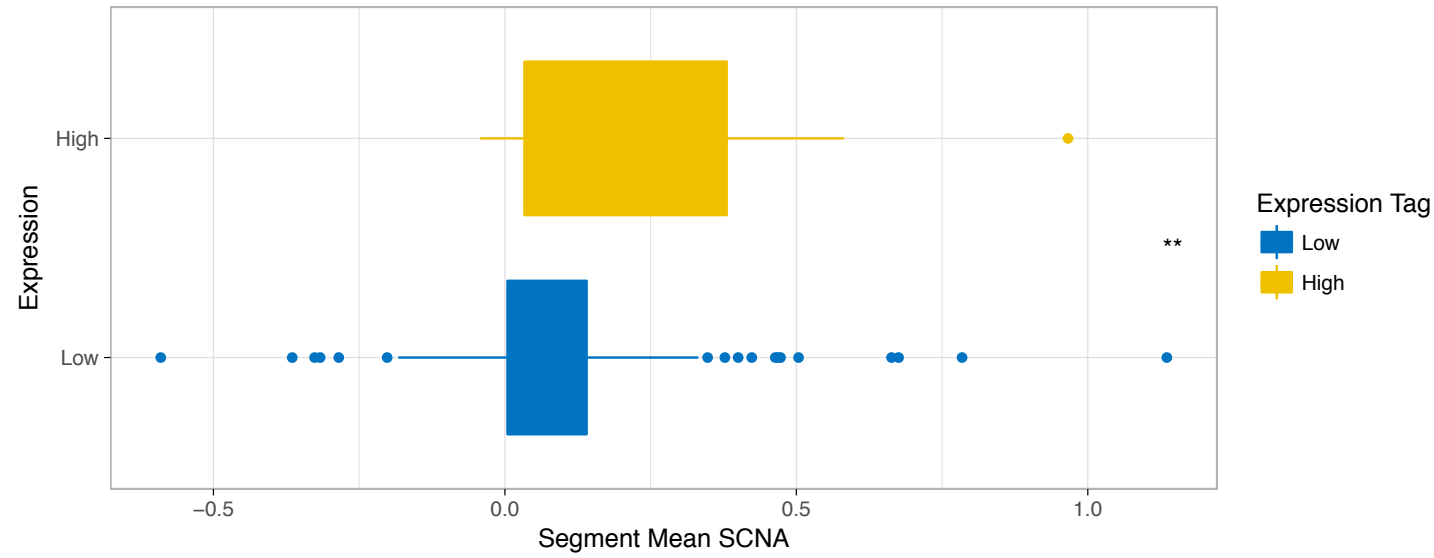Number at risk

Strata						
	101	43	20	13	9	7
	76	21	5	0	0	0

8
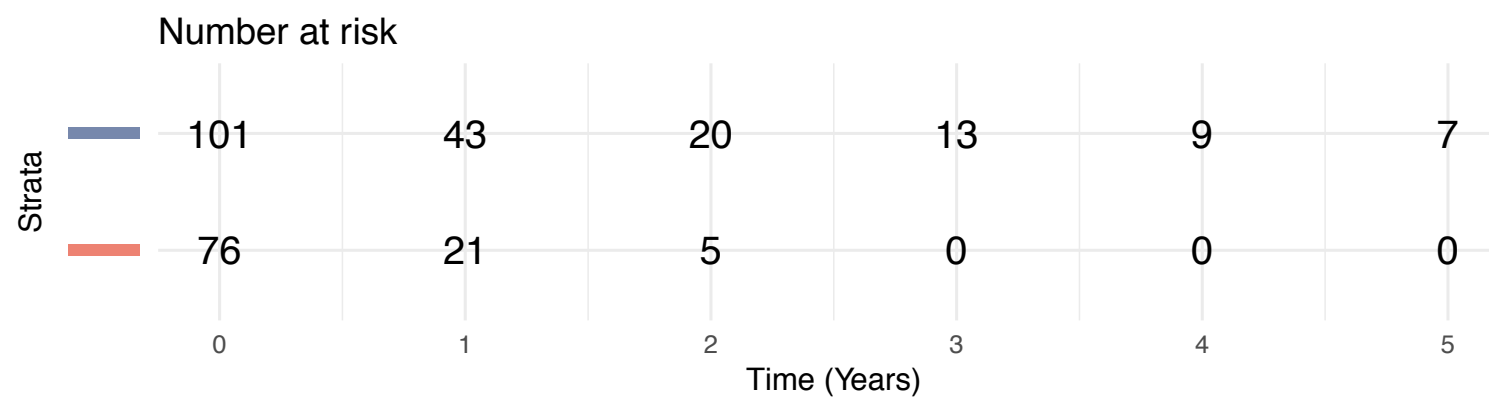
3. Copy number aberration analysis
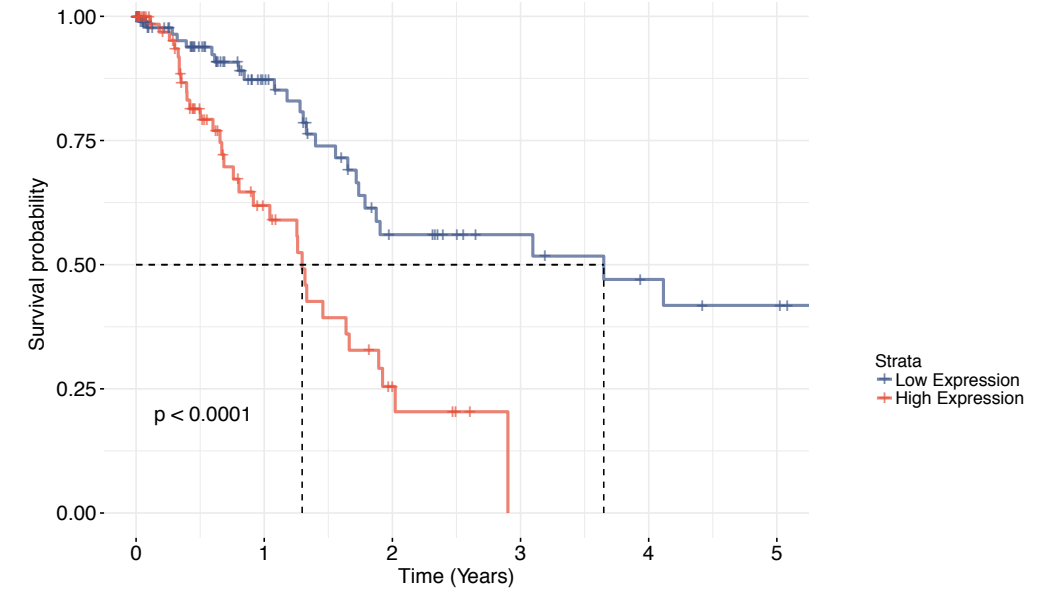
RP11−706C16.8 HR = 2.4436 Pancreatic adenocarcinoma
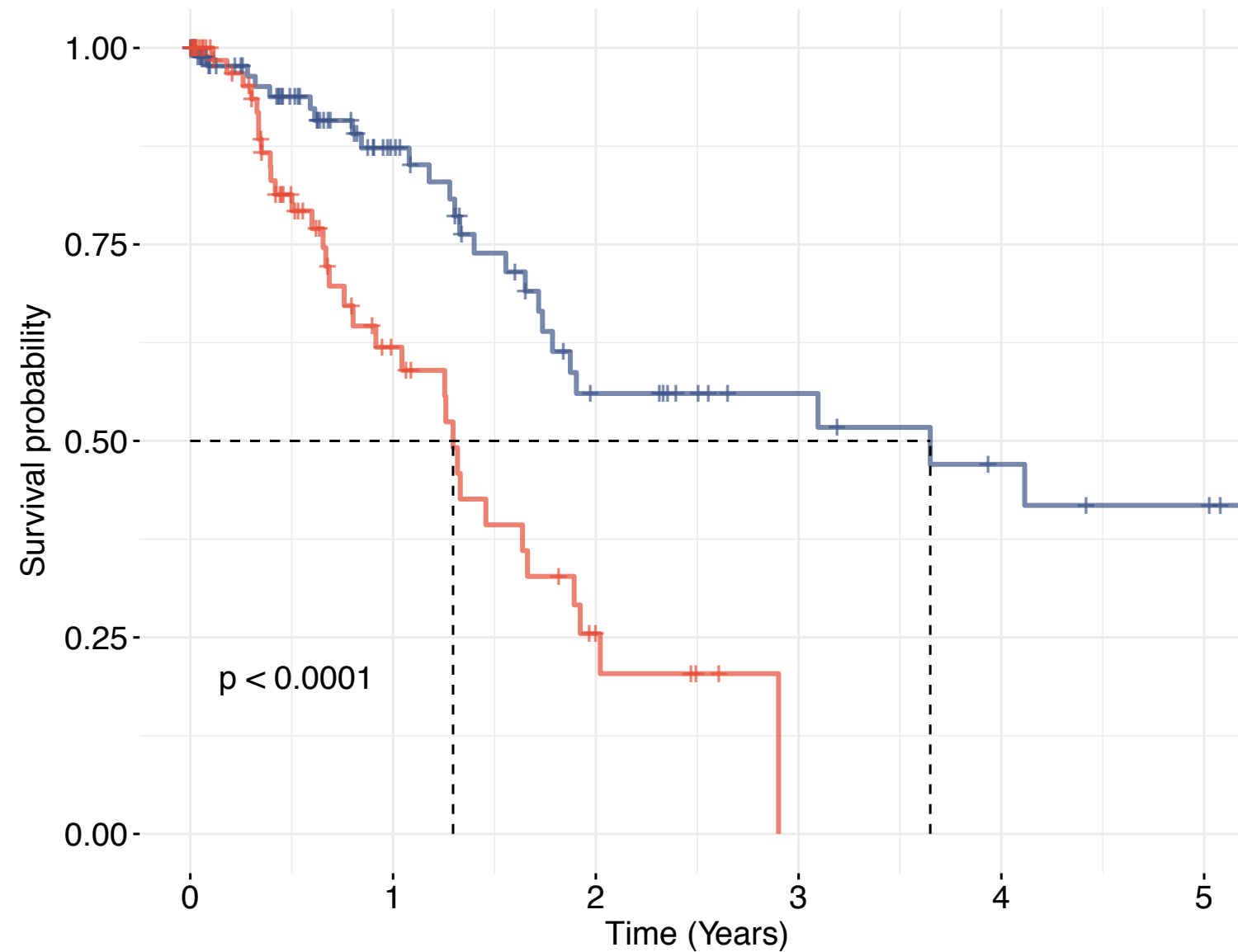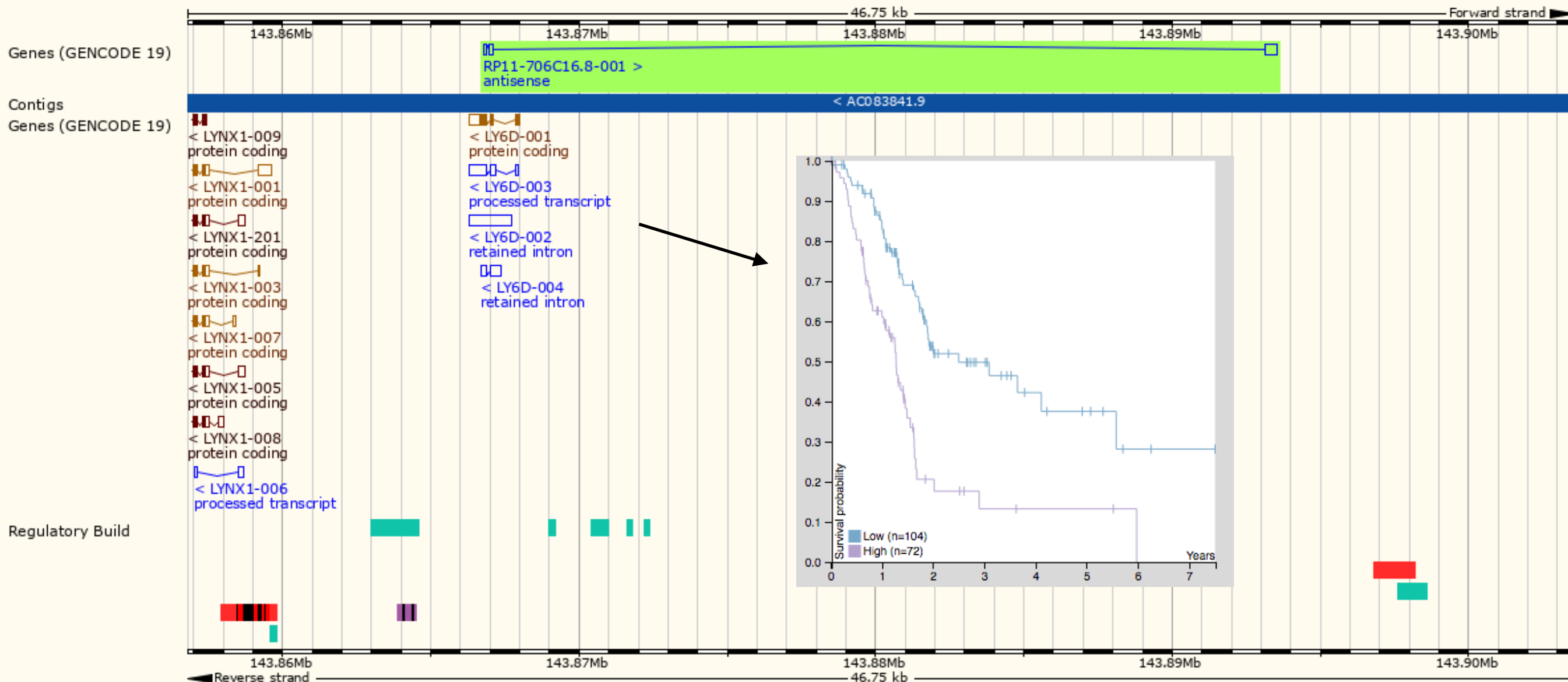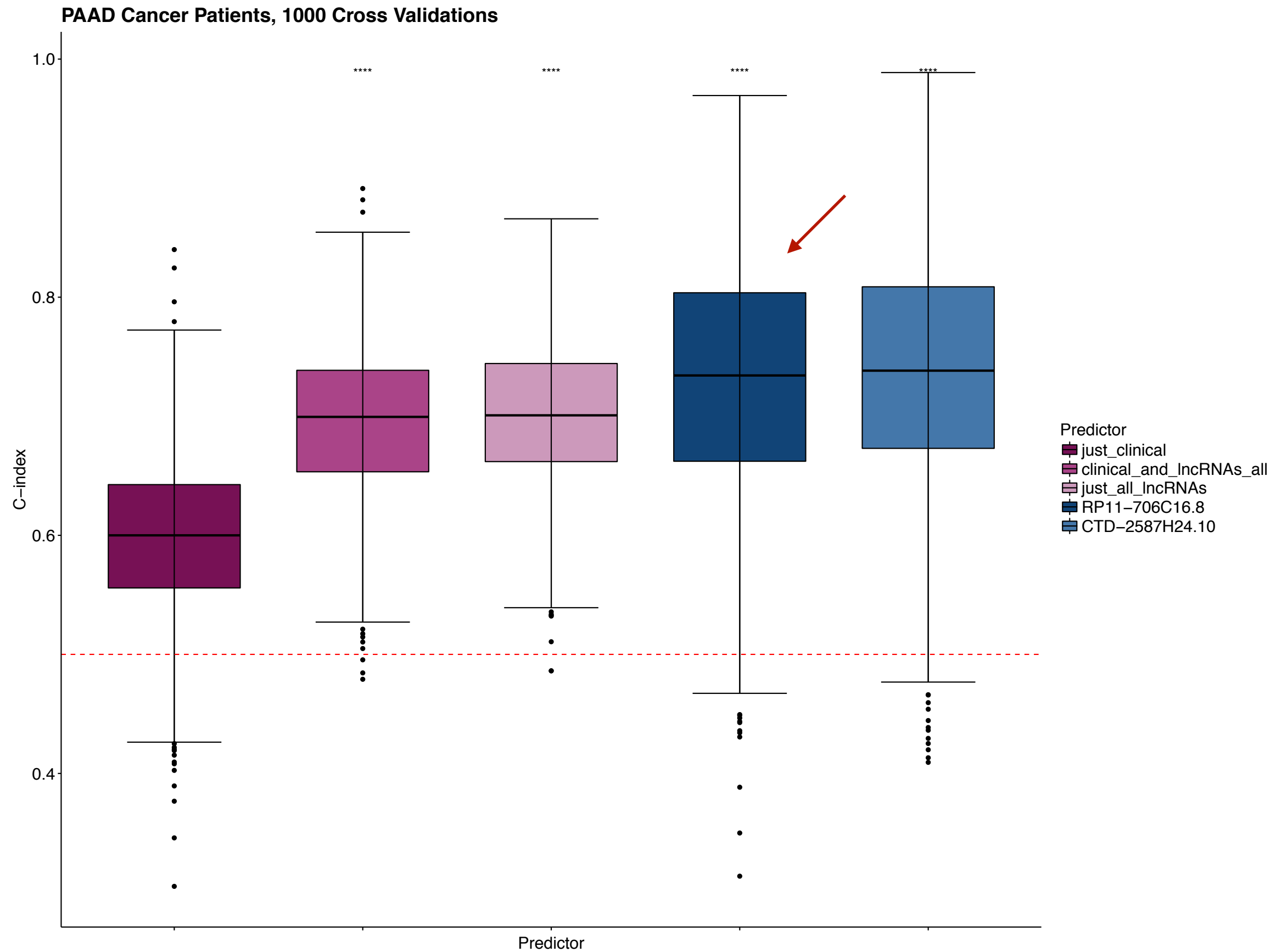
copy number aberration driving high expression?

no deletions in the high expression group

RP11-706C16.8 in Pancreatic Cancer

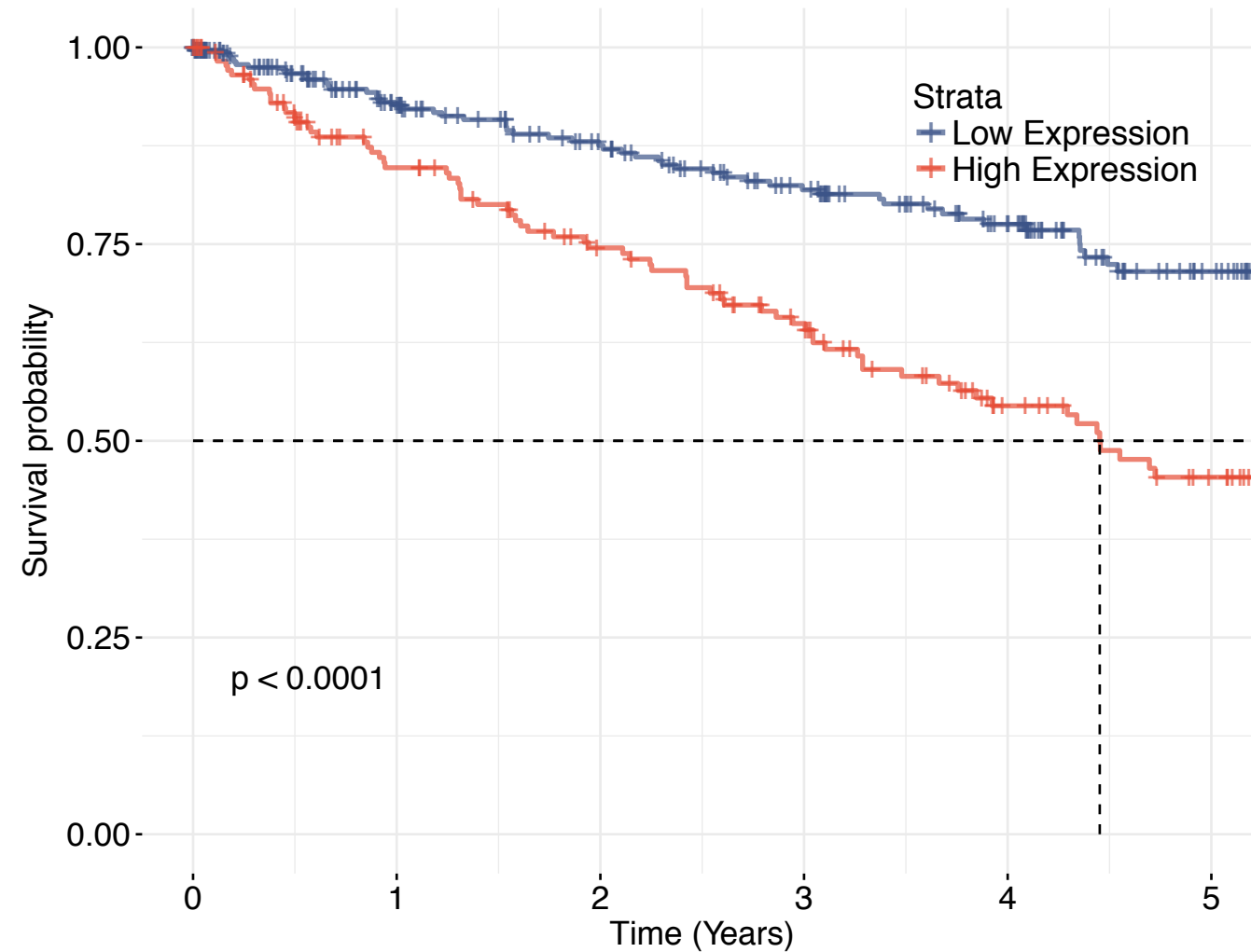RP11-706C16.8 in Pancreatic Cancer as a Predictor of Survival (TCGA)

PAAD Cancer Patients, 1000 Cross Validations

Kidney Renal Clear Cell Carcinoma
n = 481 patients in TCGA
n = 103 patients in PCAWG

1. Validated in PCAWG

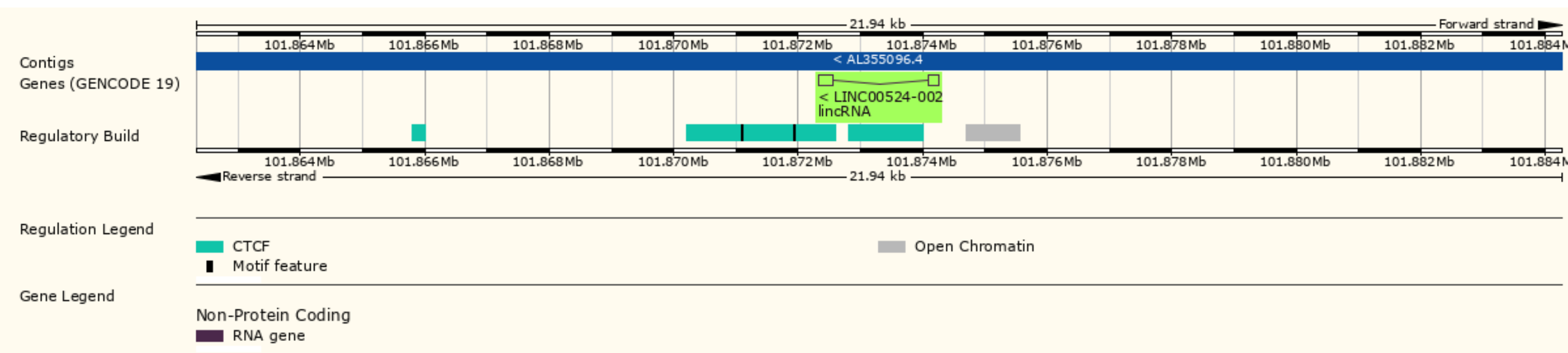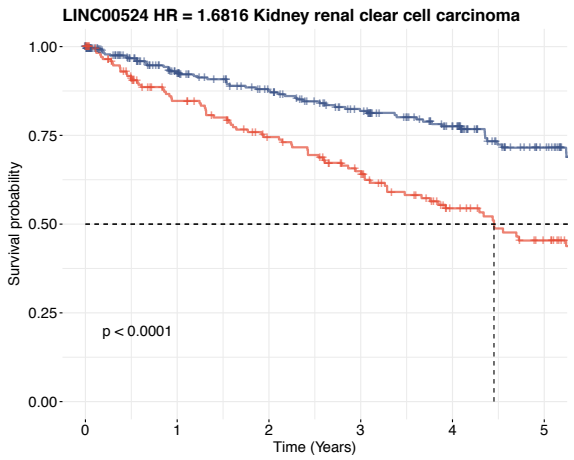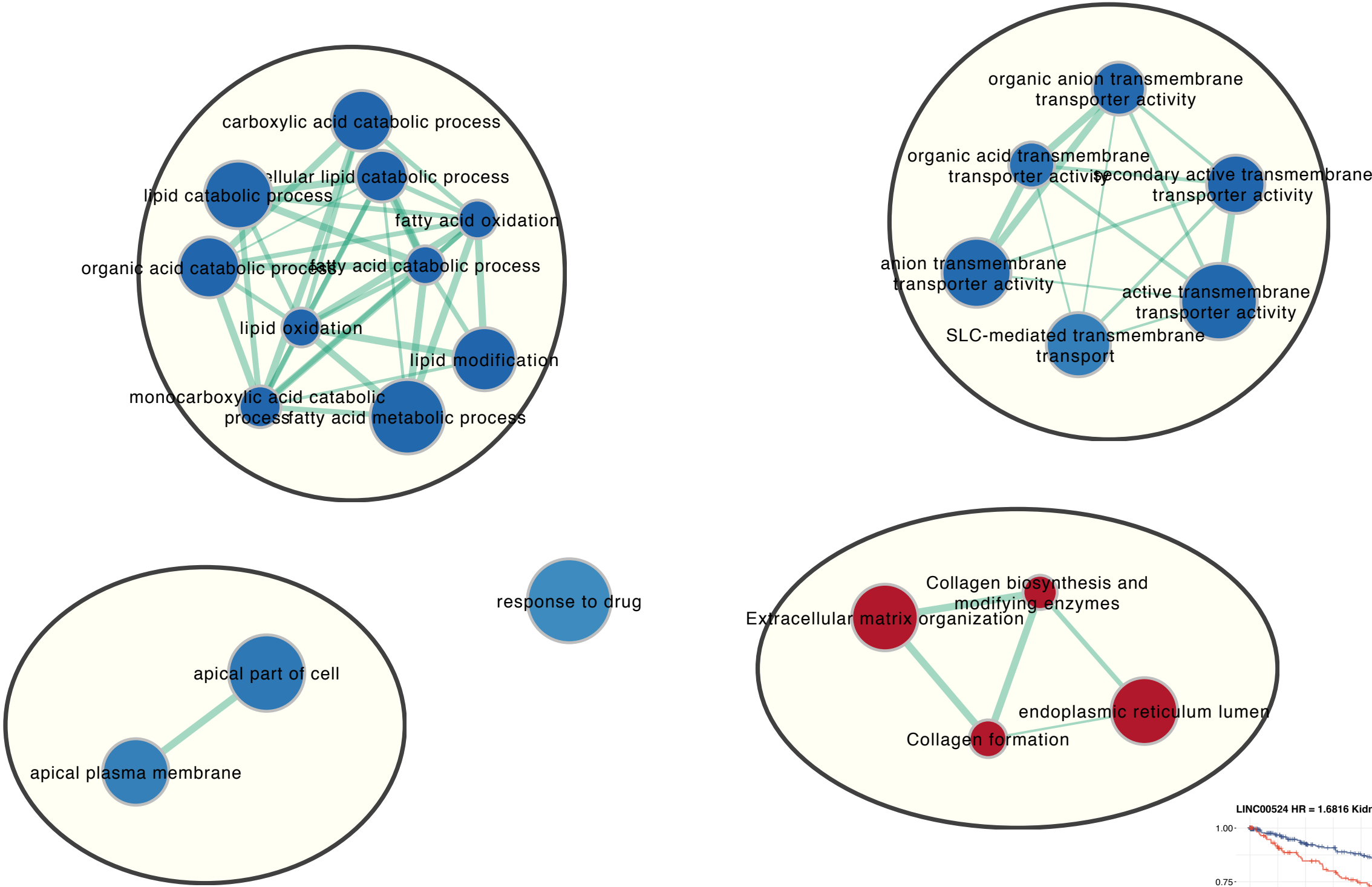No association between expression and copy number aberration or methylation
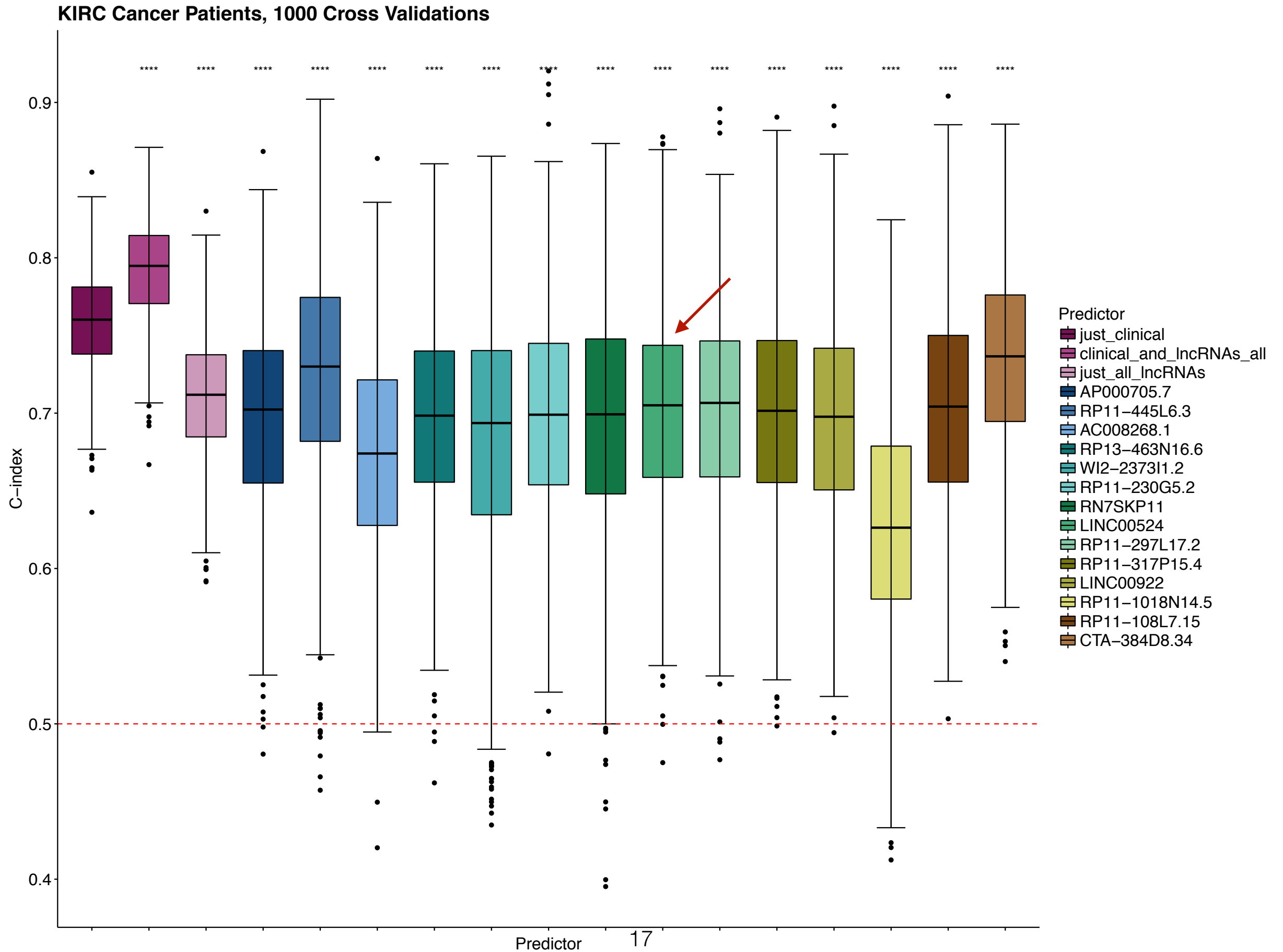
LINC00524 in KIRC

No association between expression and copy number aberration of methylation

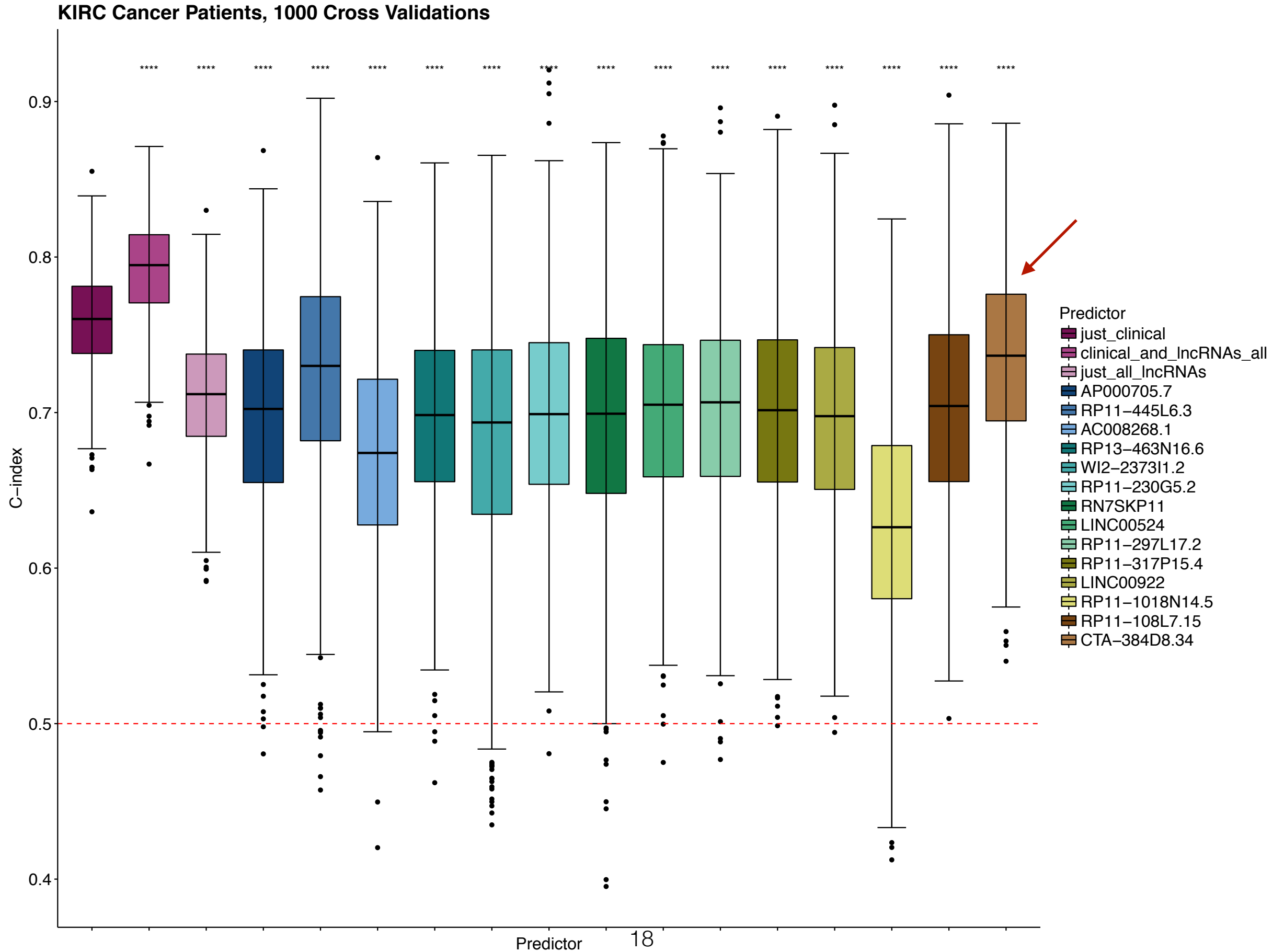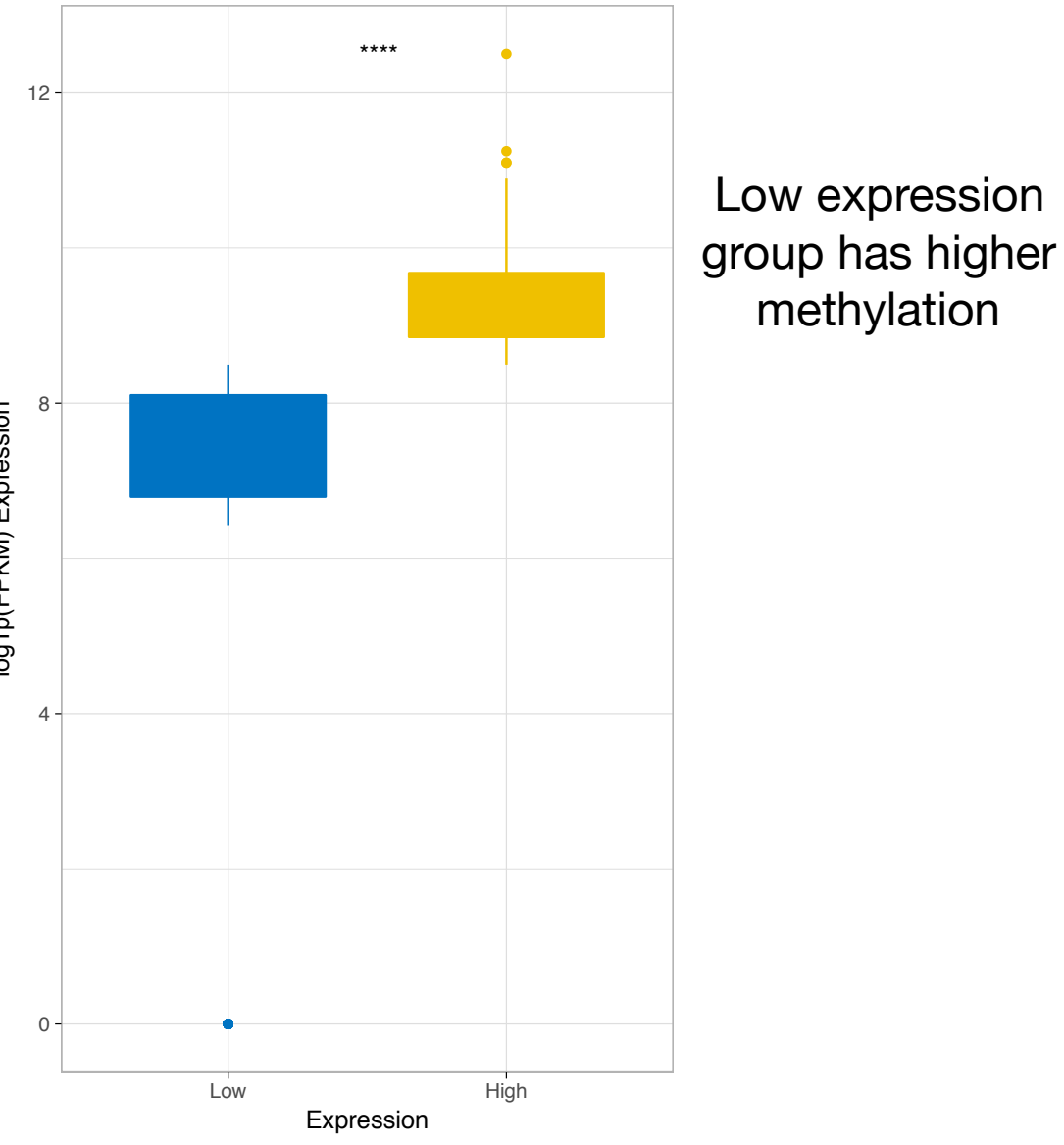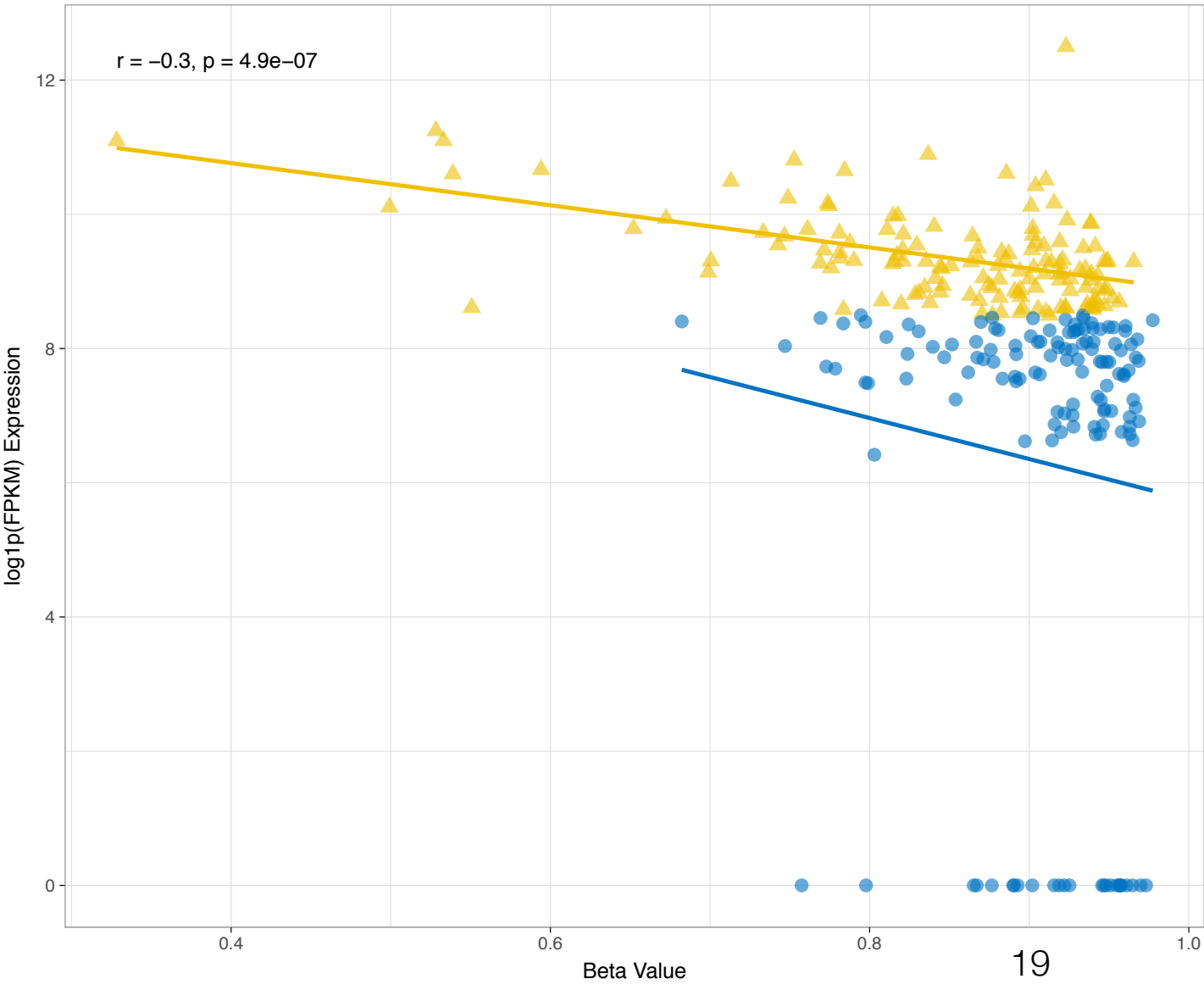# LINC00524 in KIRC Differential expression between high and low groups
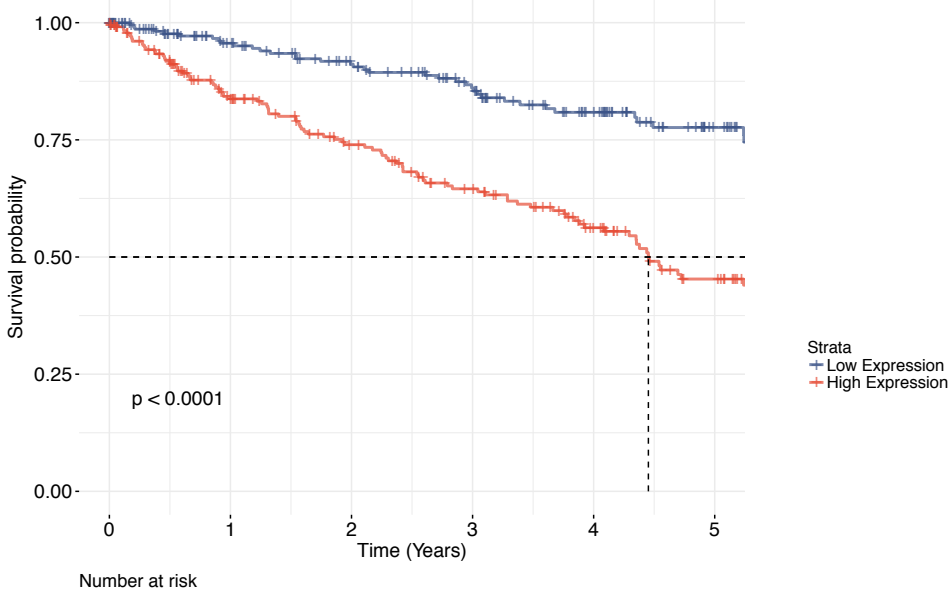
carboxylic acid catabolic process

cellular lipid catabolic process

lipid catabolic process

fatty acid oxidation

organic acid catabolic process

fatty acid catabolic process

lipid oxidation

lipid modification

monocarboxylic acid catabolic process

fatty acid metabolic process

organic anion transmembrane transporter activity

organic acid transmembrane transporter activity

secondary active transmembrane transporter activity

anion transmembrane transporter activity

active transmembrane transporter activity

SLC-mediated transmembrane transport

response to drug

apical part of cell

apical plasma membrane

Collagen biosynthesis and modifying enzymes

Extracellular matrix organization

endoplasmic reticulum lumen

Collagen formation

LINC00524 HR = 1.6816 Kidney renal clear cell carcinoma

Survival probability

p < 0.0001

Time (Years)

16

LINC00524 in KIRC as a Predictor of Survival (TCGA)

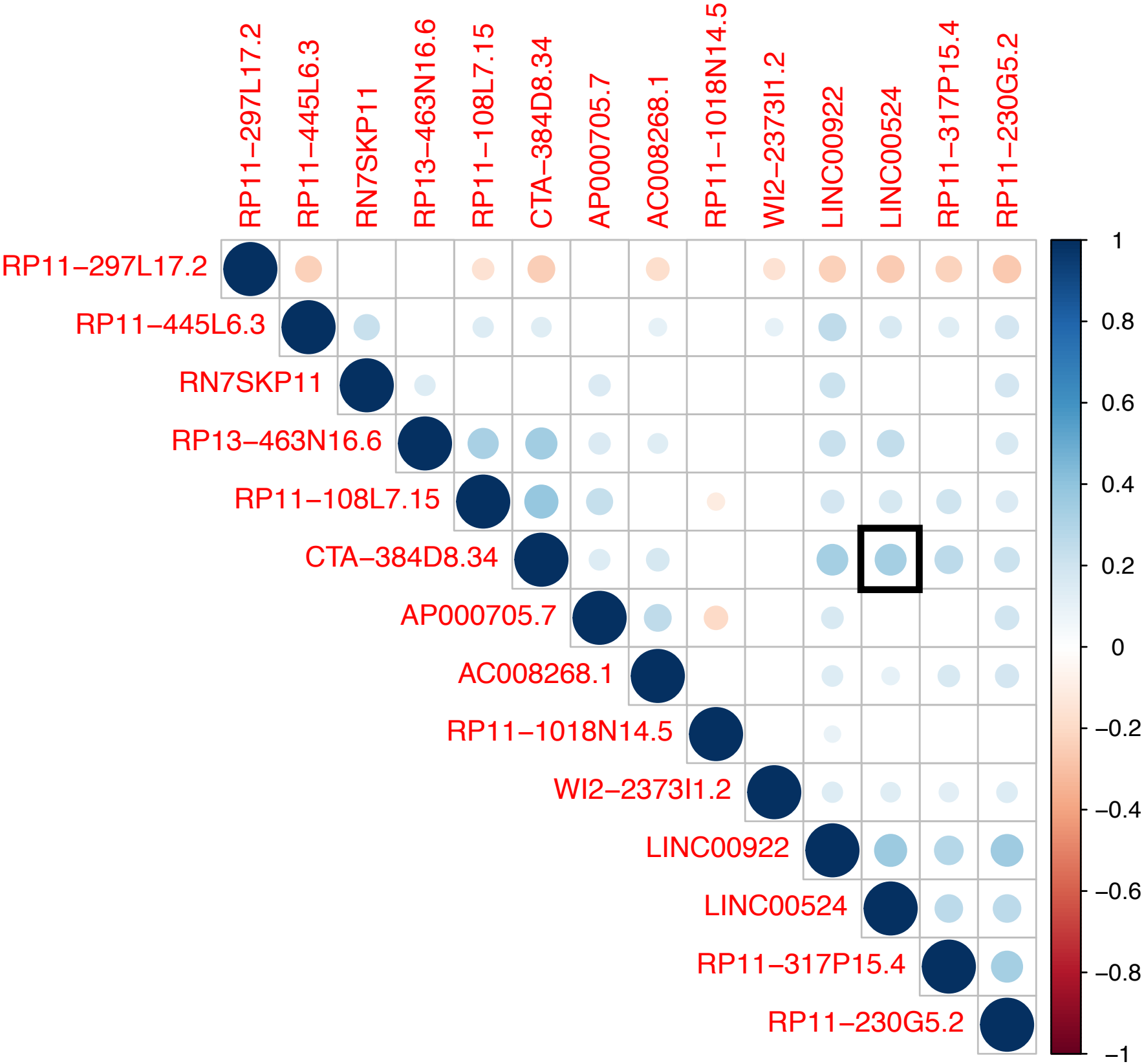KIRC Cancer Patients, 1000 Cross Validations

C-index

Predictor

Predictor
- just_clinical
- clinical_and_lncRNAs_all
- just_all_lncRNAs
- AP000705.7
- RP11−445L6.3
- AC008268.1
- RP13−463N16.6
- WI2−2373I1.2
- RP11−230G5.2
- RN7SKP11
- LINC00524
- RP11−297L17.2
- RP11−317P15.4
- LINC00922
- RP11−1018N14.5
- RP11−108L7.15
- CTA−384D8.34

17

CTA-384D8.34 in KIRC as a Predictor of Survival (TCGA)

KIRC Cancer Patients, 1000 Cross Validations

C-index

Predictor

Predictor
- just_clinical
- clinical_and_lncRNAs_all
- just_all_lncRNAs
- AP000705.7
- RP11−445L6.3
- AC008268.1
- RP13−463N16.6
- WI2−2373I1.2
- RP11−230G5.2
- RN7SKP11
- LINC00524
- RP11−297L17.2
- RP11−317P15.4
- LINC00922
- RP11−1018N14.5
- RP11−108L7.15
- CTA−384D8.34

CTA-384D8.34 in KIRC as a Predictor of Survival

CTA−384D8.34 kidney Methylation vs Exp n= 279

Expression Tag
- Low
- High

****

r = −0.3, p = 4.9e−07

Low expression group has higher methylation

****

19

CTA-384D8.34 and LINC00524 in KIRC are positively correlated

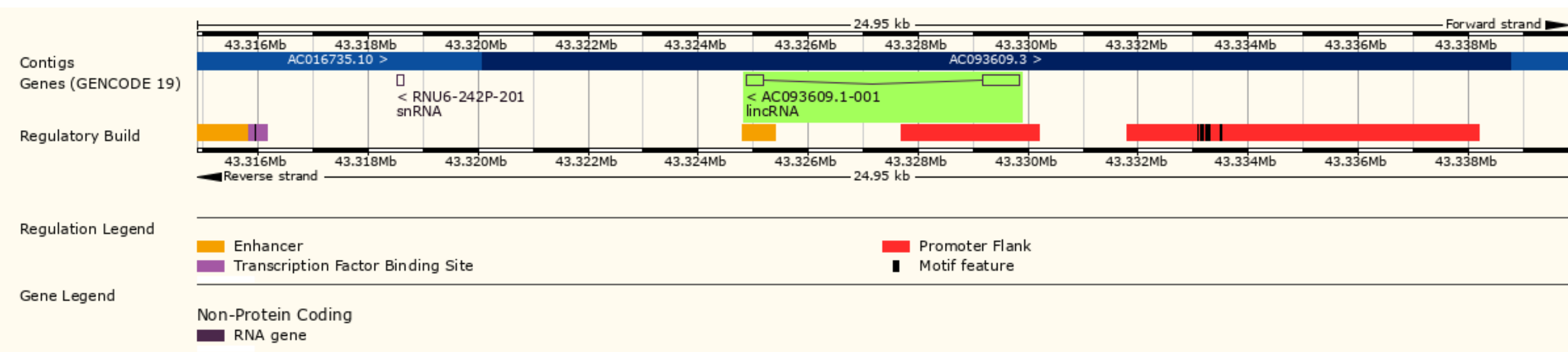# Liver Hepatocellular Carcinoma
n = 317 patients in TCGA
n = 84 patients in PCAWG

1. Validated in PCAWG

AC093609.1 HR = 0.4744 Liver hepatocellular carcinoma

Strata:
- Low Expression
- High Expression

p = 0.01

Number at risk

Strata	0	1	2	3	4	5
Low Expression	159	78	40	29	18	9
High Expression	158	81	55	39	29	15

AC093609.1 HR = 0.4348 Liver Hepatocellular carcinoma

Strata:
- Low Expression
- High Expression

p = 0.041

Number at risk

Strata	0	1	2	3	4	5
Low Expression	42	29	17	9	4	1
High Expression	42	31	25	20	8	2
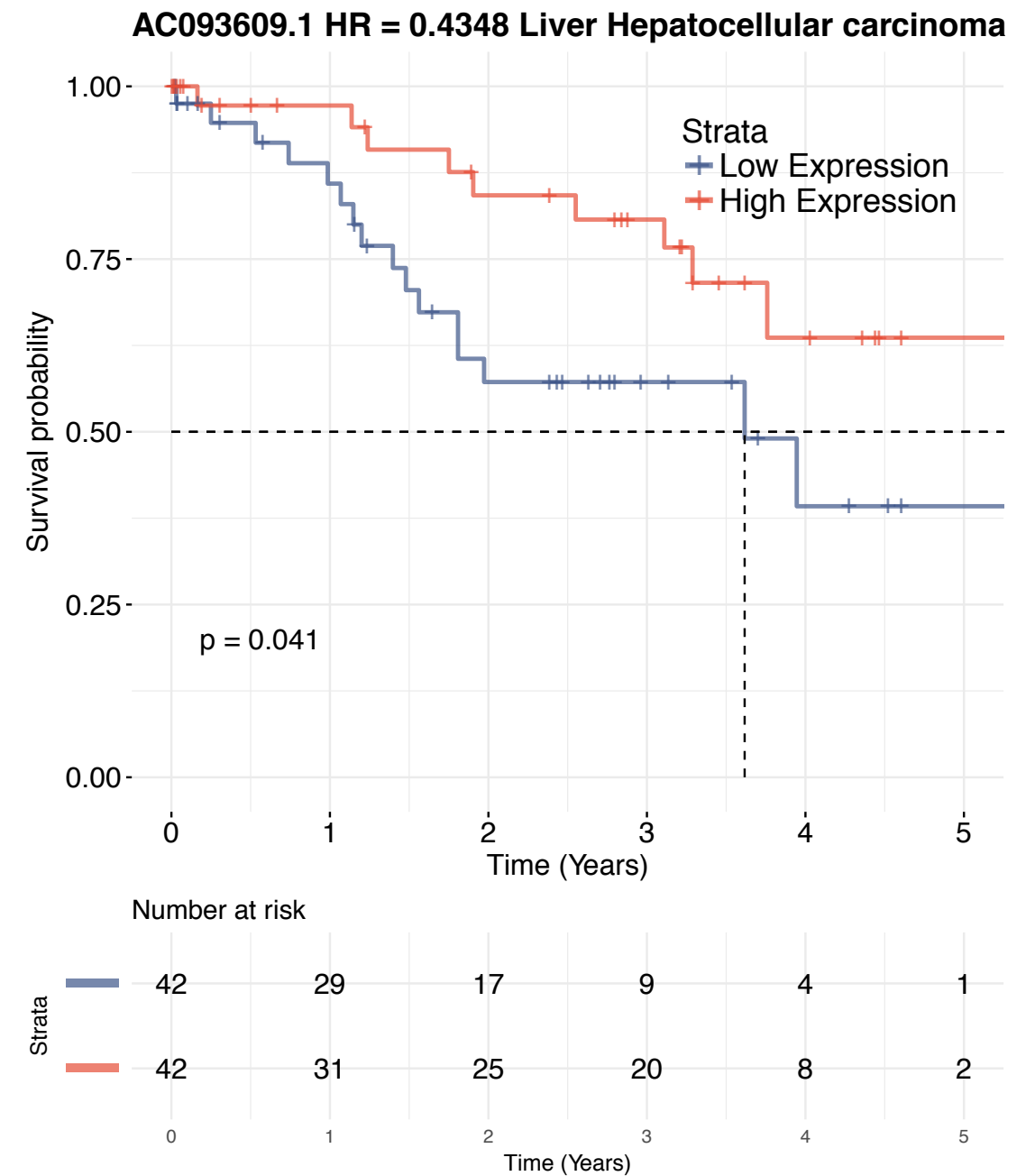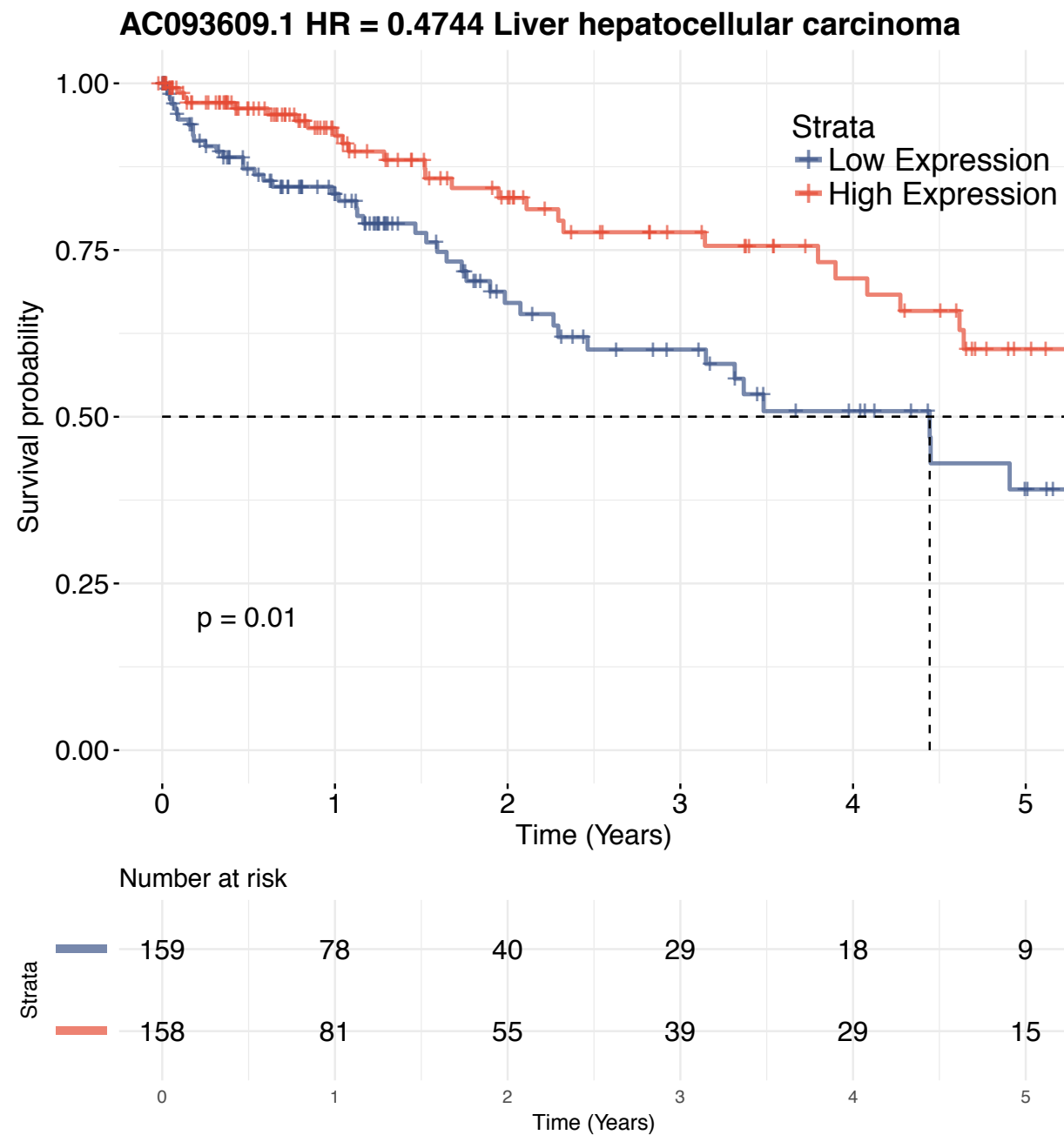
No association between expression and copy number aberration or methylation

AC093609.1 in LIHC Cancer

AC093609.1 in LIHC Cancer

LIHC Cancer Patients, 1000 Cross Validations

RP11-431M3.1 in LIHC Cancer

LIHC Cancer Patients, 1000 Cross Validations

C-index

Predictor

Predictor
- just_clinical
- clinical_and_lncRNAs_all
- just_all_lncRNAs
- AC017048.3
- AC104654.1
- AC093609.1
- RP1−154K9.2
- RP11−526A4.1
- AC114730.2
- AC006369.2
- AP001891.1
- RP11−101E14.2
- RP11−431M3.1
- RP11−867G23.10
- RP11−319E16.1
- RP11−363E7.4

RP11-431M3.1 in LIHC Cancer
Validated in PCAWG

RP11−431M3.1 HR = 2.2713 Liver hepatocellular carcinoma
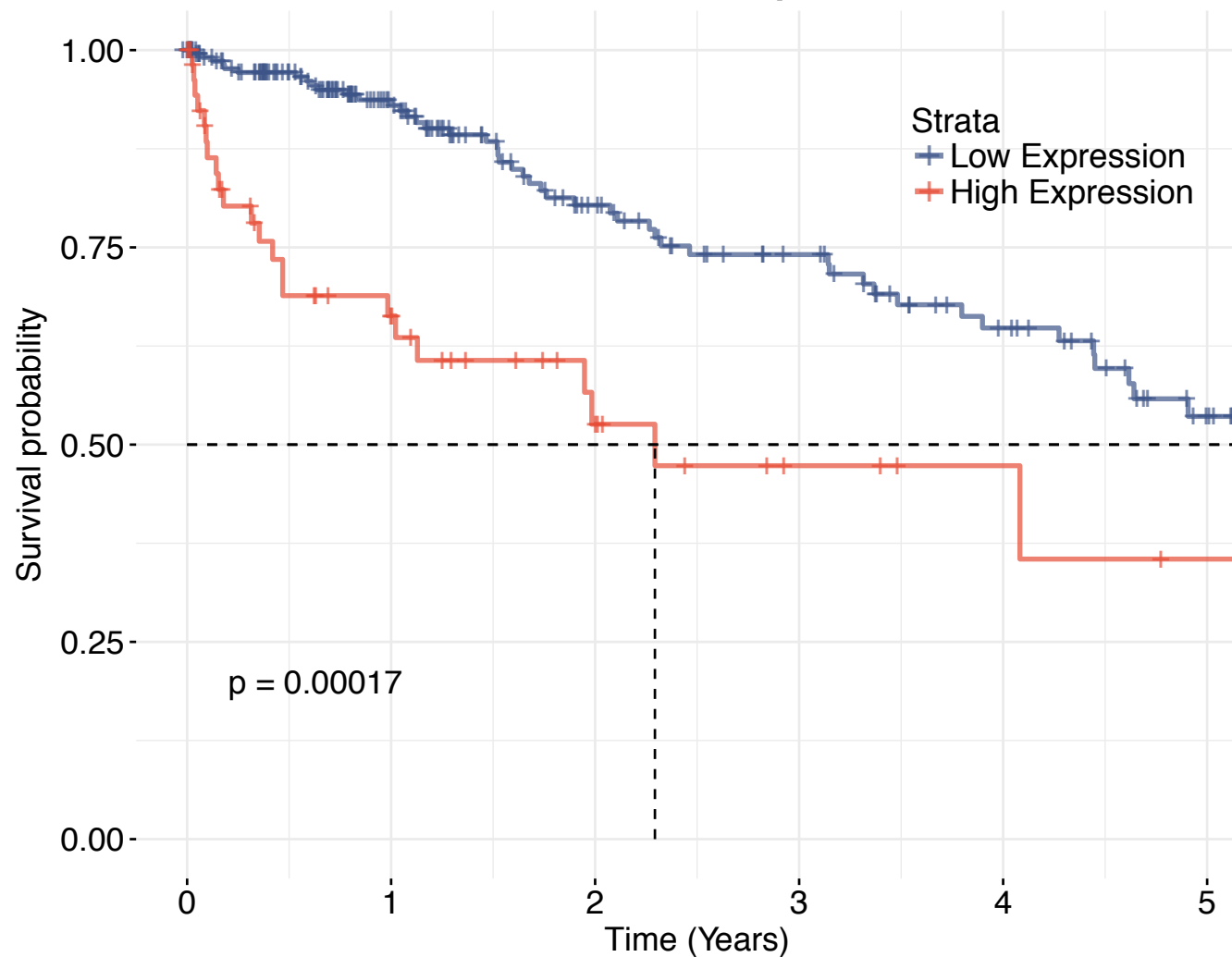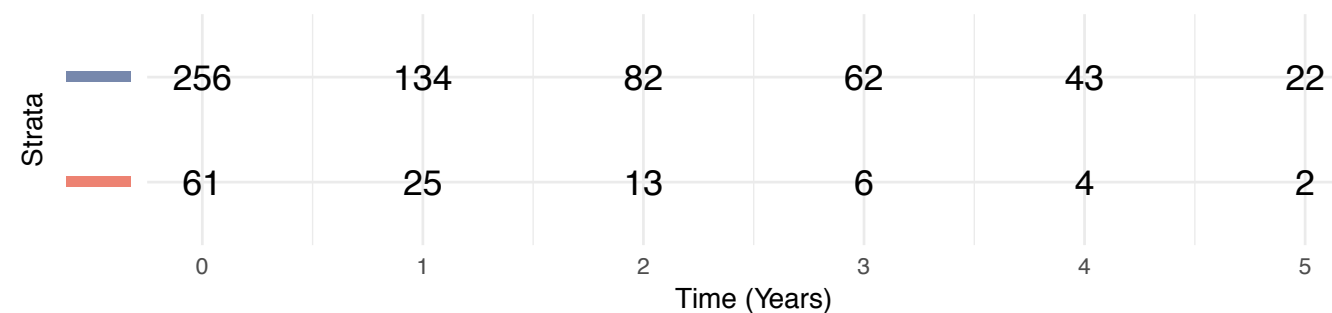
RP11−431M3.1 HR = 6.0281 Liver Hepatocellular carcinoma
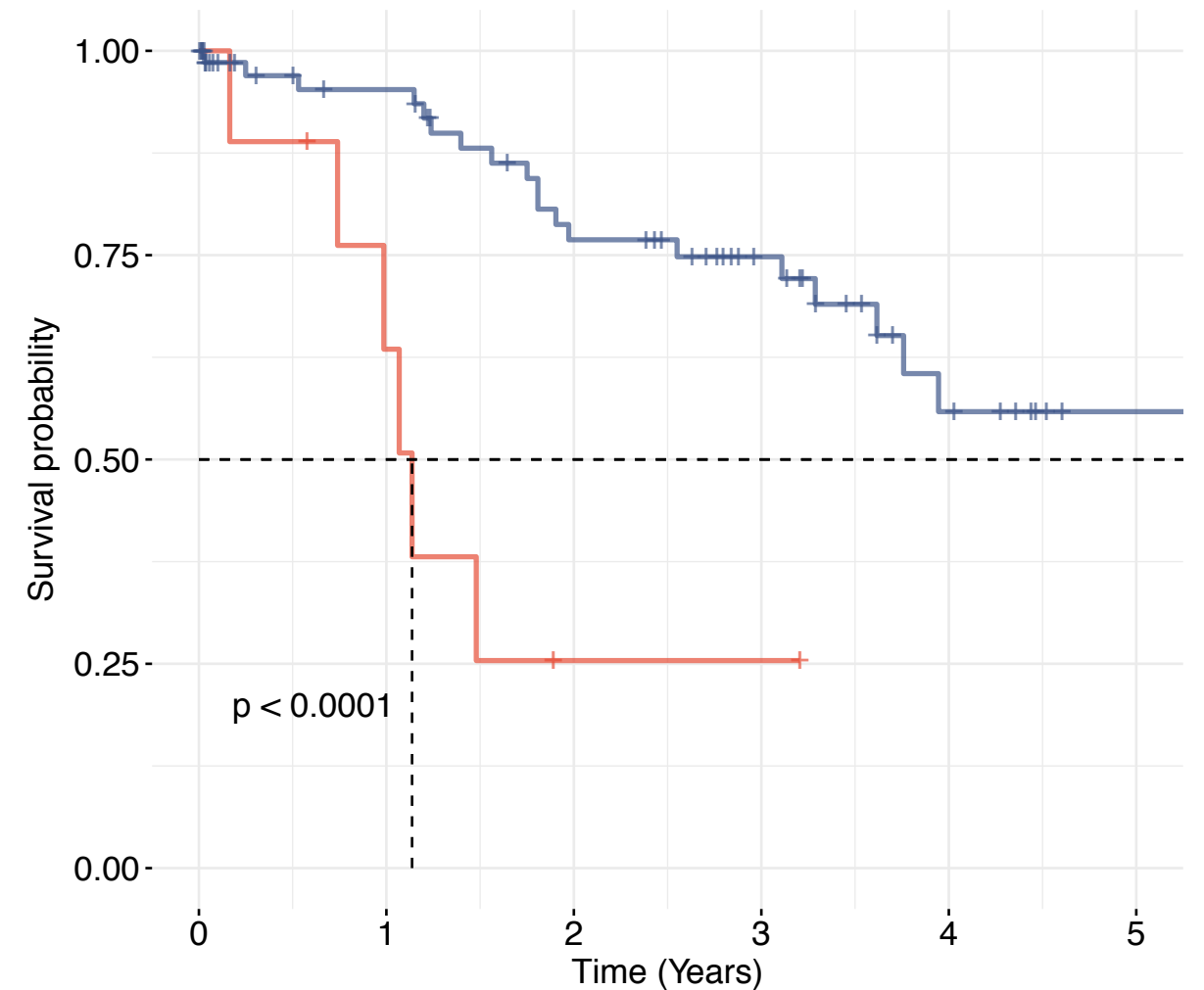
Strata
+ Low Expression
+ High Expression

p = 0.00017

p < 0.0001
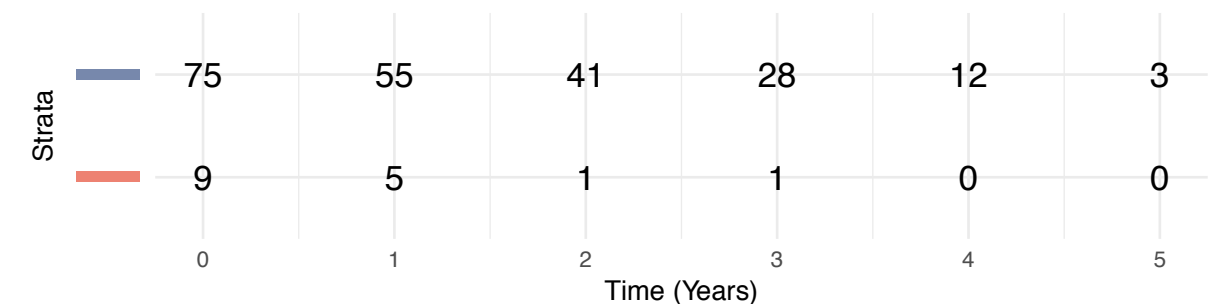
No association between expression and copy number aberration or methylation
Does not overlap any regulatory regions

26

RP11-431M3.1 in LIHC

RP11-431M3.1 HR = 2.2713 Liver hepatocellular carcinoma

Strata
+ Low Expression
+ High Expression

p = 0.00017

Number at risk

Strata

256	134	82	62	43	22
61	25	13	6	4	2
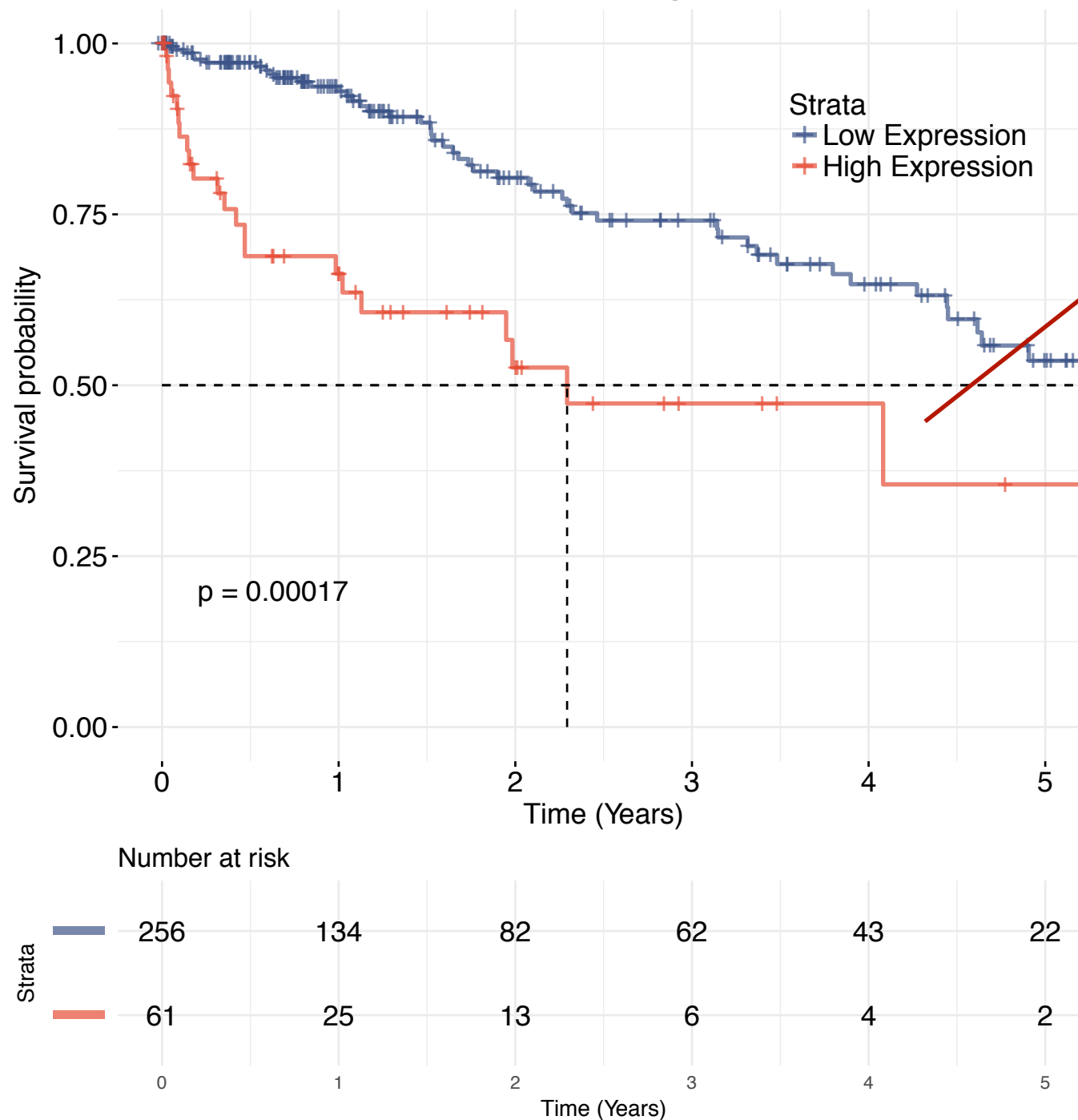
Upregulated pathways:
– Chromosome localization and cell cycle
– Nuclear segregation
– DNA damage checkpoint
– Chromatin remodelling

<u>Preliminary Functional Prediction:</u> Nuclear lncRNA? Most likely not in cis because there is nothing around it

Possible explanation for difference in expression:
- Mutations affecting enhancer or promoter of lncRNA
- Something inhibiting it, look at negative correlations

27

AC006369.2 in LIHC Cancer

LIHC Cancer Patients, 1000 Cross Validations
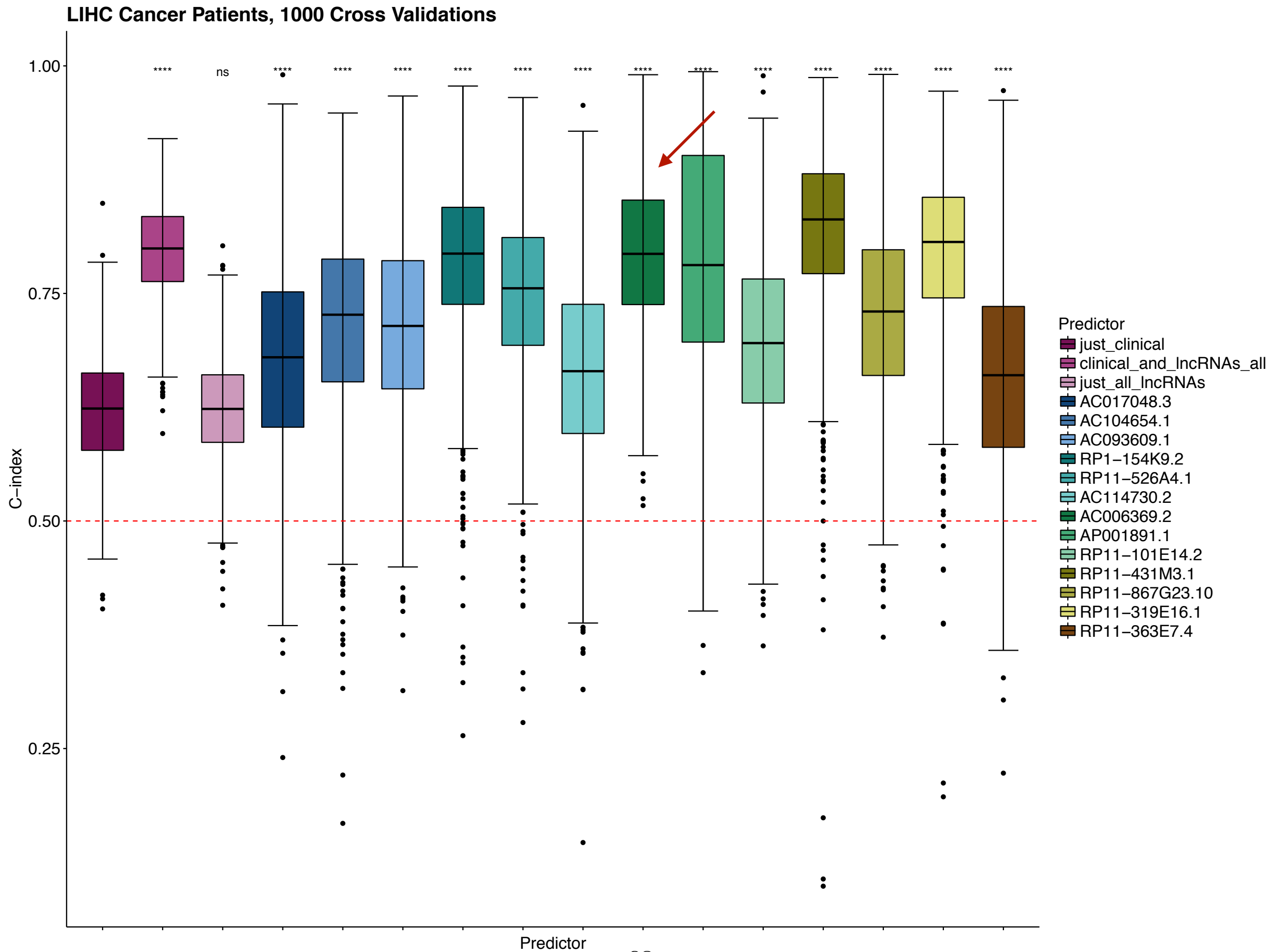
AC006369.2 in LIHC Cancer

AC006369.2 HR = 0.4402 Liver hepatocellular carcinoma
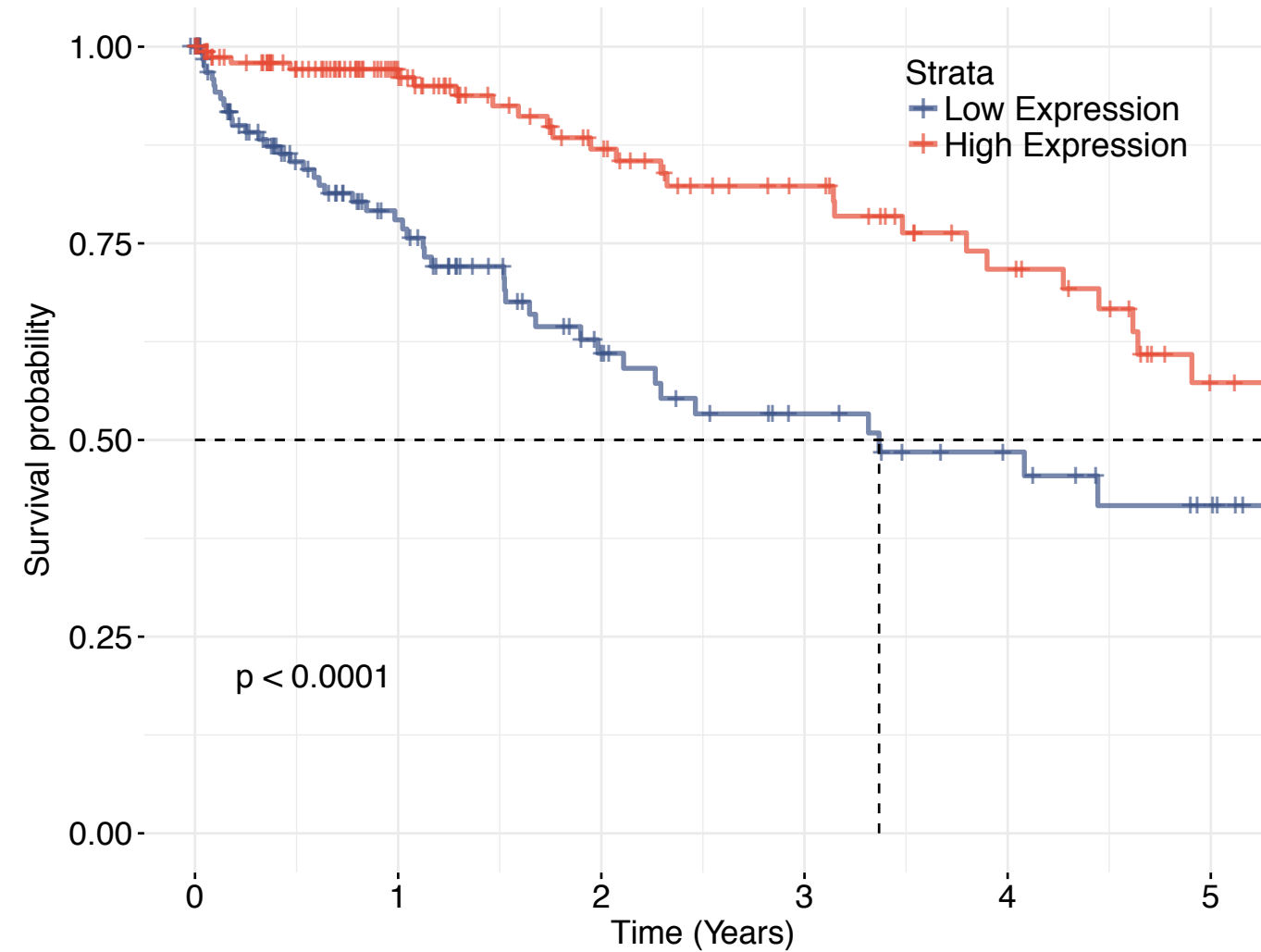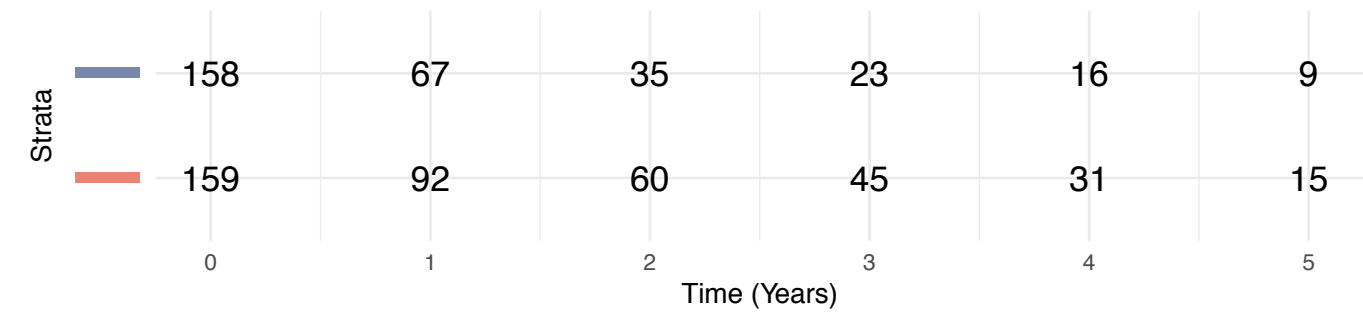
AC006369.2 HR = 0.4544 Liver Hepatocellular carcinoma

p < 0.0001

p = 0.053

Number at risk

Strata	158	67	35	23	16	9
	159	92	60	45	31	15

Strata	42	27	16	11	7	2
	42	33	26	18	5	1

AC006369.2 in LIHC Cancer

*Need to check where in the gene, the probe is
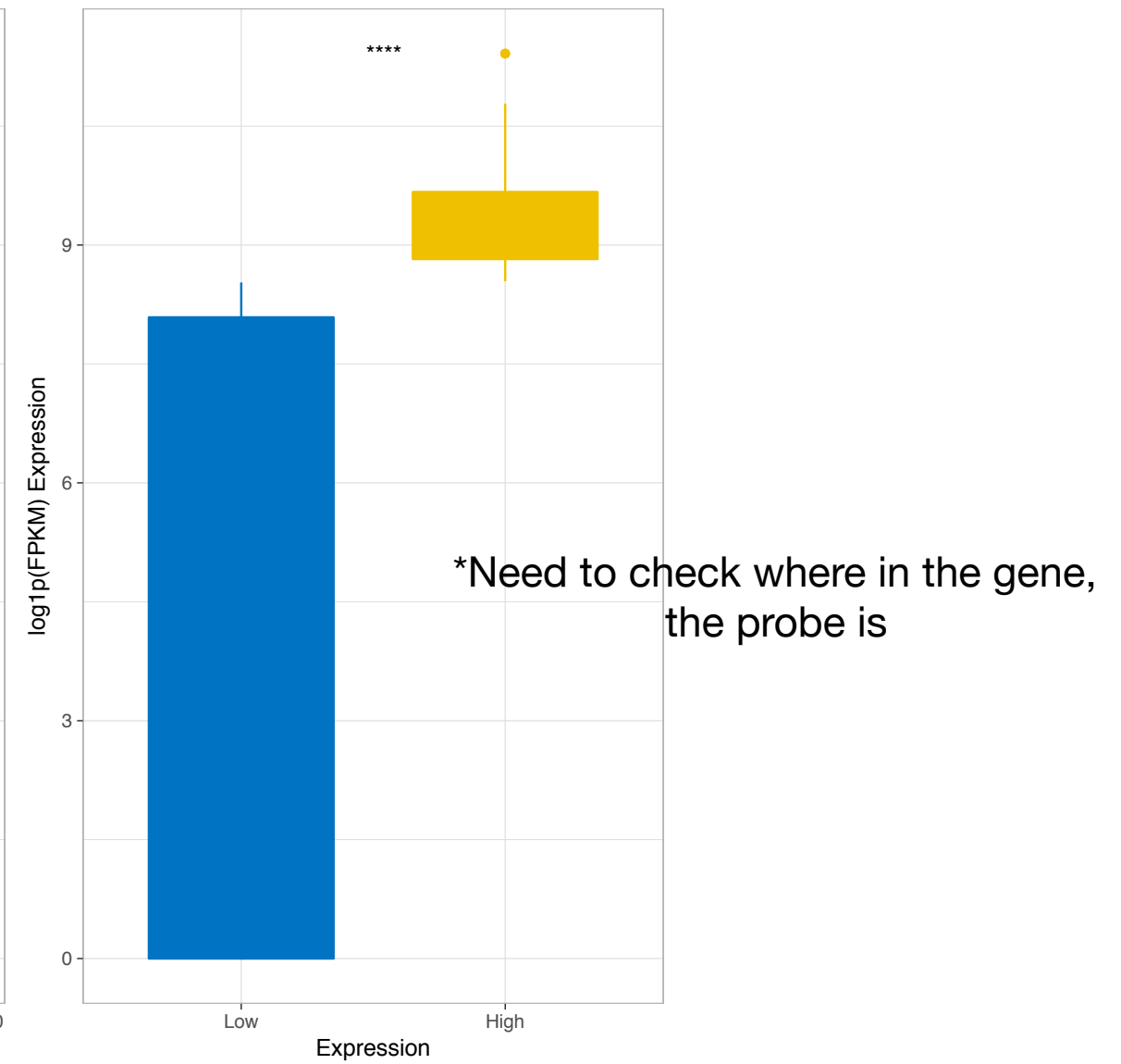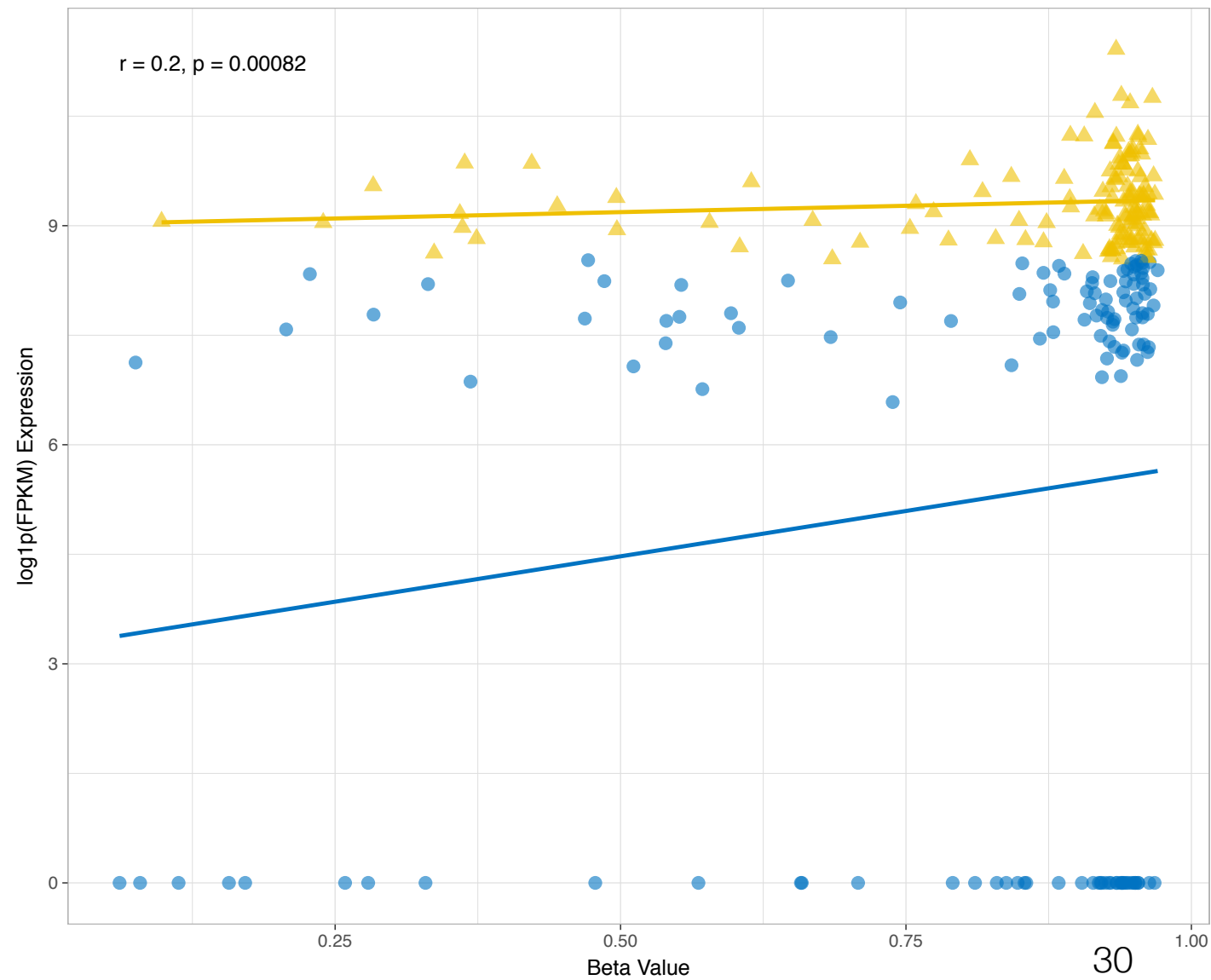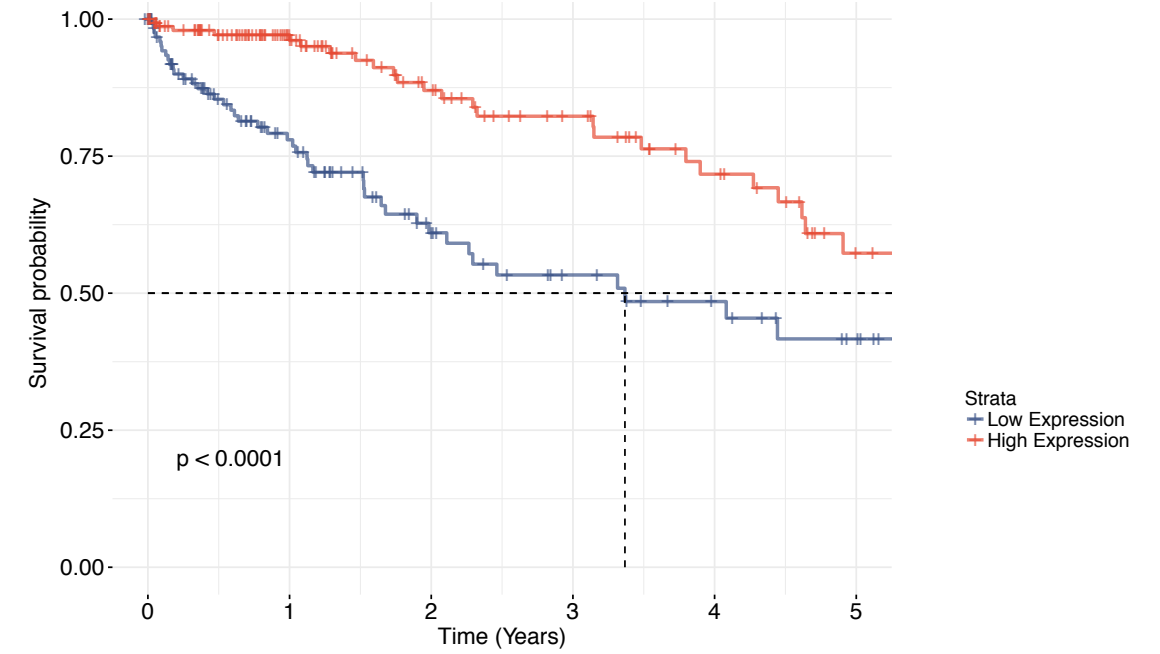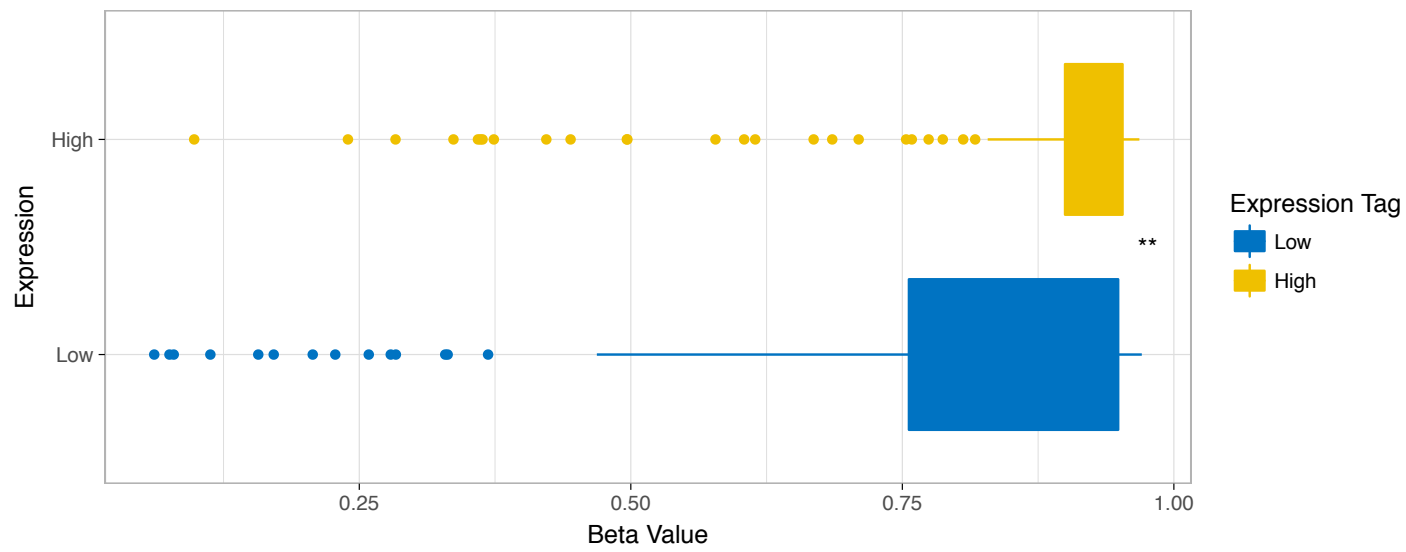
Ovarian serous cystadenocarcinoma
n = 317 patients in TCGA
n = 84 patients in PCAWG

AP001057.1 in OV

AP001057.1 HR = 0.5711 Ovarian serous cystadenocarcinoma
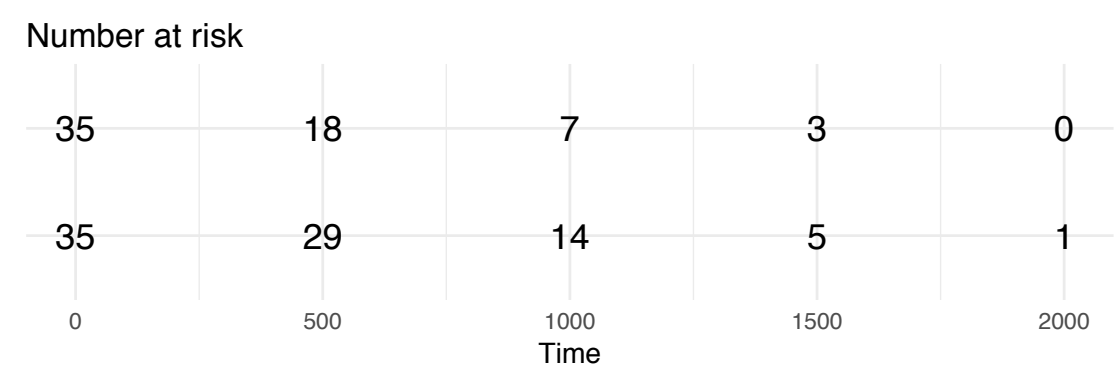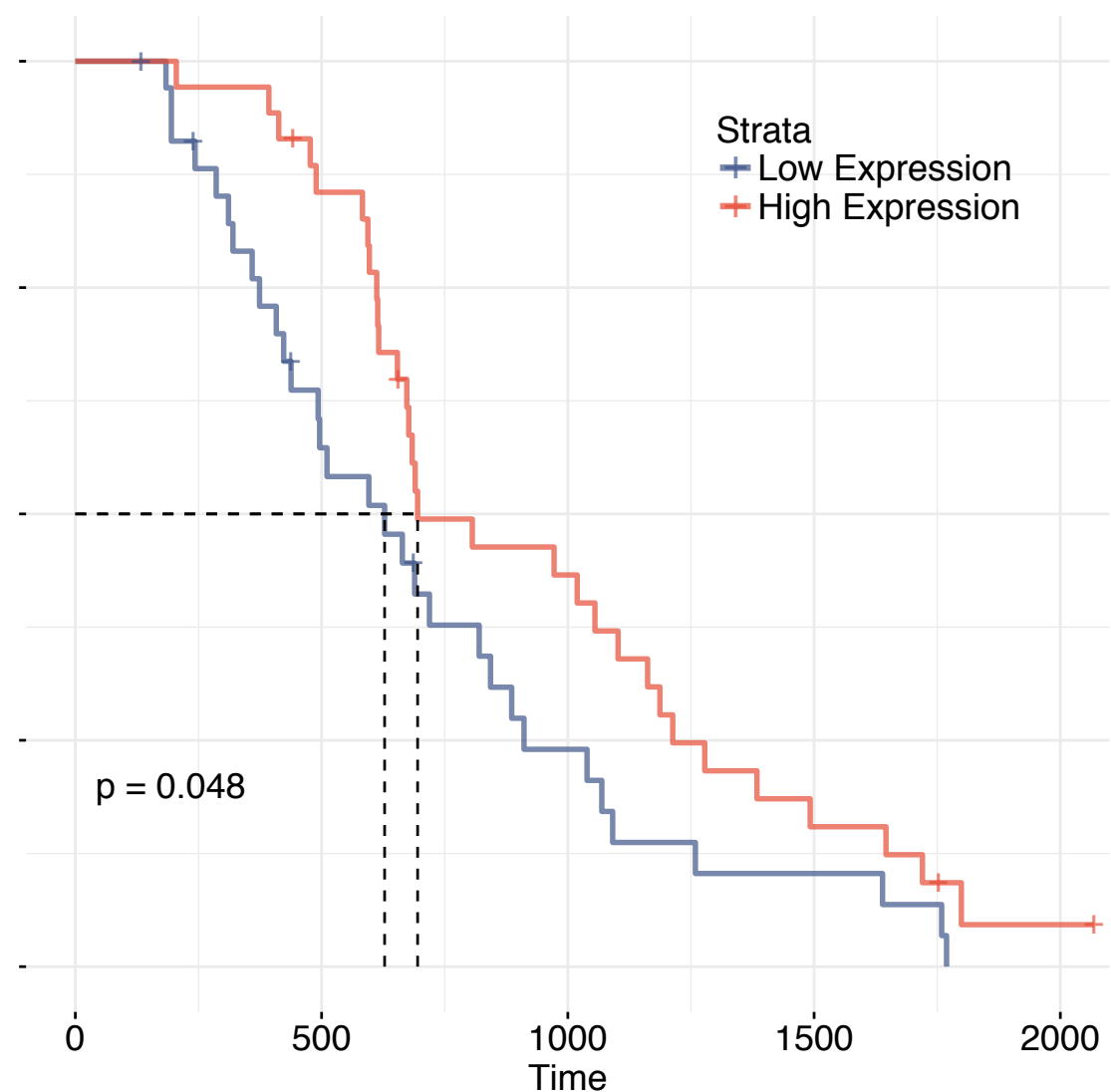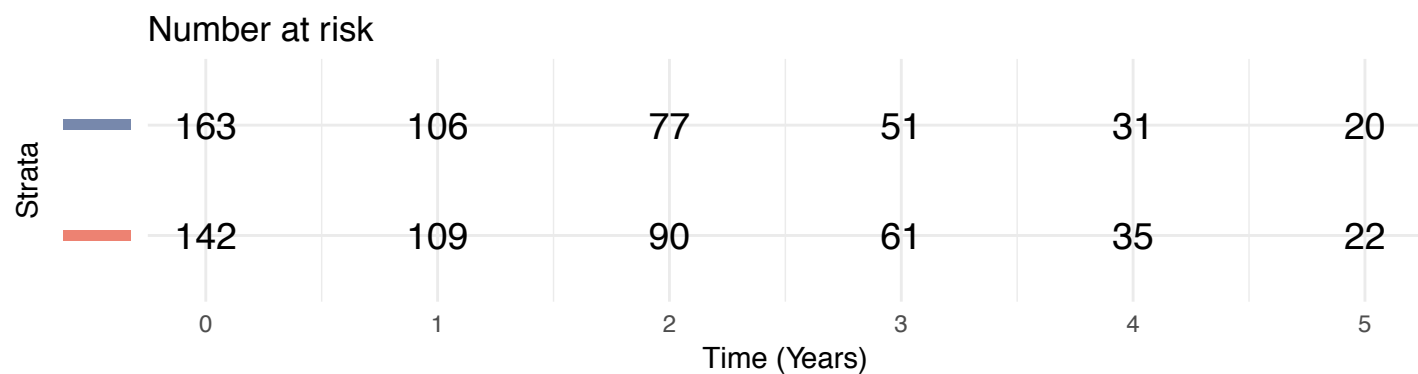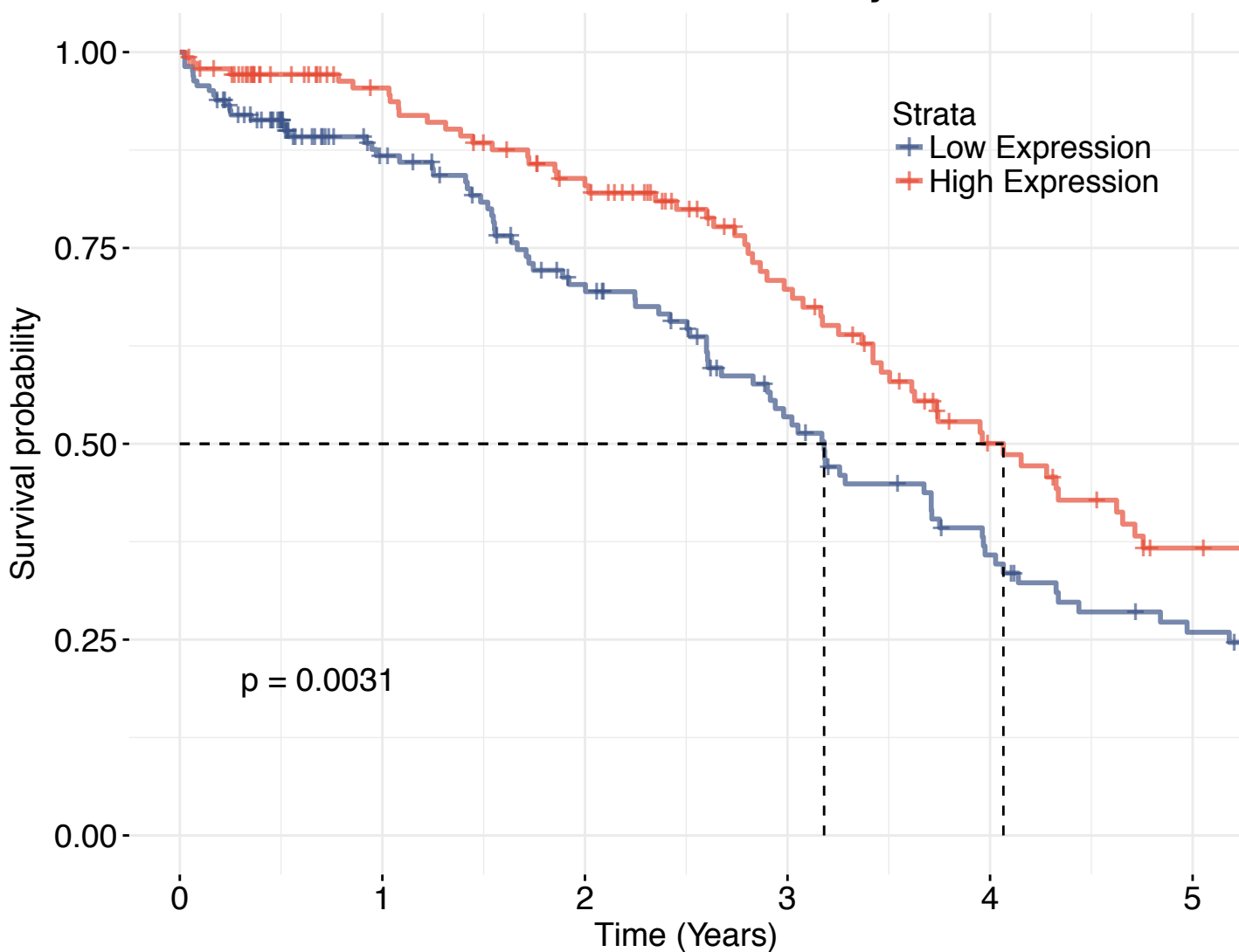
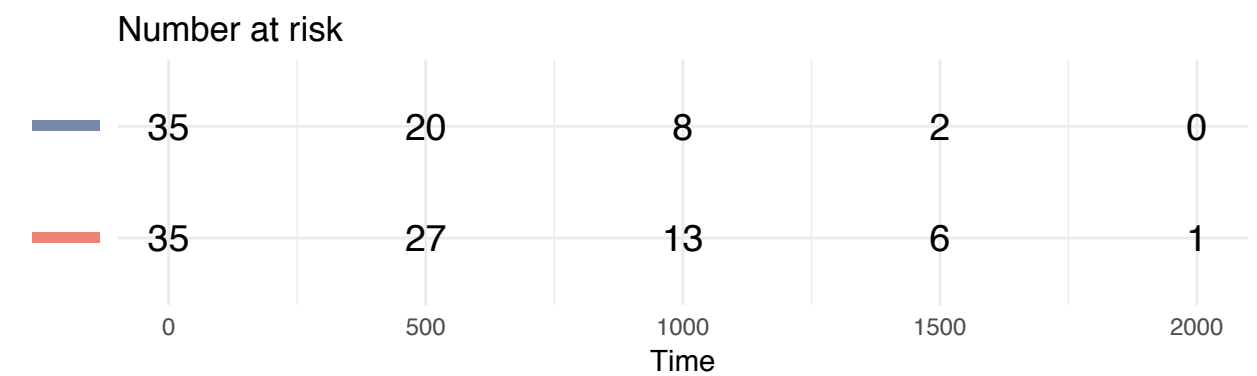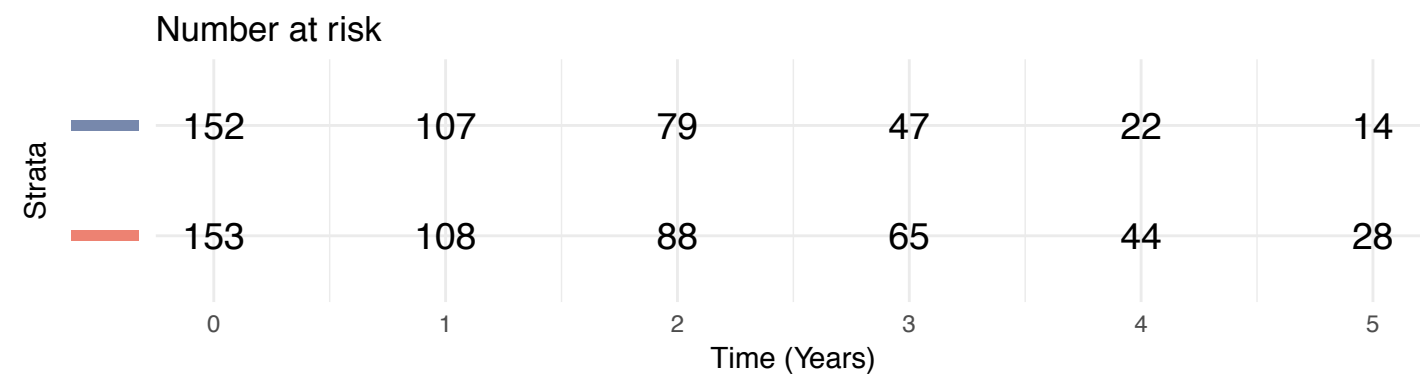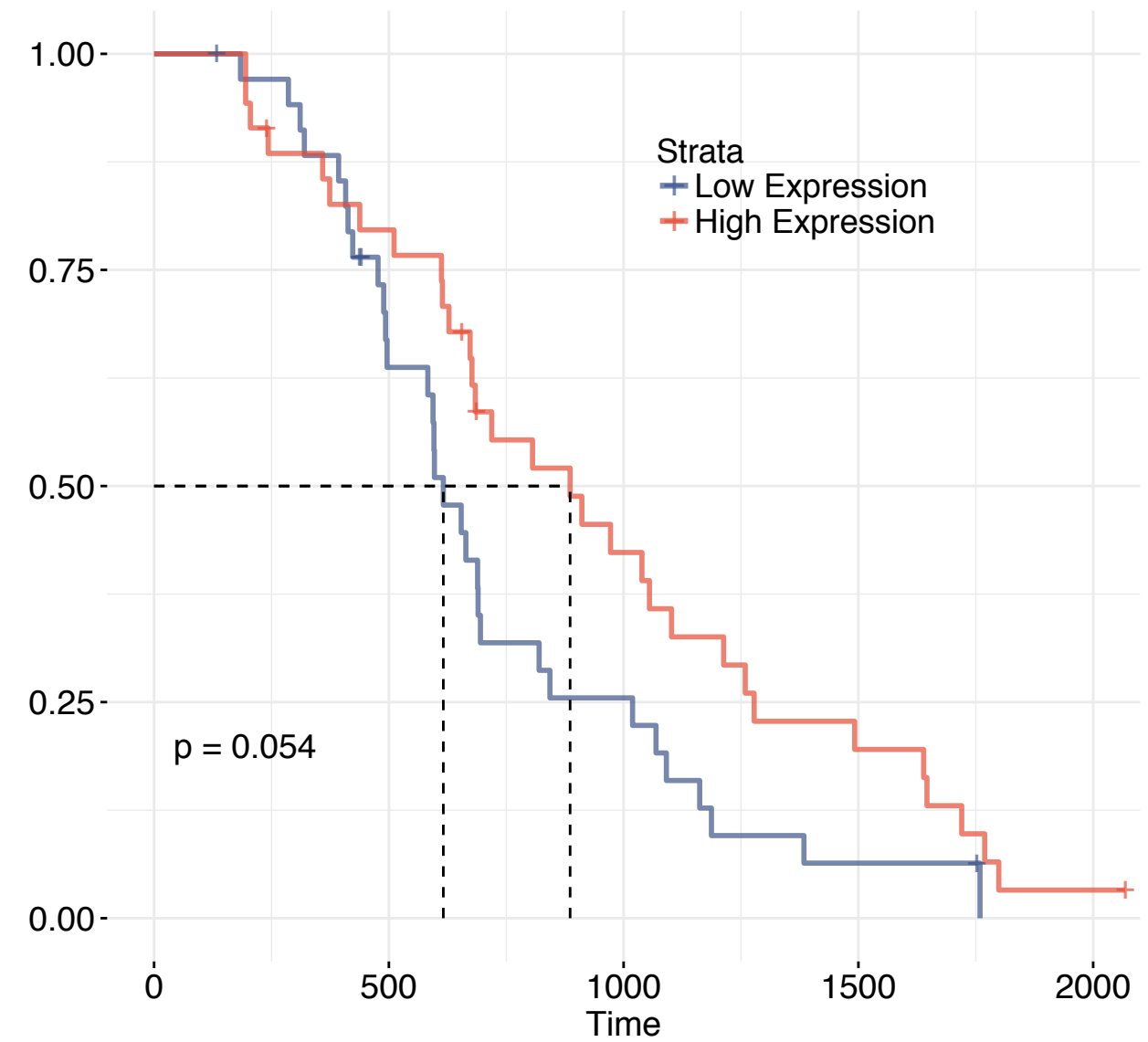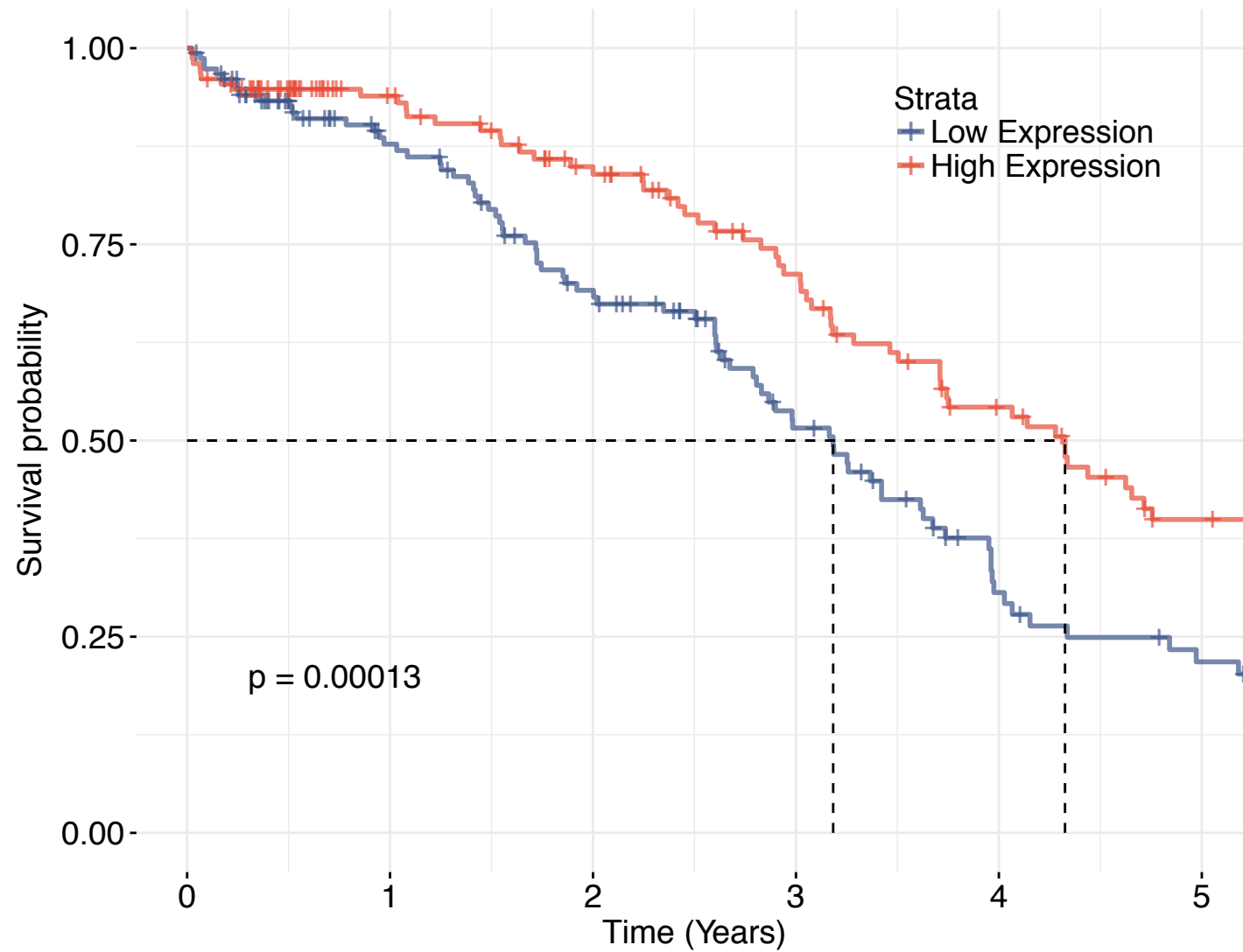No Differentially expressed genes between the two groups

32

AP001057.1 in OV

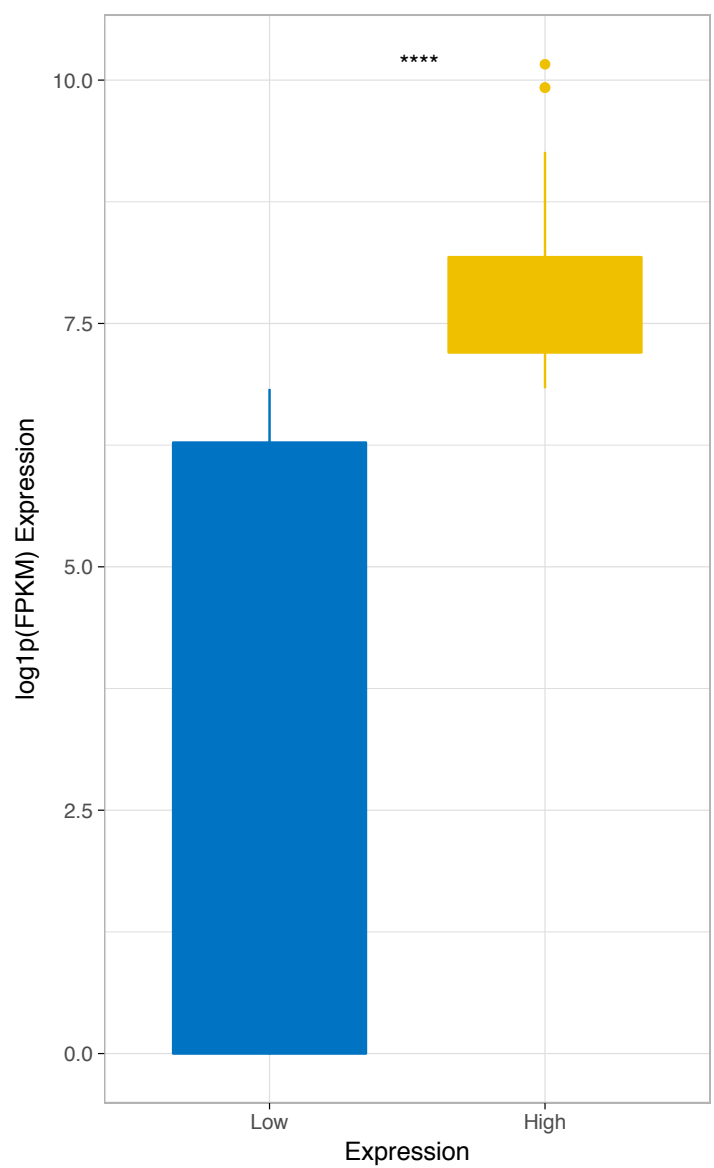AP001057.1 ovary CNA vs Exp n= 290
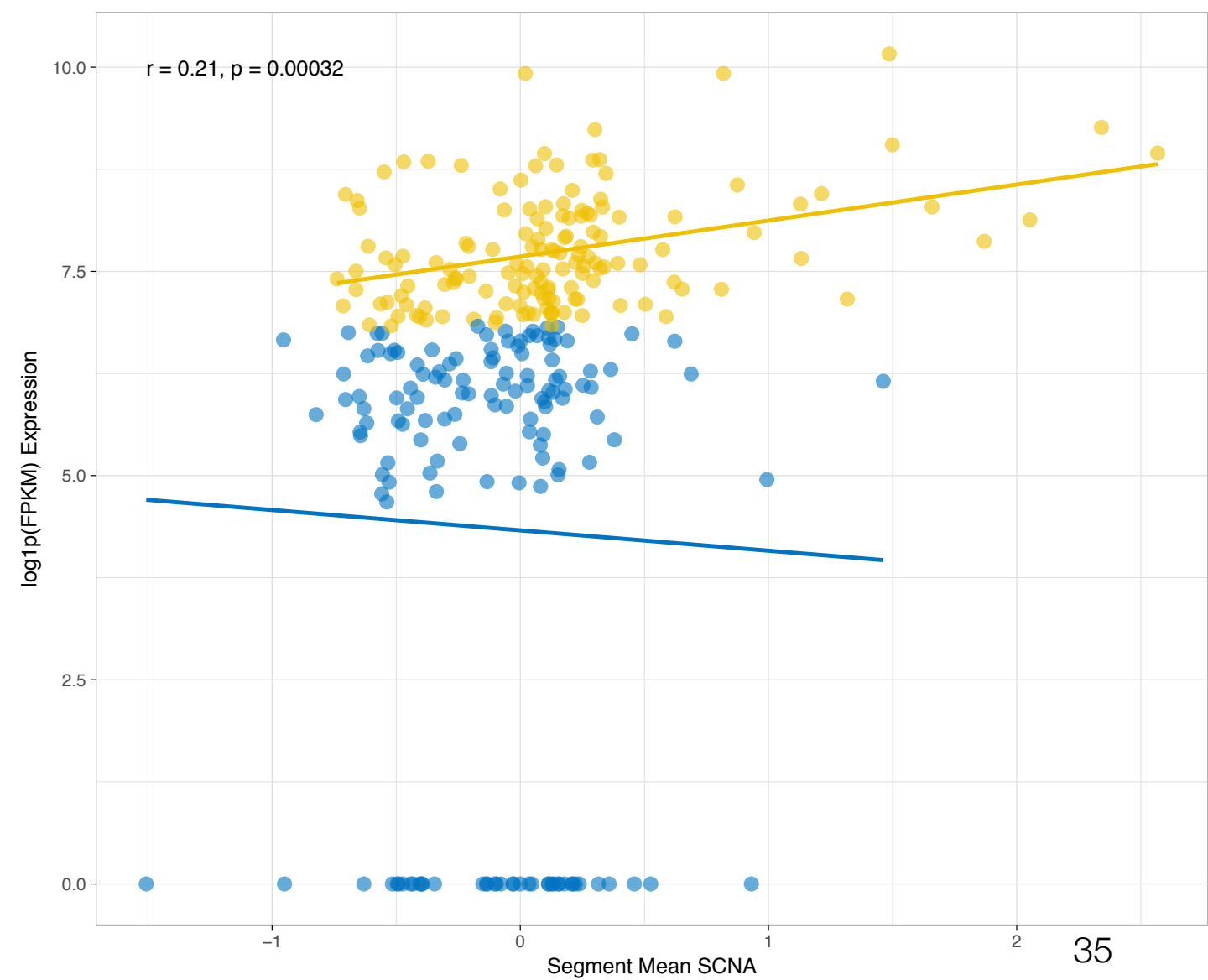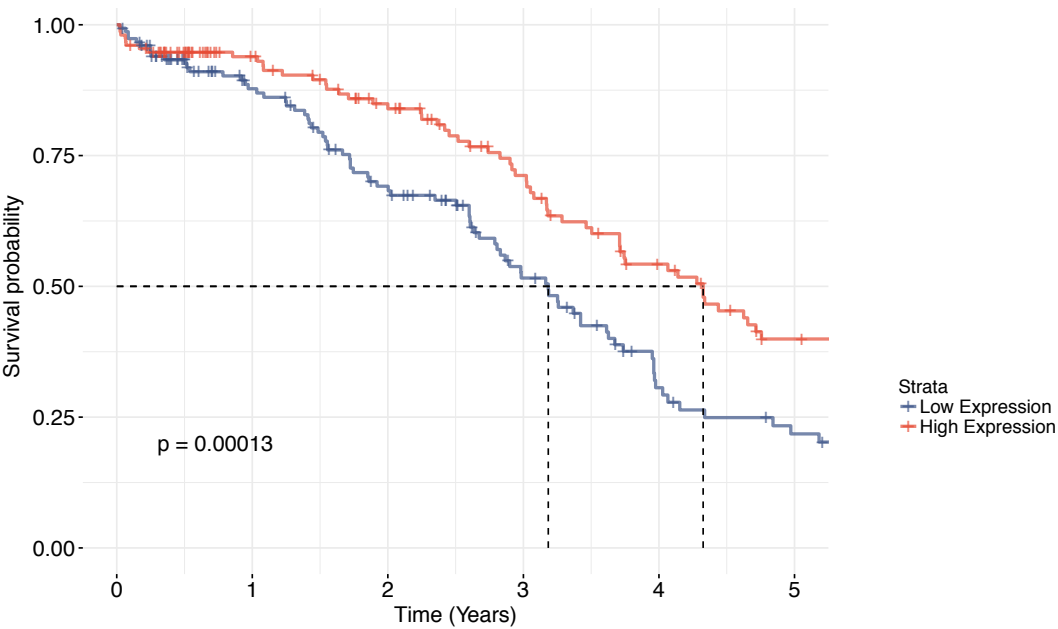
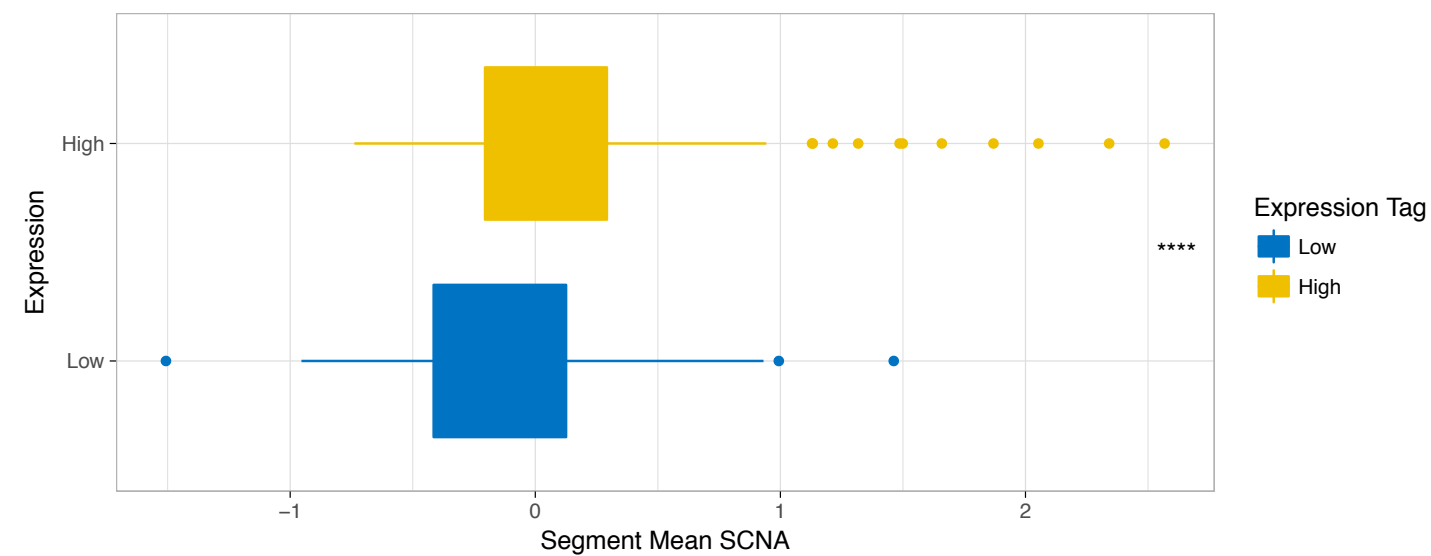r = 0.18, p = 0.0021

p = 0.0031

****
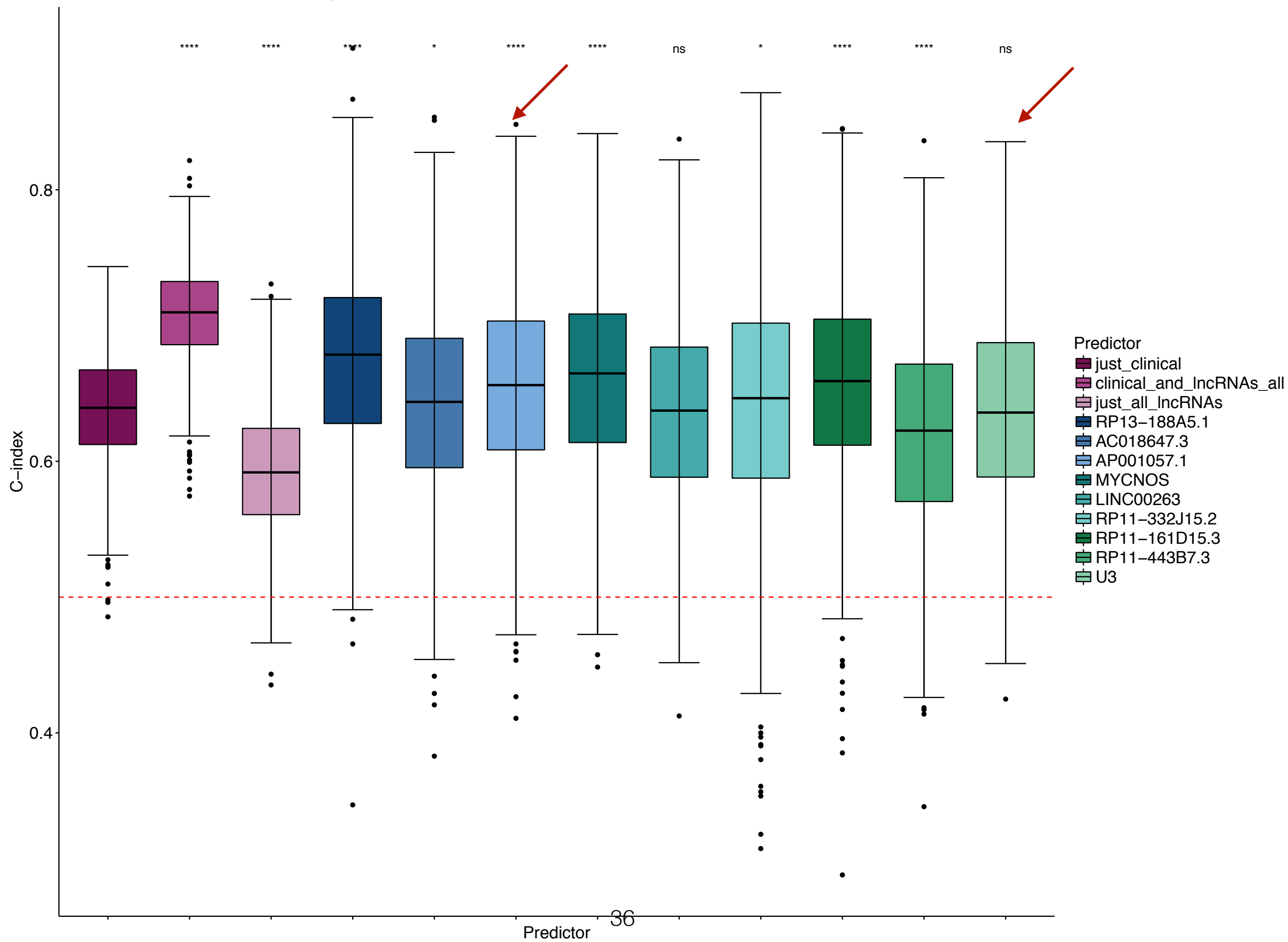
33
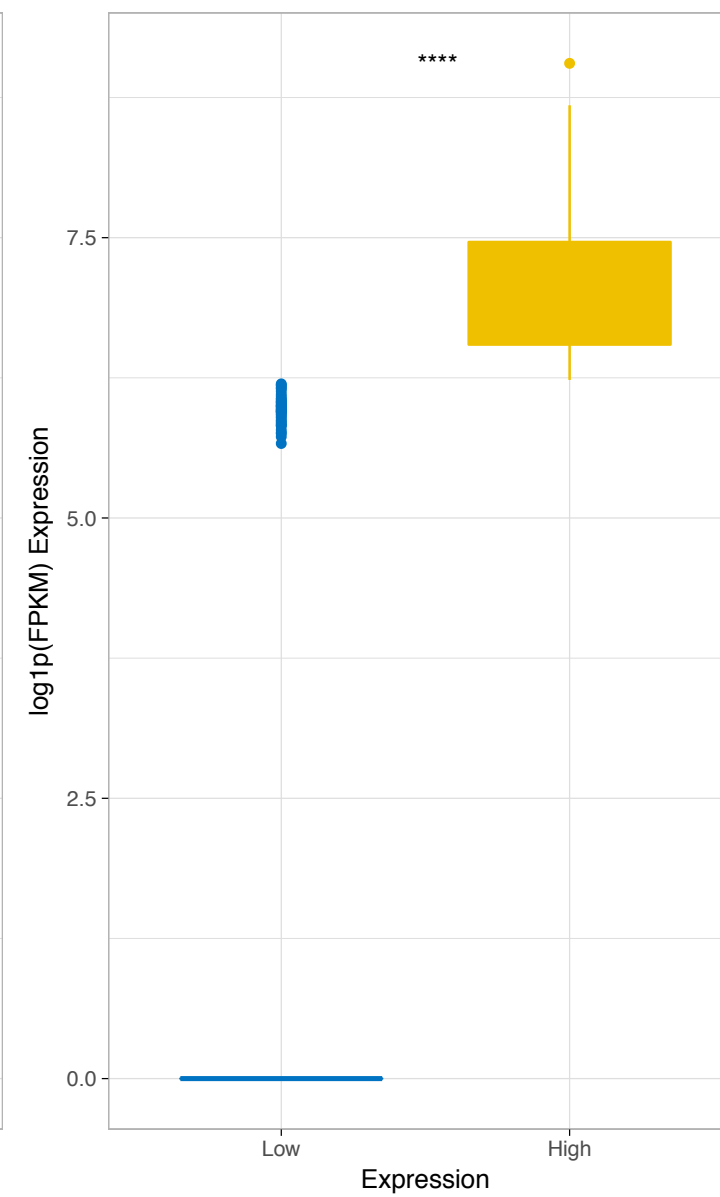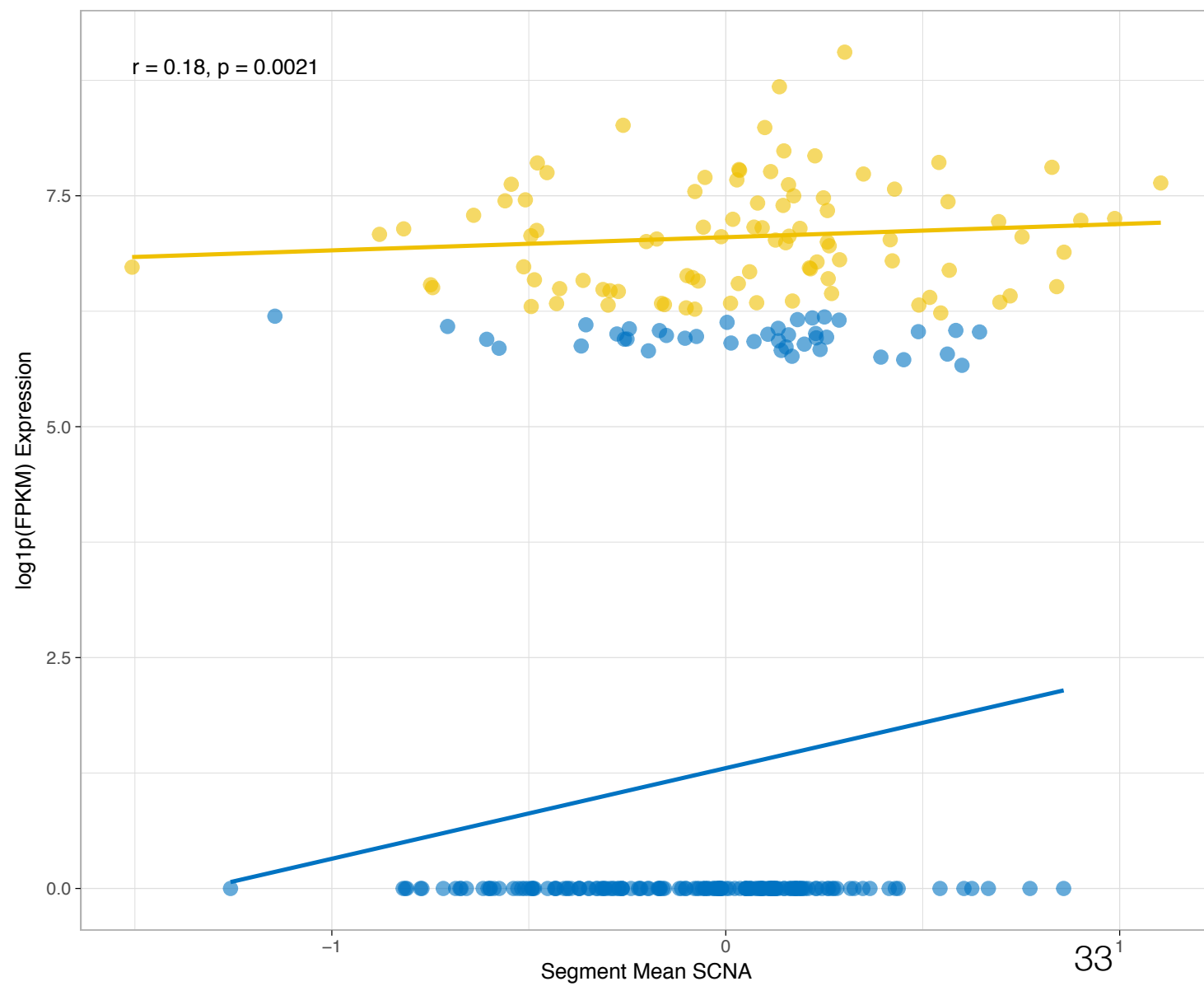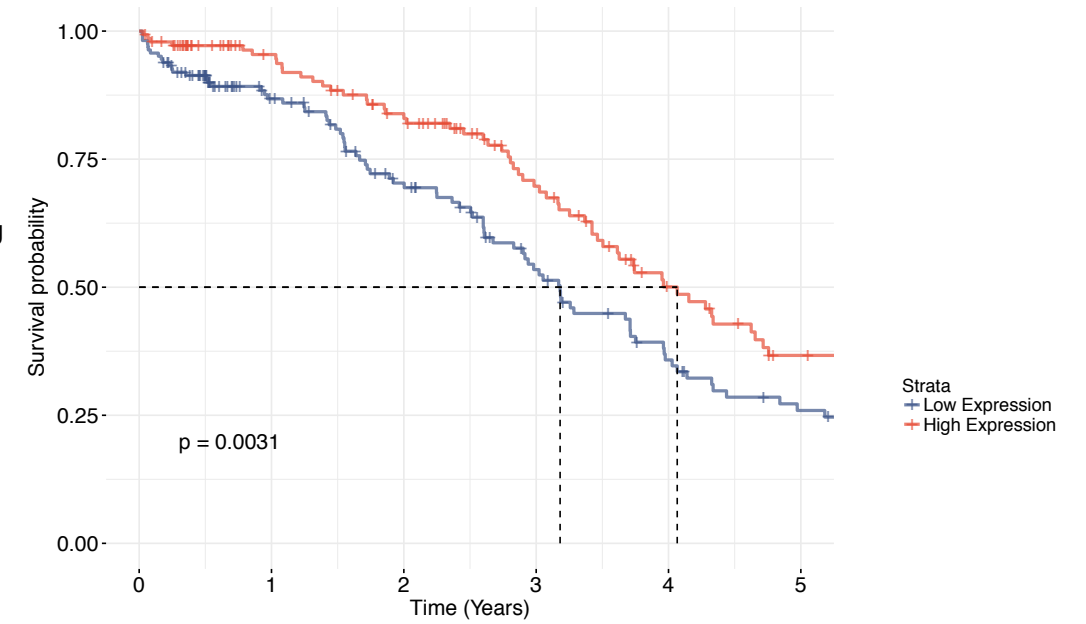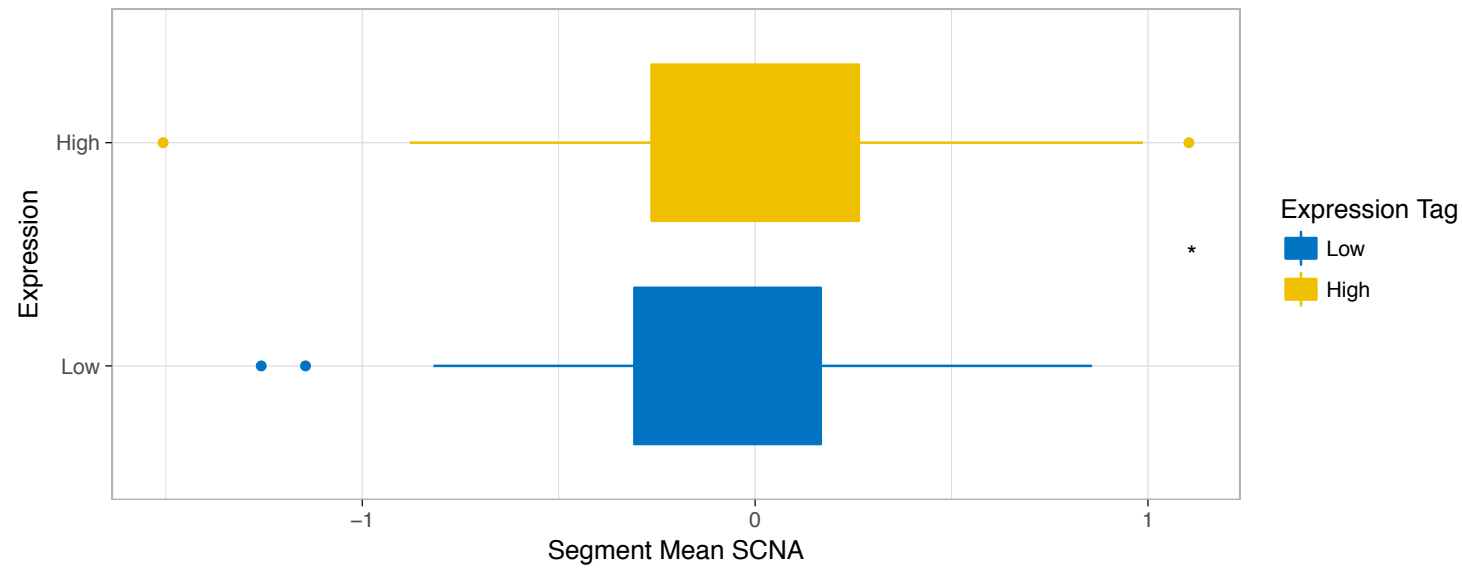
U3 in OV

U3 HR = 0.5372 Ovarian serous cystadenocarcinoma

U3 in OV

U3 ovary CNA vs Exp n= 291

Expression Tag
- Low
- High

****

r = 0.21, p = 0.00032

p = 0.00013
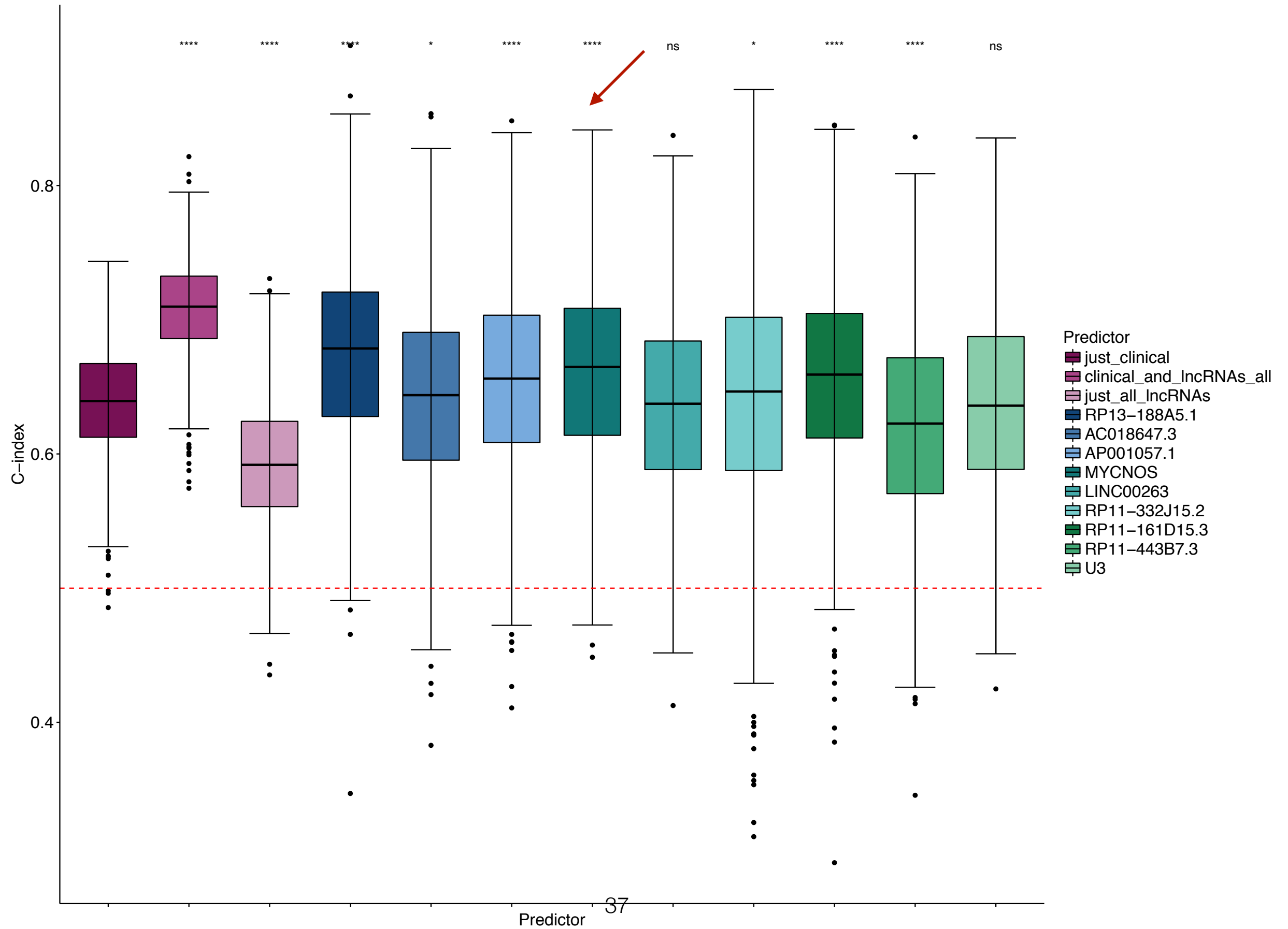
Strata
- Low Expression
- High Expression

****

35

AP001057.1 & U3 as predictors of survival in OV

305 Ovarian Cancer Patients, 1000 Cross Validations

C-index

Predictor

Predictor
- just_clinical
- clinical_and_lncRNAs_all
- just_all_lncRNAs
- RP13−188A5.1
- AC018647.3
- AP001057.1
- MYCNOS
- LINC00263
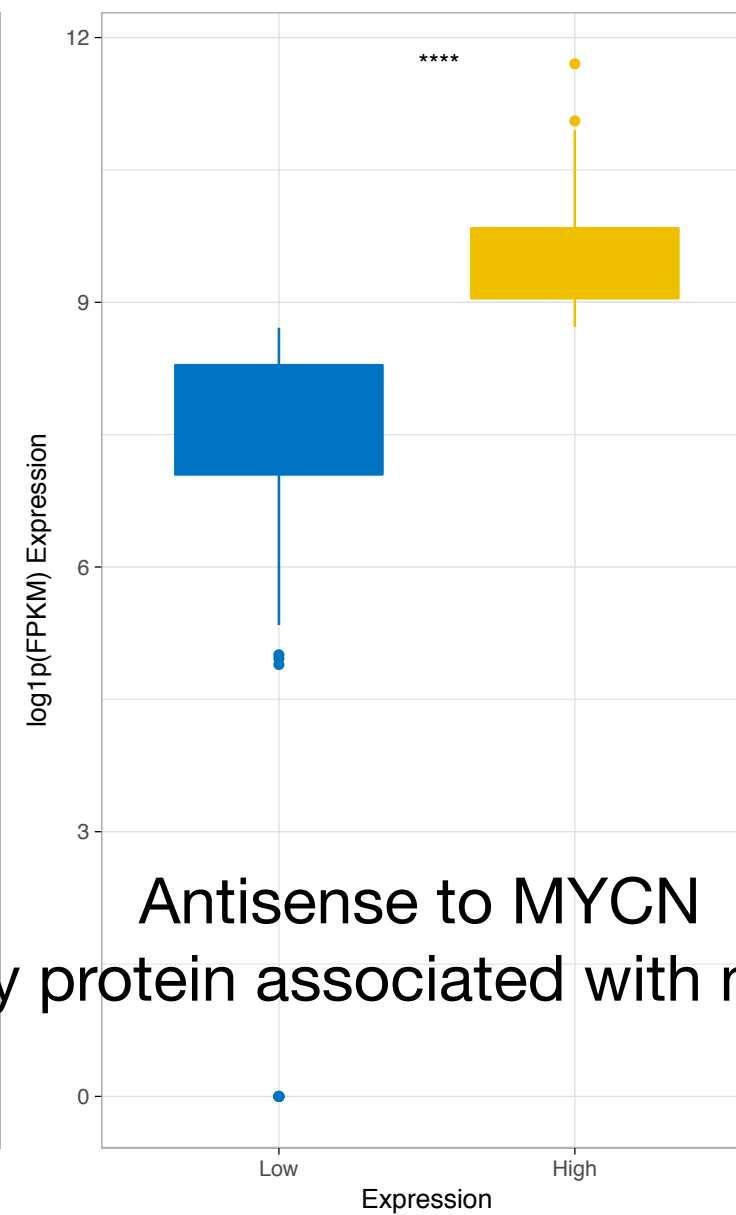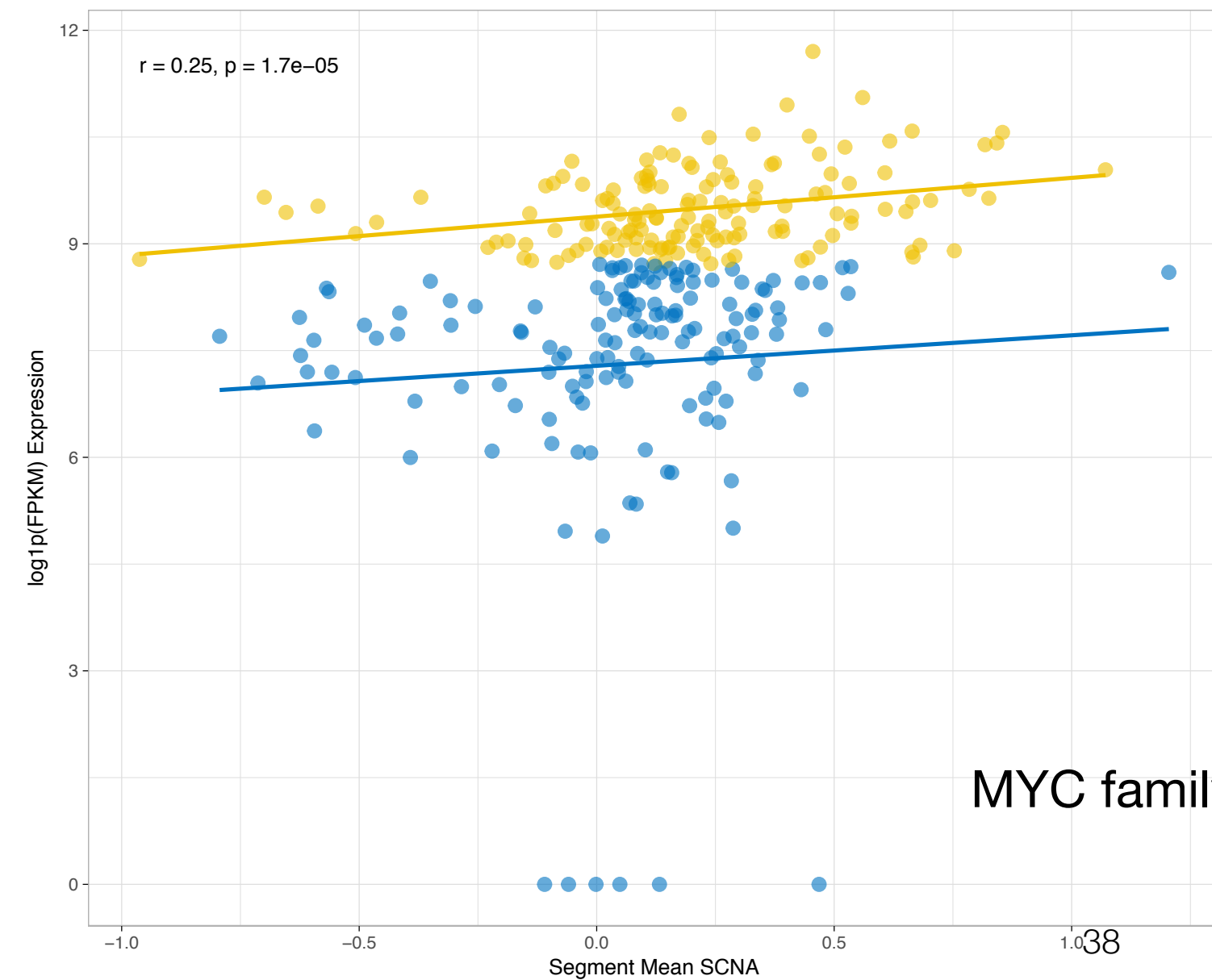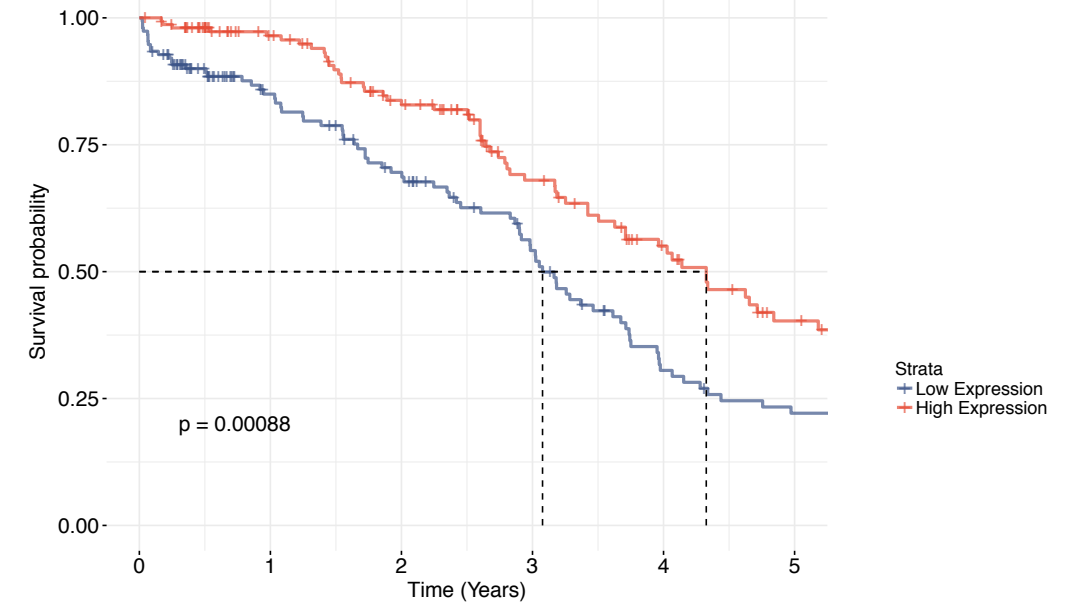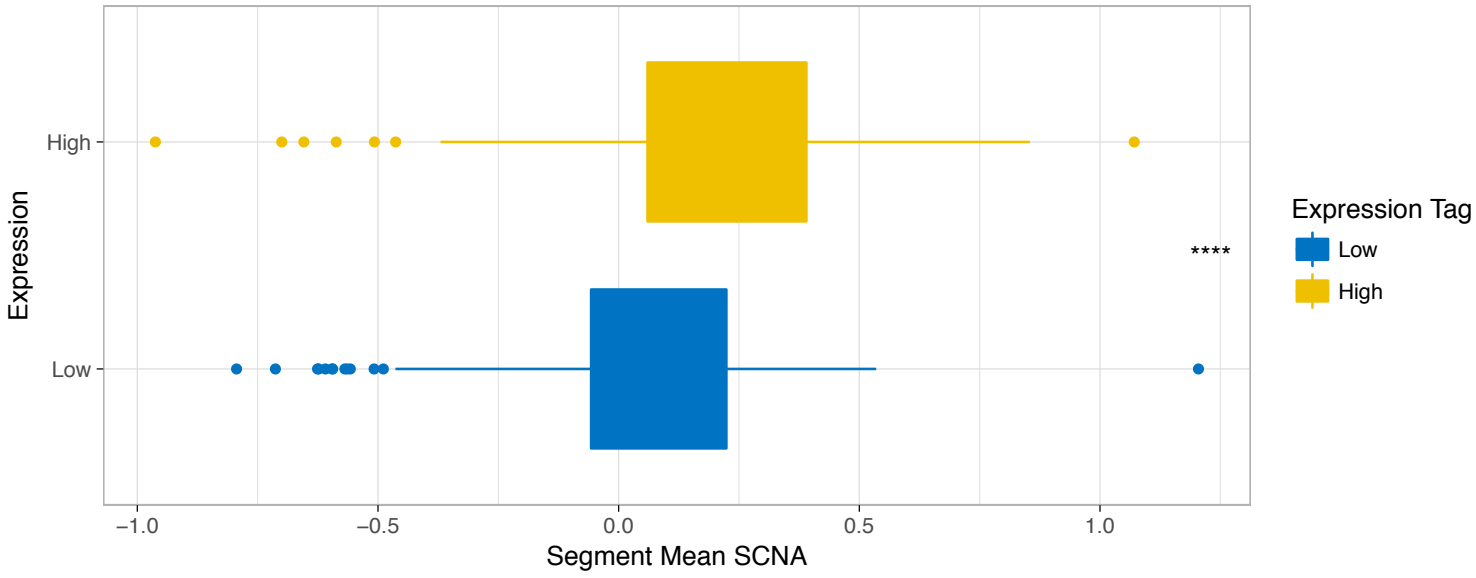- RP11−332J15.2
- RP11−161D15.3
- RP11−443B7.3
- U3

36

MYCNOS in OV

305 Ovarian Cancer Patients, 1000 Cross Validations

MYCNOS in OV

MYCNOS ovary CNA vs Exp n= 291

r = 0.25, p = 1.7e−05

p = 0.00088

Expression Tag
Low
High

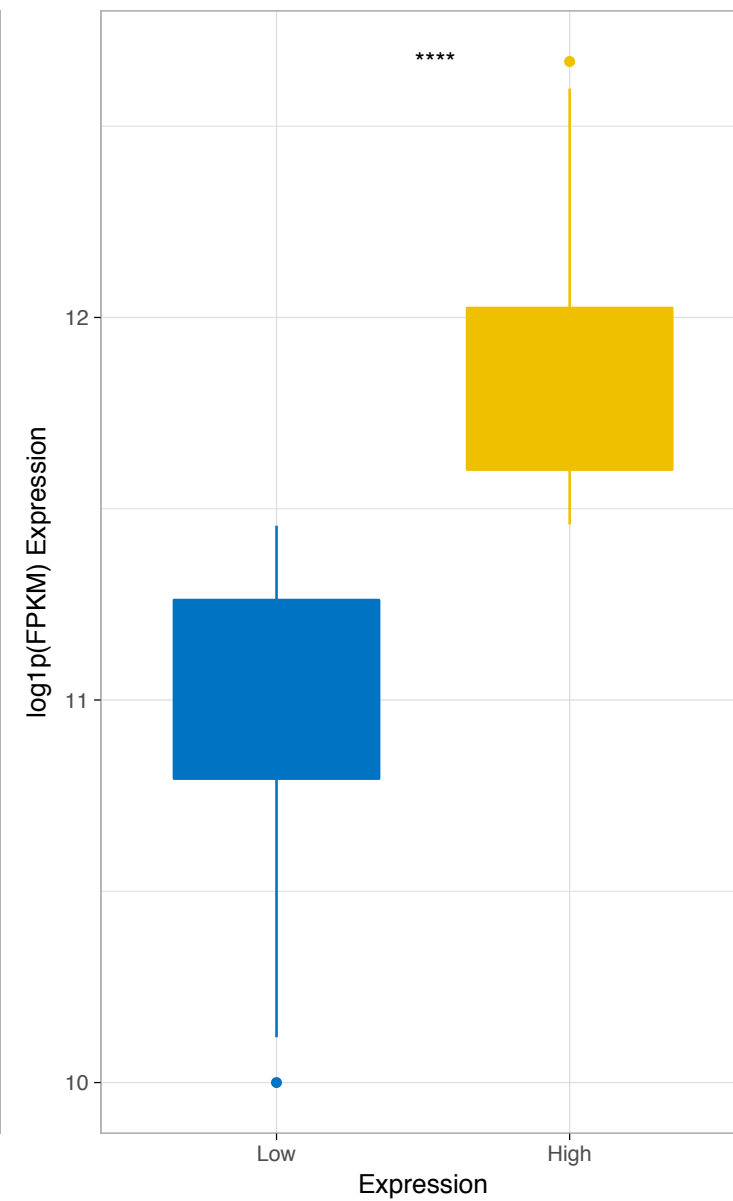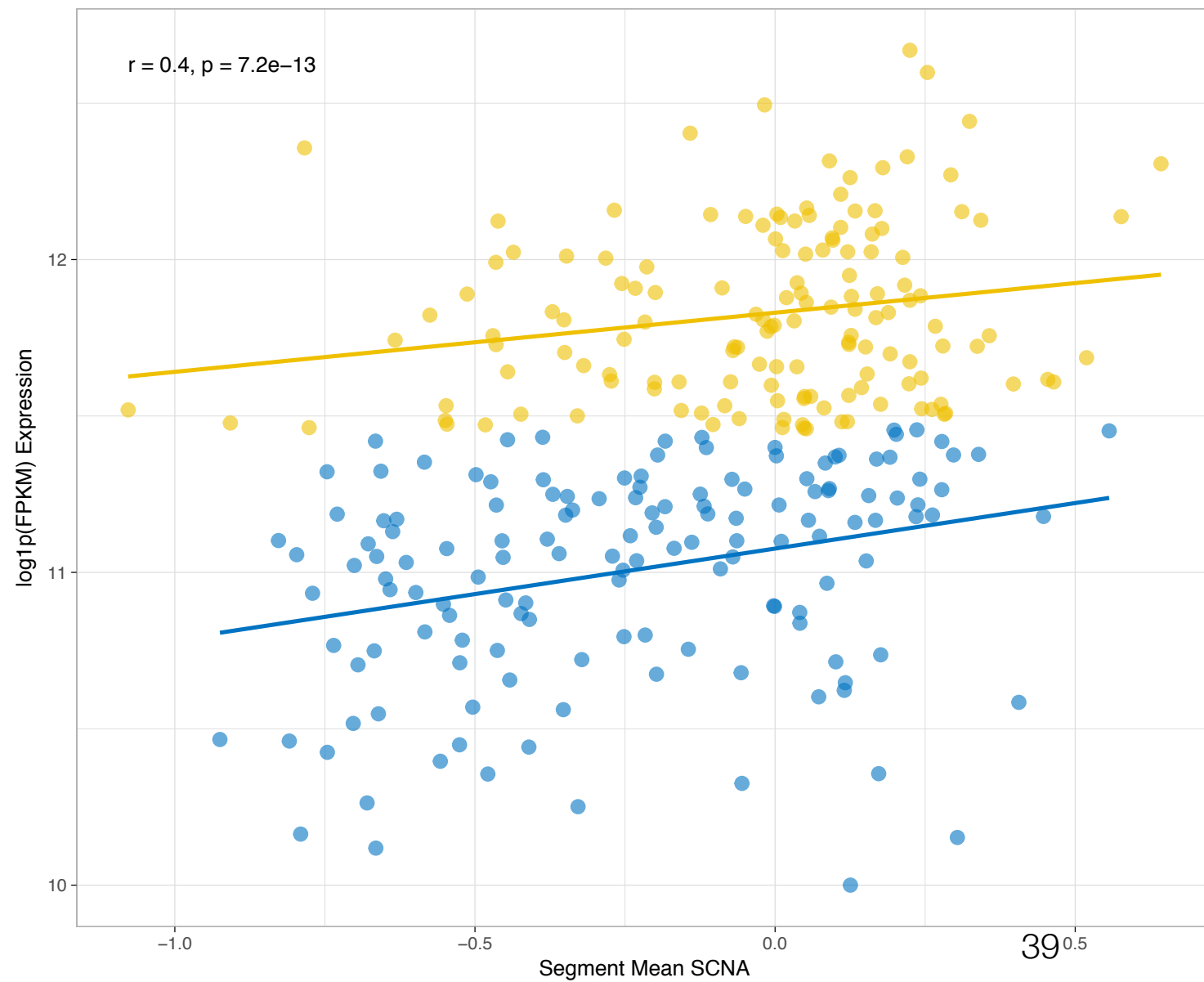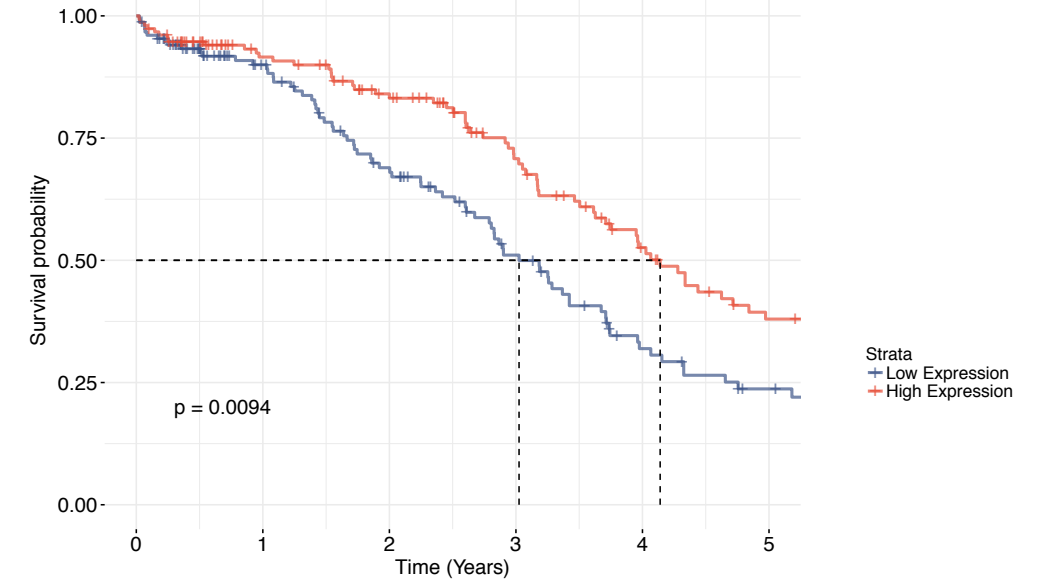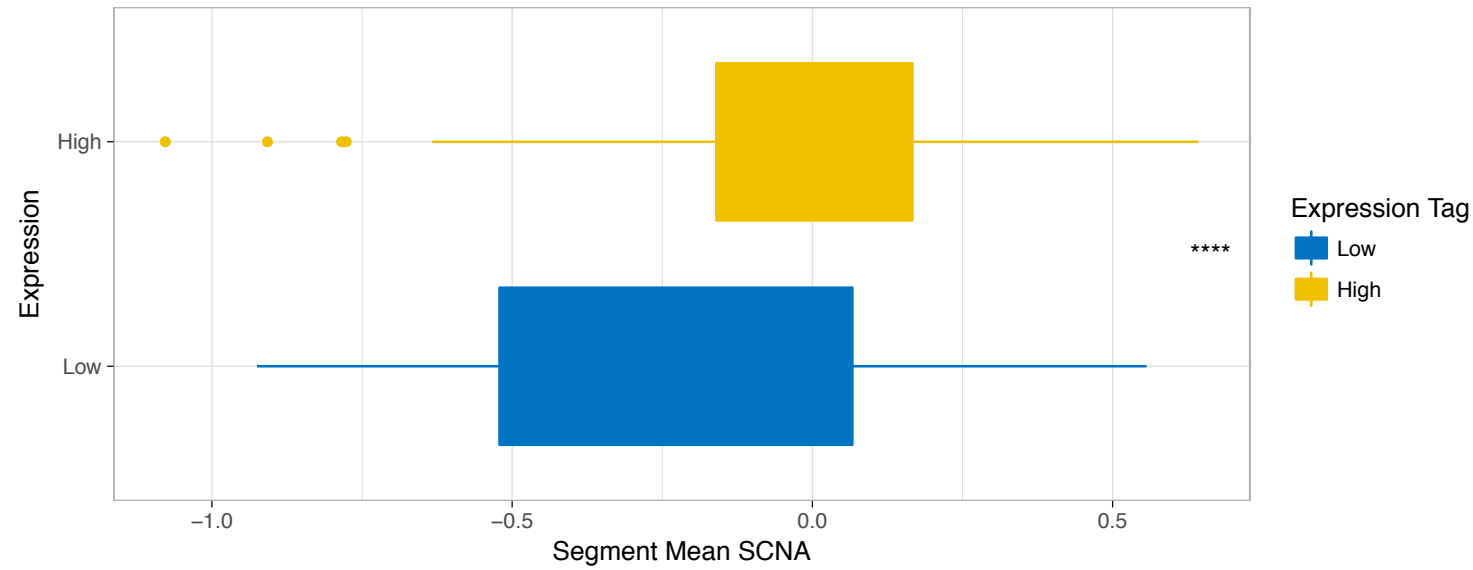Strata
Low Expression
High Expression

****

Antisense to MYCN
MYC family protein associated with neuroblastoma

38

LINC00263 in OV

Summary

- Identified candidates in 4 cancer types where at least one validated in an independent dataset

- Identified several candidates that potentially have a genetic/epigenetic molecular profile associated with their expression

Currently:
- lncRNA ~ miRNA co-expression
- lncRNA co-expression with distal genes connected through loops