

Thesis: Genomic characterization of clinically relevant lncRNAs in multiple cancer types

Second Committee Meeting

Karin Isaev

Supervisor: Dr. Juri Reimand

September 19th, 2017

Thesis: Genomic characterization of clinically relevant lncRNAs in multiple cancer types

1. Review of lncRNAs
2. Summary of the last committee meeting
3. Summary of progress since the last committee meeting
4. Future directions

Thesis: Genomic characterization of clinically relevant lncRNAs in multiple cancer types

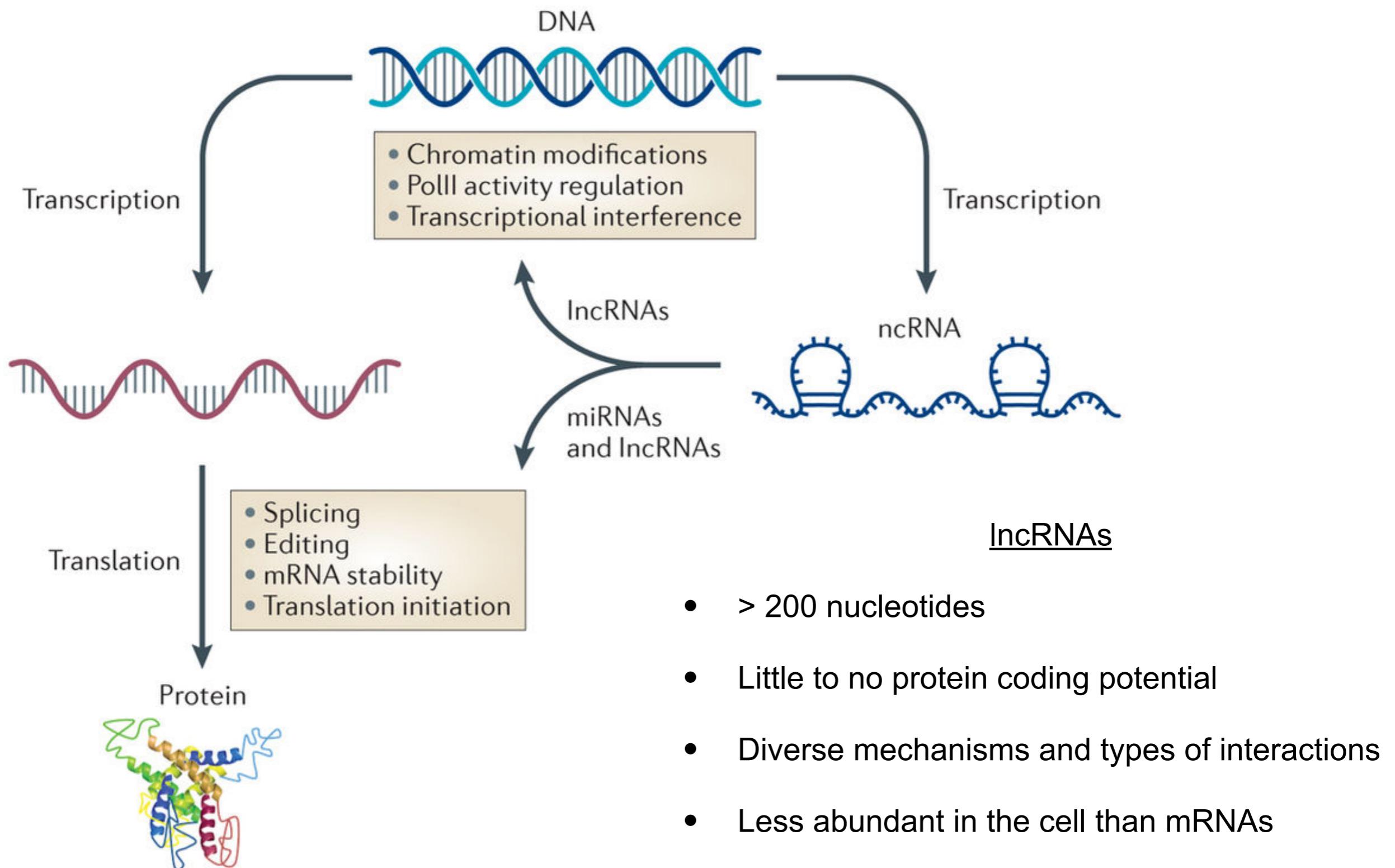
1. Review of lncRNAs

2. Summary of the last committee meeting

3. Summary of progress since the last committee meeting

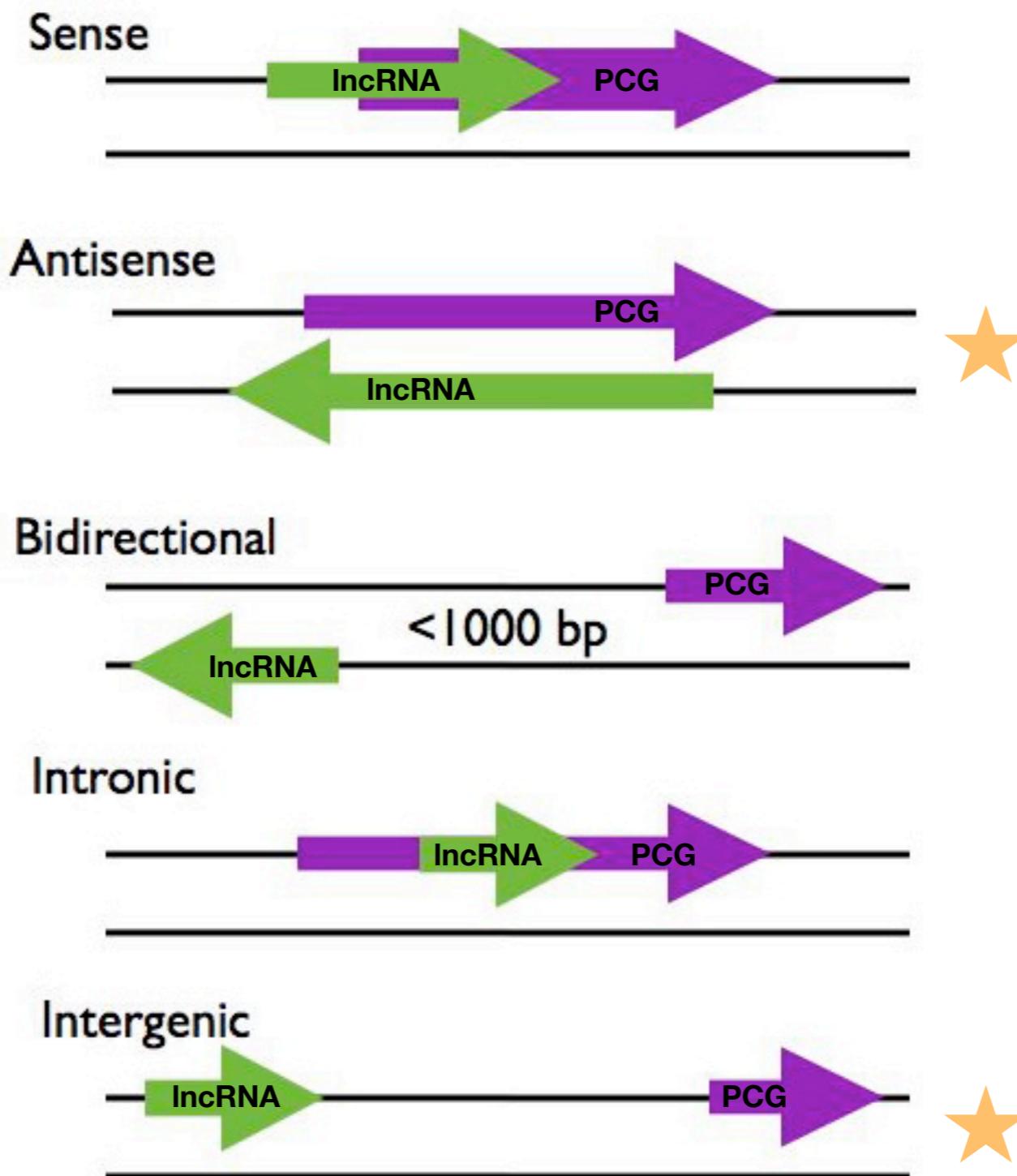
4. Future directions

Central Dogma in the context of lncRNAs



lncRNAs are grouped based on proximity to protein coding genes

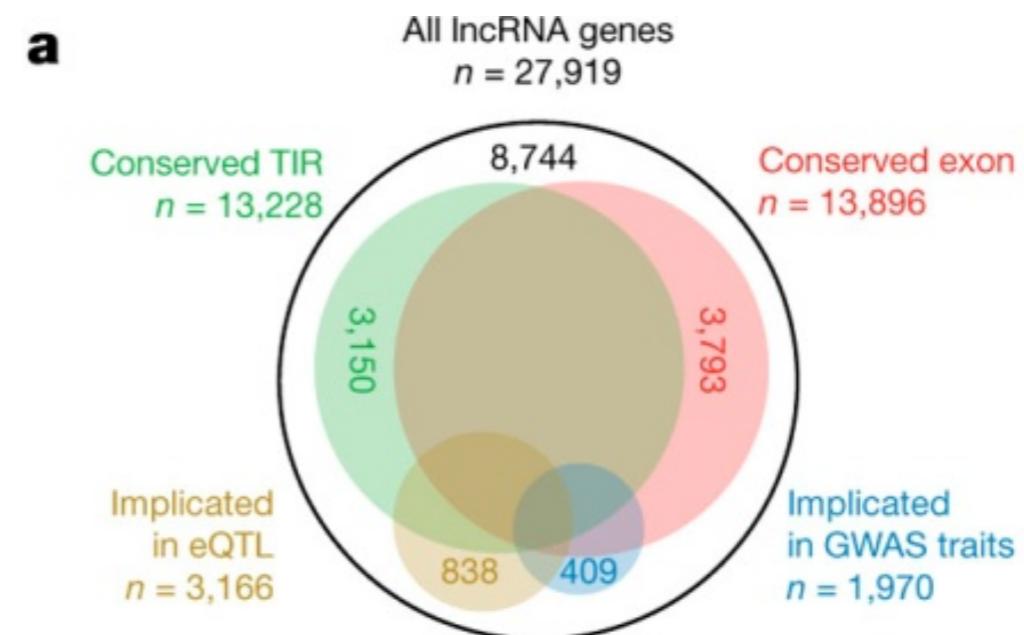
* Protein Coding Gene (PCG)



Functional evidence of human lncRNAs

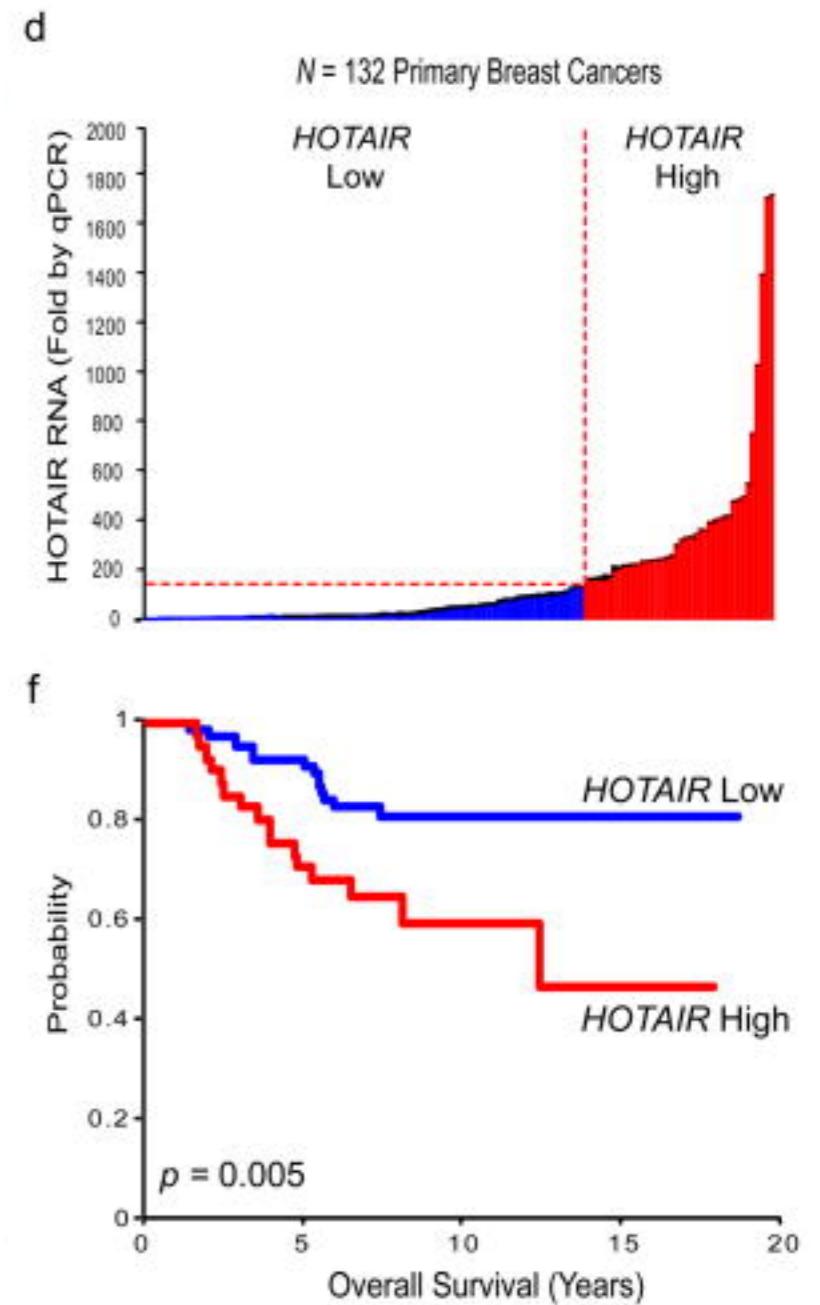
FANTOM CAT contains ~20,000 potentially functional lncRNAs with annotated 5'ends and epigenome evidence

- Used Cap Analysis of Gene Expression (CAGE)
- lncRNAs overlapping eQTL-associated SNPs of mRNAs are co-expressed
 - Potential role as transcriptional regulators



Rationale for studying lncRNAs in Cancer

- lncRNA tissue specificity greater than mRNAs
- Previously annotated prognostic lncRNAs:
 - **HOTAIR** in Breast cancer (metastasis)
 - **FAL1** in Ovarian cancer (poor prognosis)
 - Genome wide survey of lncRNA somatic copy number alterations
- Interact with DNA, RNA and protein
 - Attractive targets for therapeutic intervention



Systematic analysis of lncRNAs as predictors of survival across multiple cancers remains to be done

Thesis: Genomic characterization of clinically relevant lncRNAs in multiple cancer types

1. Review of lncRNAs

2. Summary of the last committee meeting

3. Summary of progress since the last committee meeting

4. Future directions

Summary of the last committee meeting

Hypothesis: lncRNAs are clinically informative and regulate cancer genes and pathways

Aim 1:

Identify lncRNA molecular profiles associated with survival in multiple cancer types



Aim 2:

Characterize lncRNA roles through integration of -omics data and association to cancer pathways

Summary of the last committee meeting

Summary of the last committee meeting

Aim 1

Summary of the last committee meeting

Aim 1

- Using the PCAWG dataset, lncRNA candidates will be identified in multiple cancer types through associating their gene expression with patient survival.

Summary of the last committee meeting

Aim 1

- Using the PCAWG dataset, lncRNA candidates will be identified in multiple cancer types through associating their gene expression with patient survival.

Aim 2

Summary of the last committee meeting

Aim 1

- Using the PCAWG dataset, lncRNA candidates will be identified in multiple cancer types through associating their gene expression with patient survival.

Aim 2

- Through differential expression, co-expression and pathway enrichment analysis, lncRNA candidates will be associated with cellular roles.

Summary of the last committee meeting

Aim 1

- Using the PCAWG dataset, lncRNA candidates will be identified in multiple cancer types through associating their gene expression with patient survival.

Aim 2

- Through differential expression, co-expression and pathway enrichment analysis, lncRNA candidates will be associated with cellular roles.

Preliminary analysis:

Summary of the last committee meeting

Aim 1

- Using the PCAWG dataset, lncRNA candidates will be identified in multiple cancer types through associating their gene expression with patient survival.

Aim 2

- Through differential expression, co-expression and pathway enrichment analysis, lncRNA candidates will be associated with cellular roles.

Preliminary analysis:

- Identified *PVT1* as prognostic in kidney cancer
 - 2016 Oncotarget paper identified *PVT1* hypomethylation was associated with increased expression and survival correlation in Clear Cell Renal Carcinoma

Summary of the last committee meeting

Aim 1

- Using the PCAWG dataset, lncRNA candidates will be identified in multiple cancer types through associating their gene expression with patient survival.

Aim 2

- Through differential expression, co-expression and pathway enrichment analysis, lncRNA candidates will be associated with cellular roles.

Preliminary analysis:

- Identified *PVT1* as prognostic in kidney cancer
 - 2016 Oncotarget paper identified *PVT1* hypomethylation was associated with increased expression and survival correlation in Clear Cell Renal Carcinoma
- Used to establish method that can be applied systematically, to all cancer types in dataset

Summary of the last committee meeting

Notes:

- Improve cancer specific filtration of candidate lncRNAs
- Identify methods to overcome low abundance and possible nearby effect of a protein coding gene
- Integrate FANTOM5 CAT lncRNA annotations

Thesis: Genomic characterization of clinically relevant lncRNAs in multiple cancer types

1. Review of lncRNAs
2. Summary of the last committee meeting
3. Summary of progress since the last committee meeting
4. Future directions

Thesis: Genomic characterization of clinically relevant lncRNAs in multiple cancer types

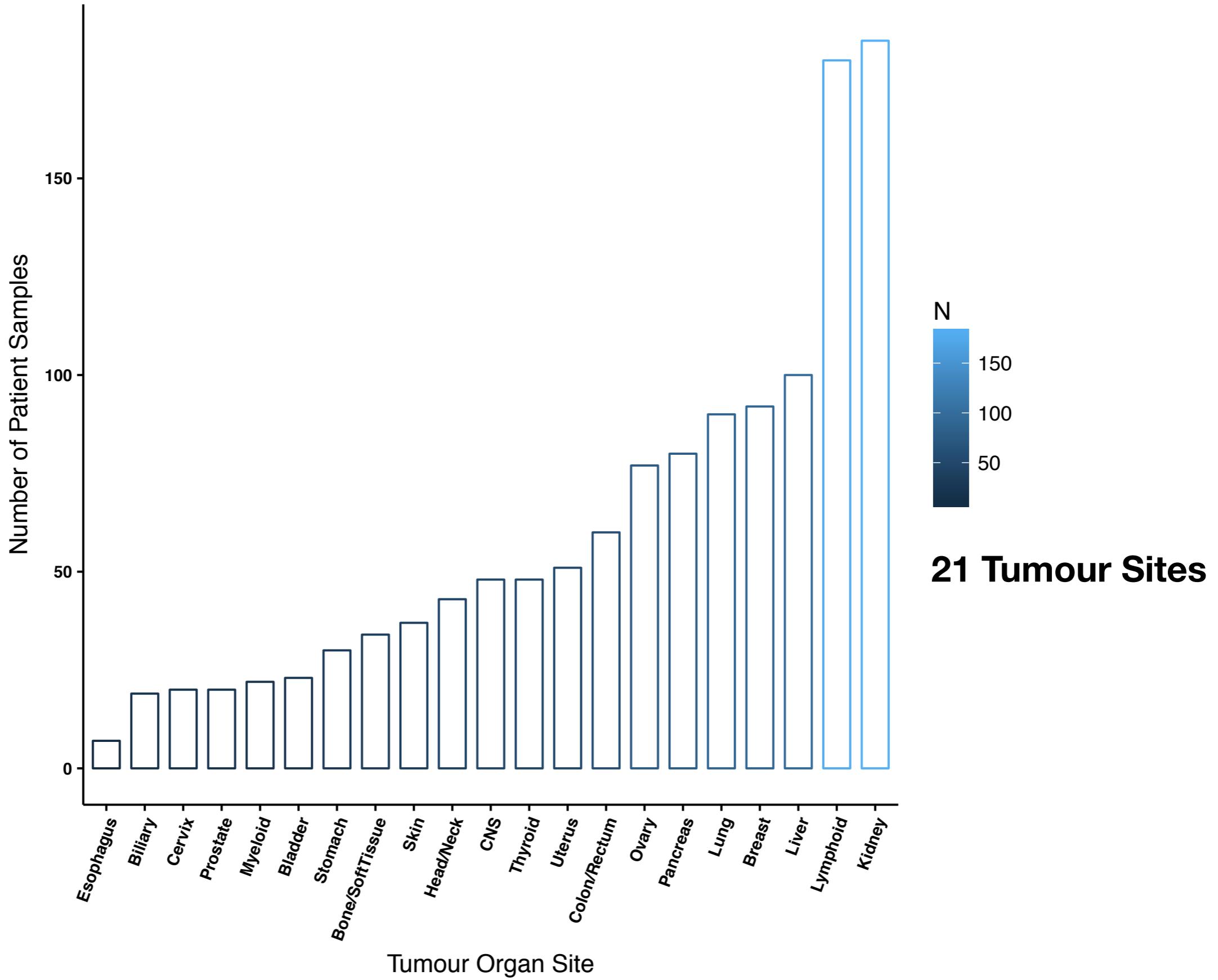
1. Review of lncRNAs
2. Summary of the last committee meeting
- 3. Summary of progress since the last committee meeting**
4. Future directions

Summary of progress

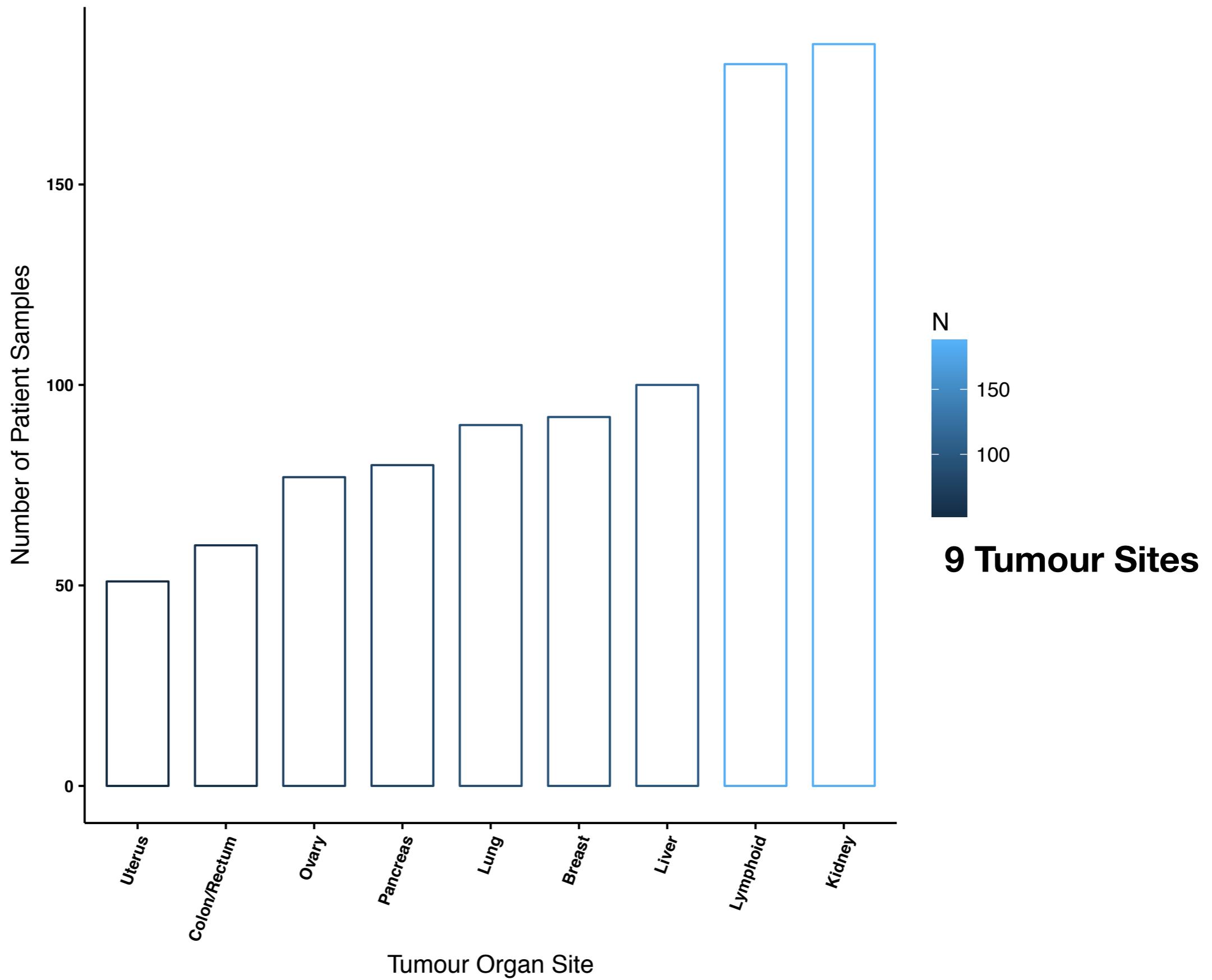
1. Identified high expressing lncRNA candidates in multiple cancer types
2. Survival analysis of lncRNAs using the Cox Proportional Hazards model
3. Differential expression analysis of candidate lncRNAs between tumours (PCAWG) and normal tissues (GTEx)
4. In depth analysis of two lncRNA candidates in Ovarian cancer

- 1. Identified high expressing lncRNA candidates in multiple cancer types**
2. Survival analysis of lncRNAs using the Cox Proportional Hazards model
3. Differential expression analysis of candidate lncRNAs between tumours (PCAWG) and normal tissues (GTEx)
4. In depth analysis of two lncRNA candidates in Ovarian cancer

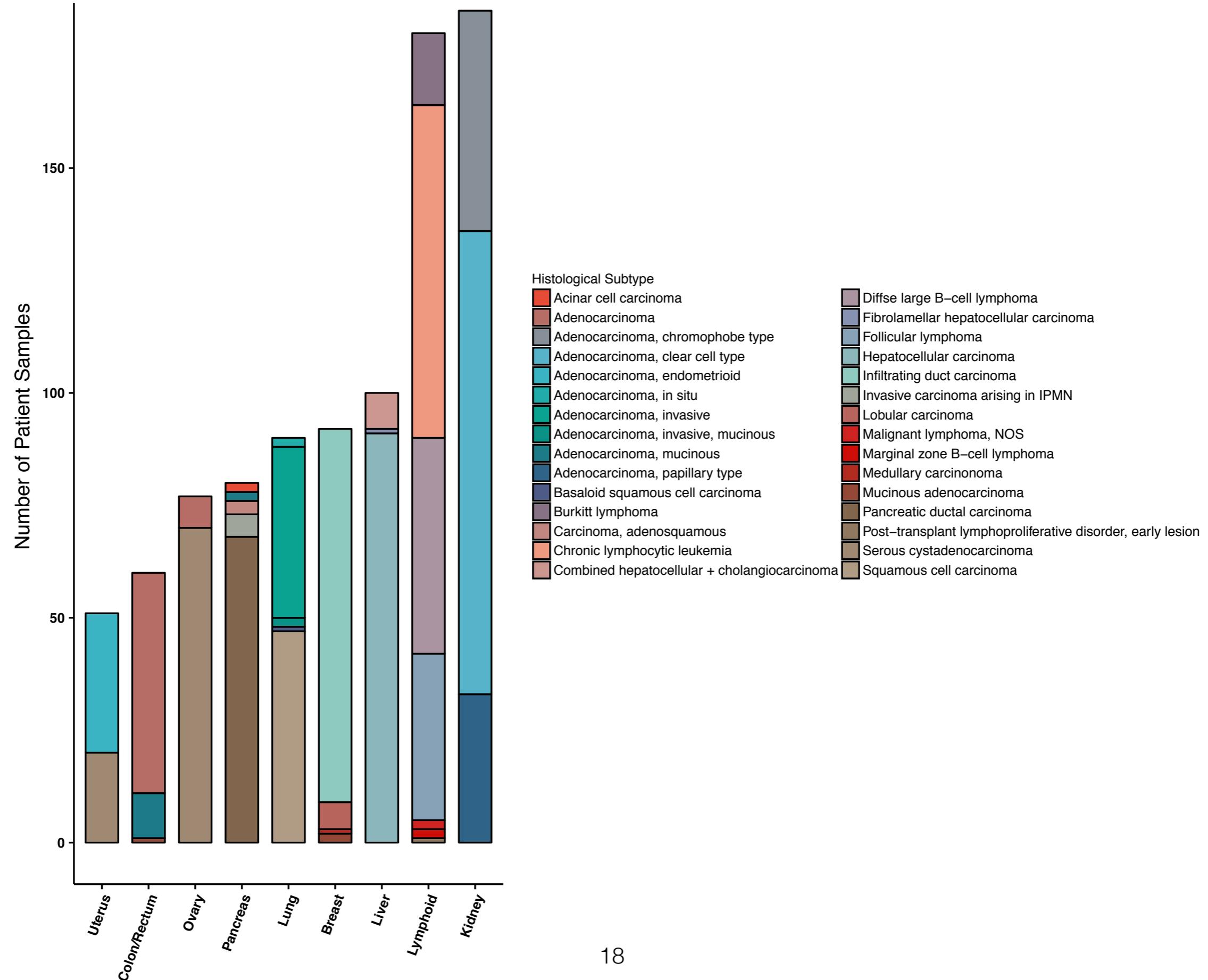
1,267 PCAWG patients with tumour RNA-Seq data



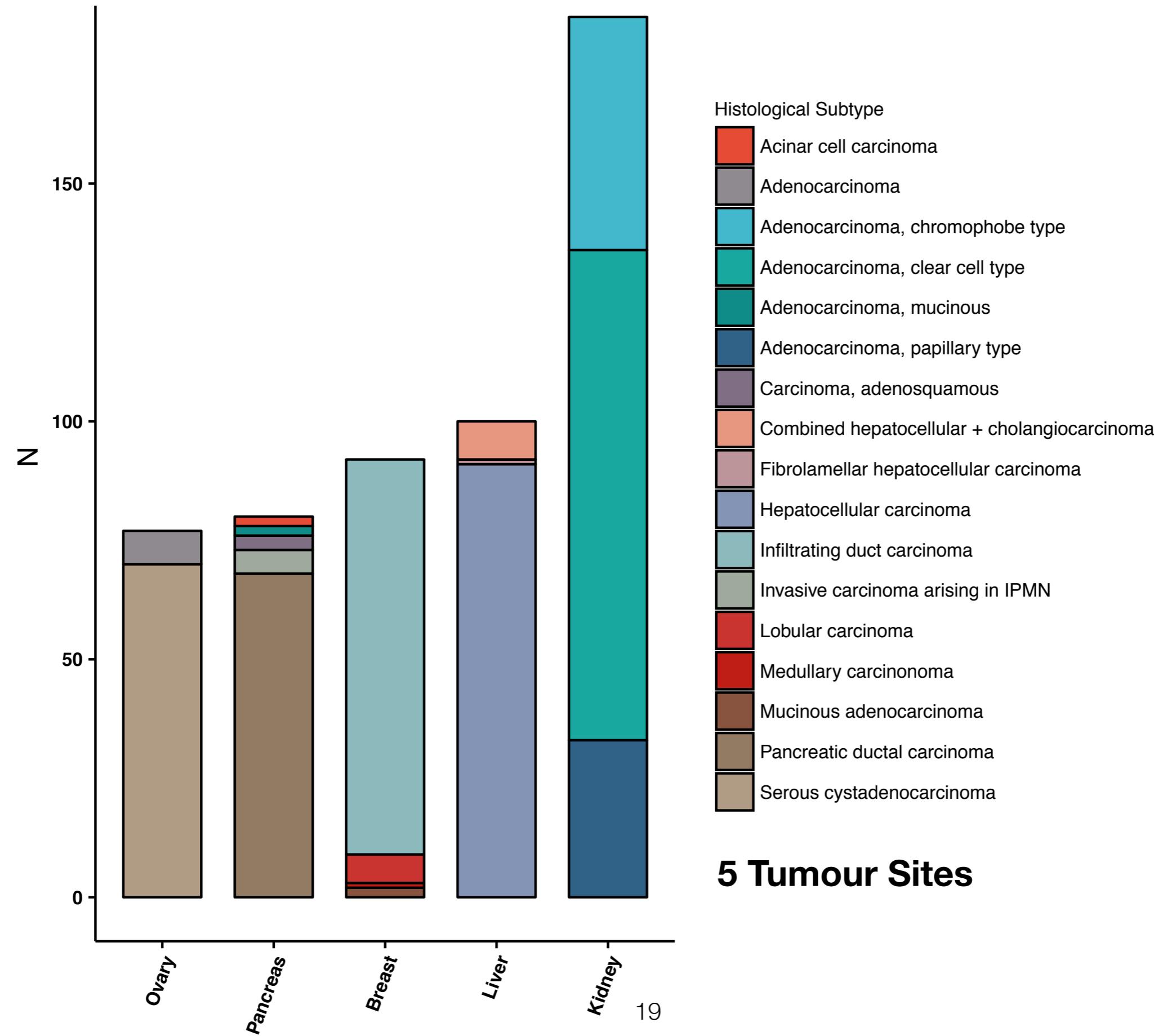
Cancers that have at least 50 patient samples



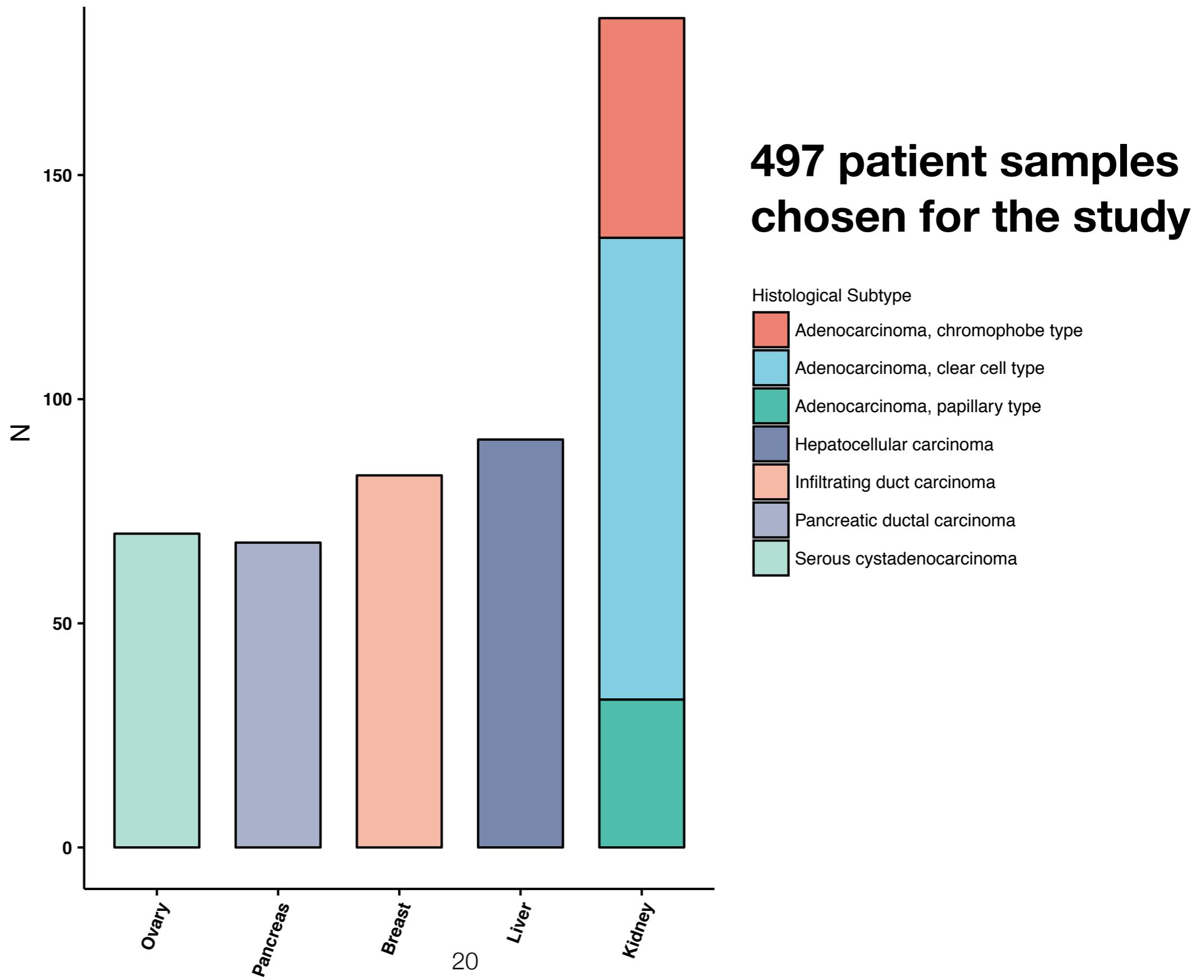
Distribution of histological subtypes within each cancer with 50 or more samples



Cancers with at least one histological subtype with 50 or more samples



Finally, remove histological subtypes with 10 or less patients



Filtering lncRNAs with low expression in all cancer types

12,598 Ensemble lncRNAs (**intergenic and antisense**) with RNA-Seq data



Obtain median expression value of each lncRNA within every cancer type



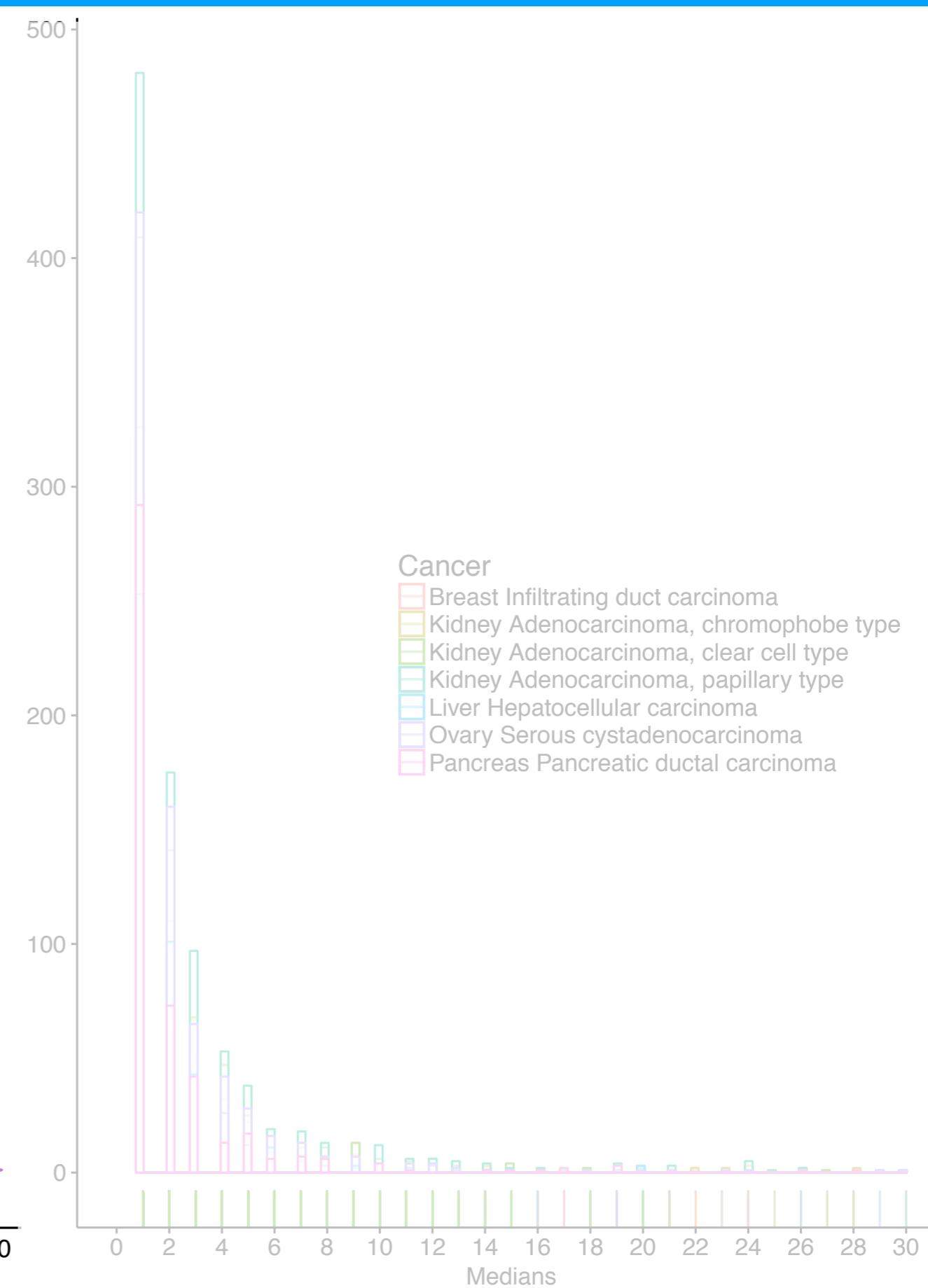
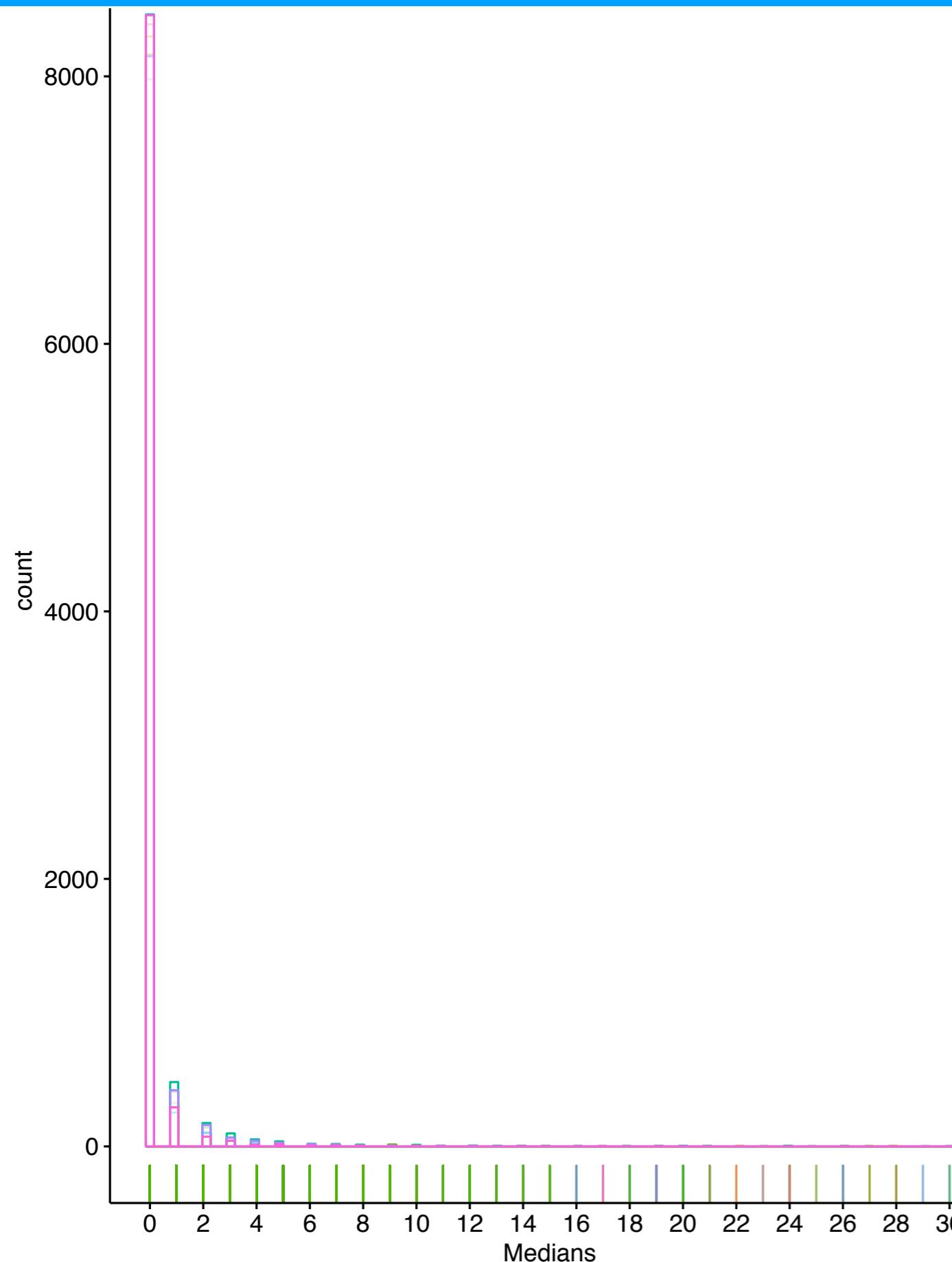
8,940 have median FPKM expression greater than 0 in at least one cancer type



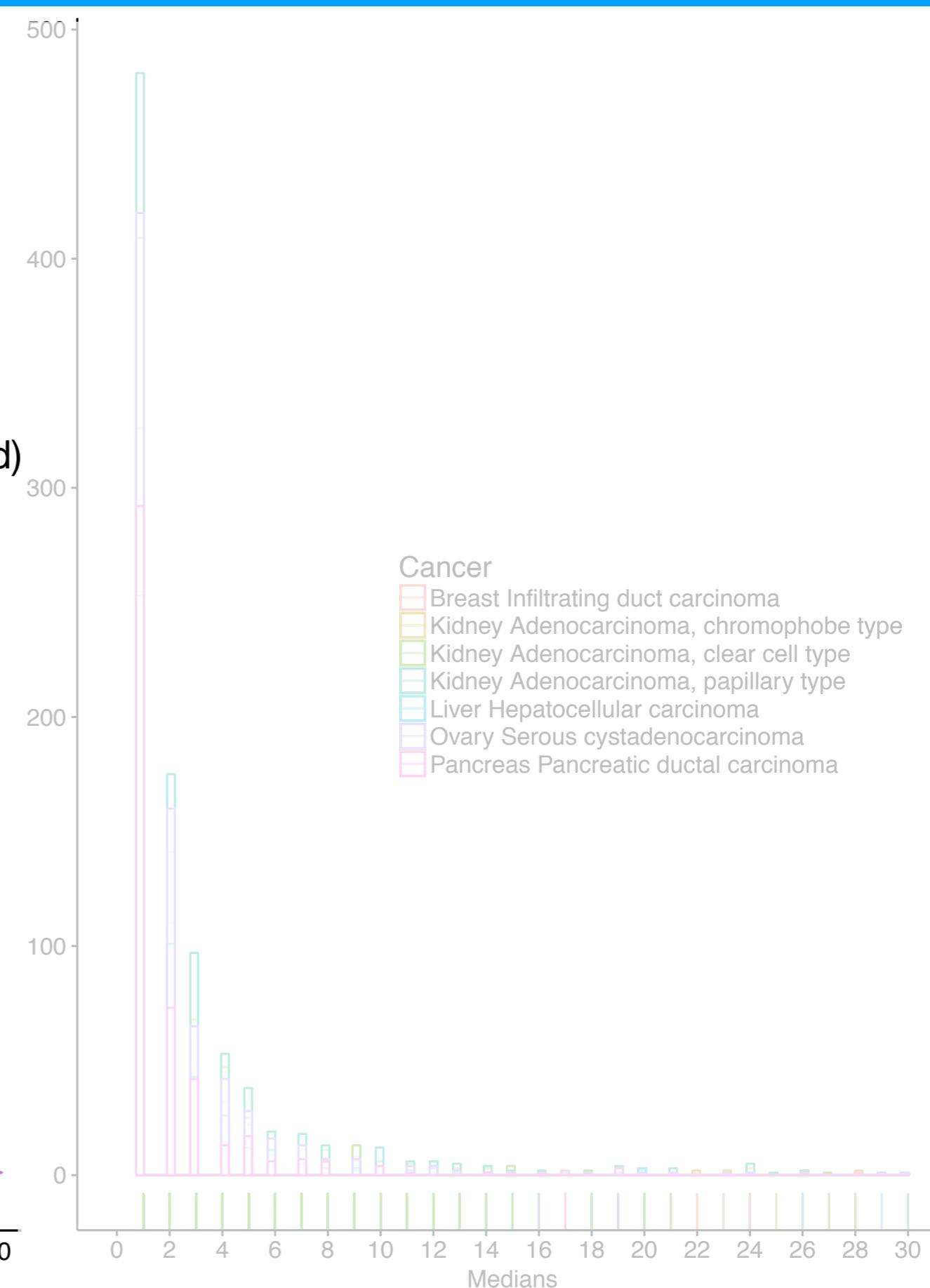
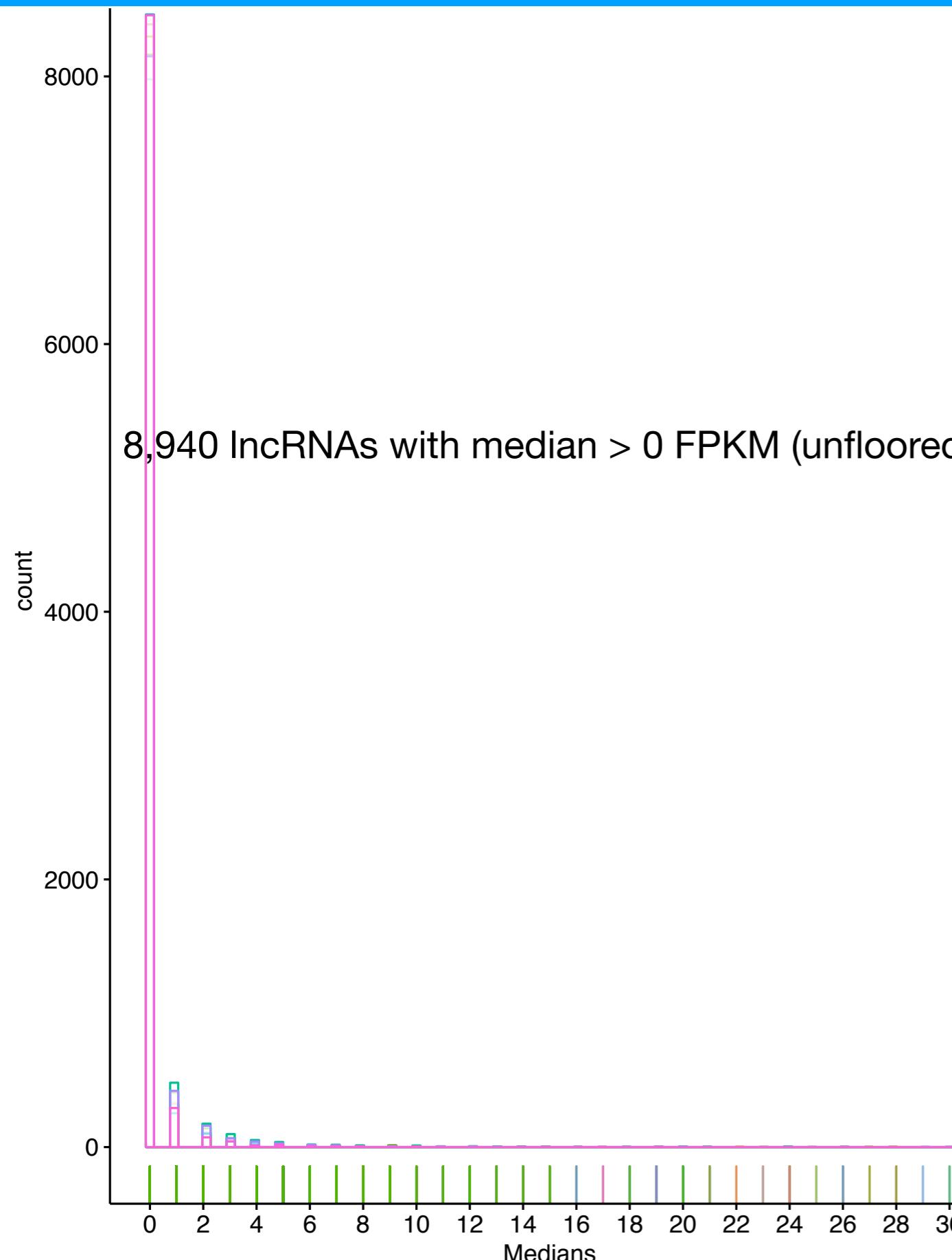
Floor median values eg. 0.7 → 0

1,604 lncRNAs with median expression greater than 1 FPKM in at least one cancer type

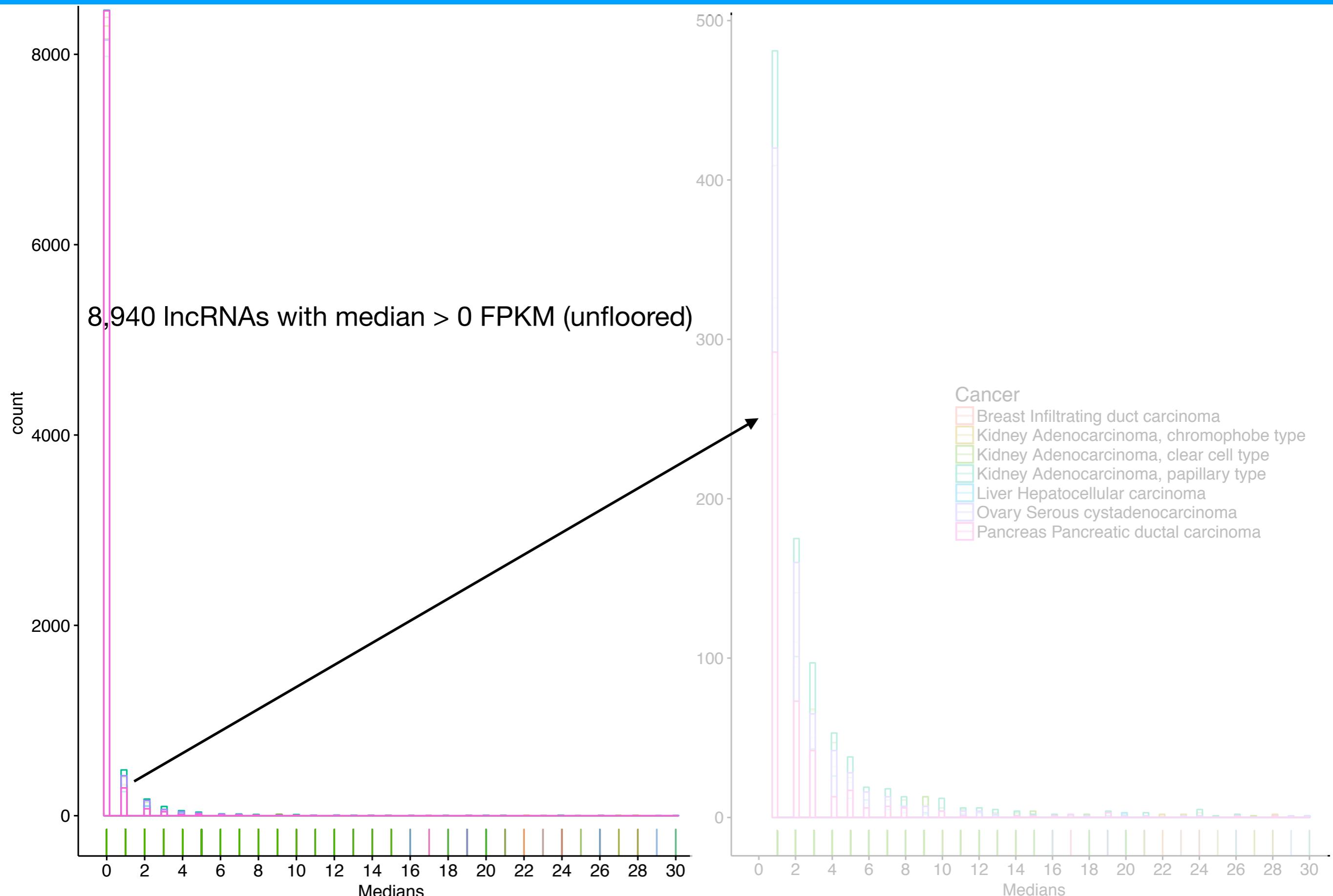
Distribution of median expression values of 8,940 lncRNAs



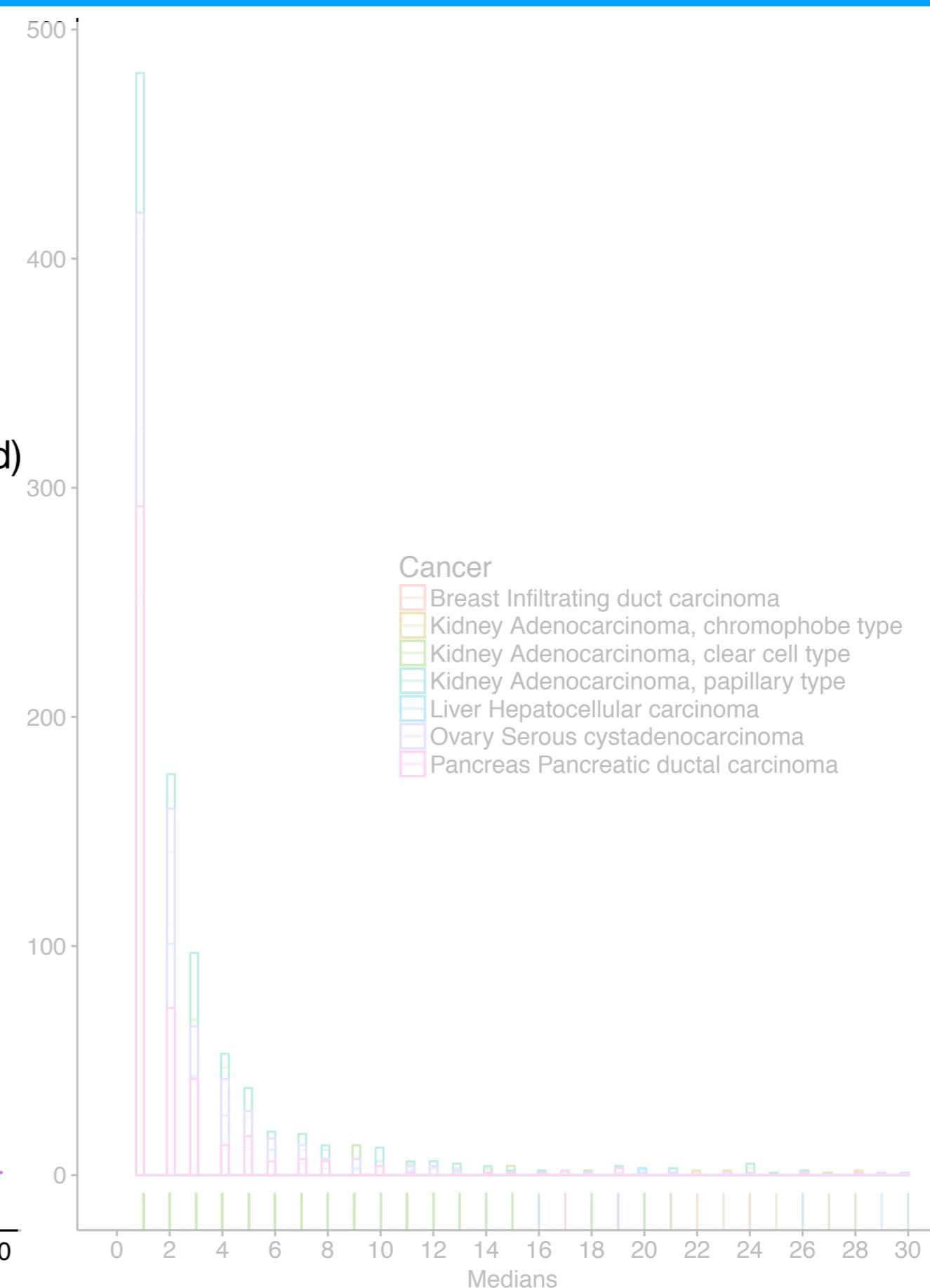
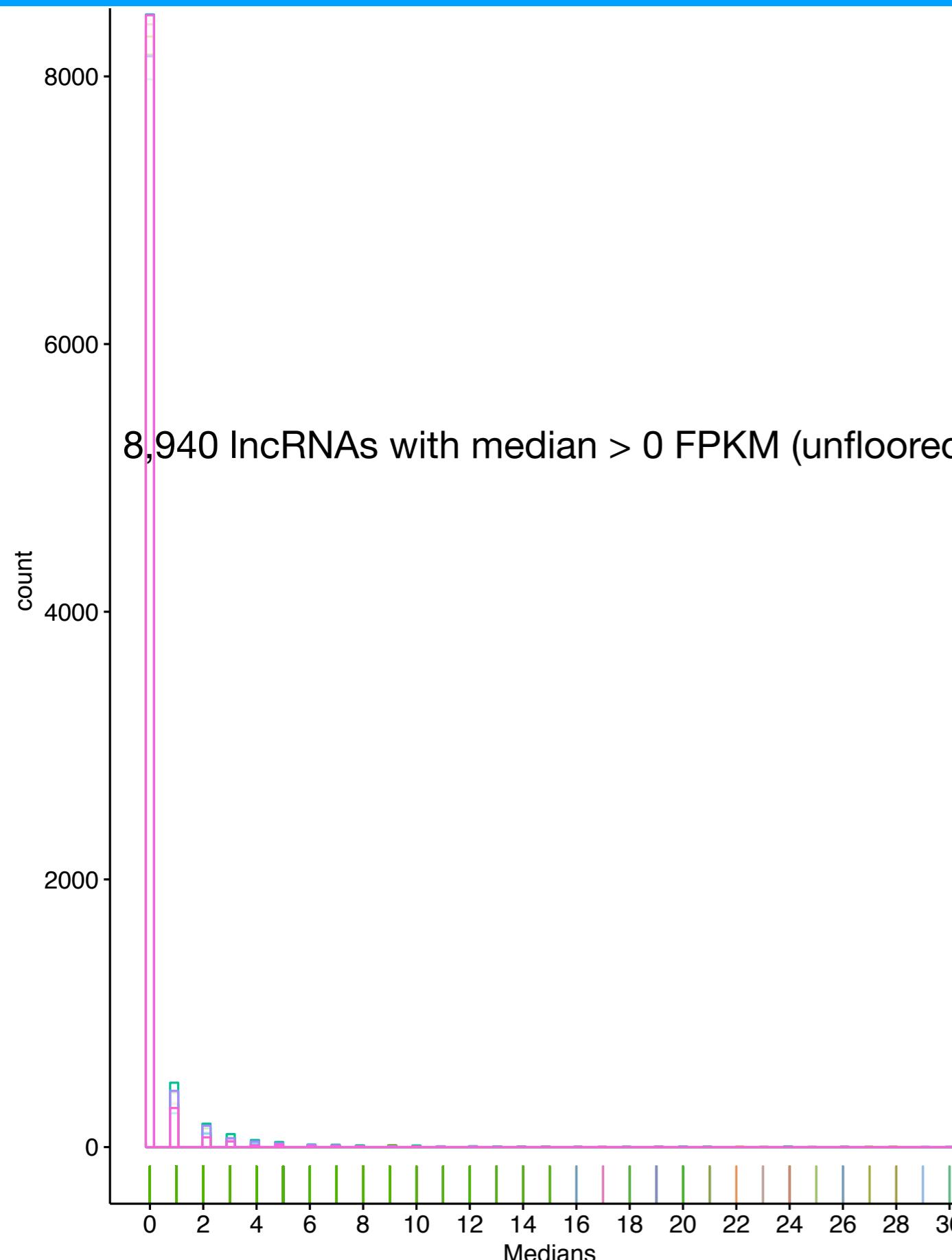
Distribution of median expression values of 8,940 lncRNAs



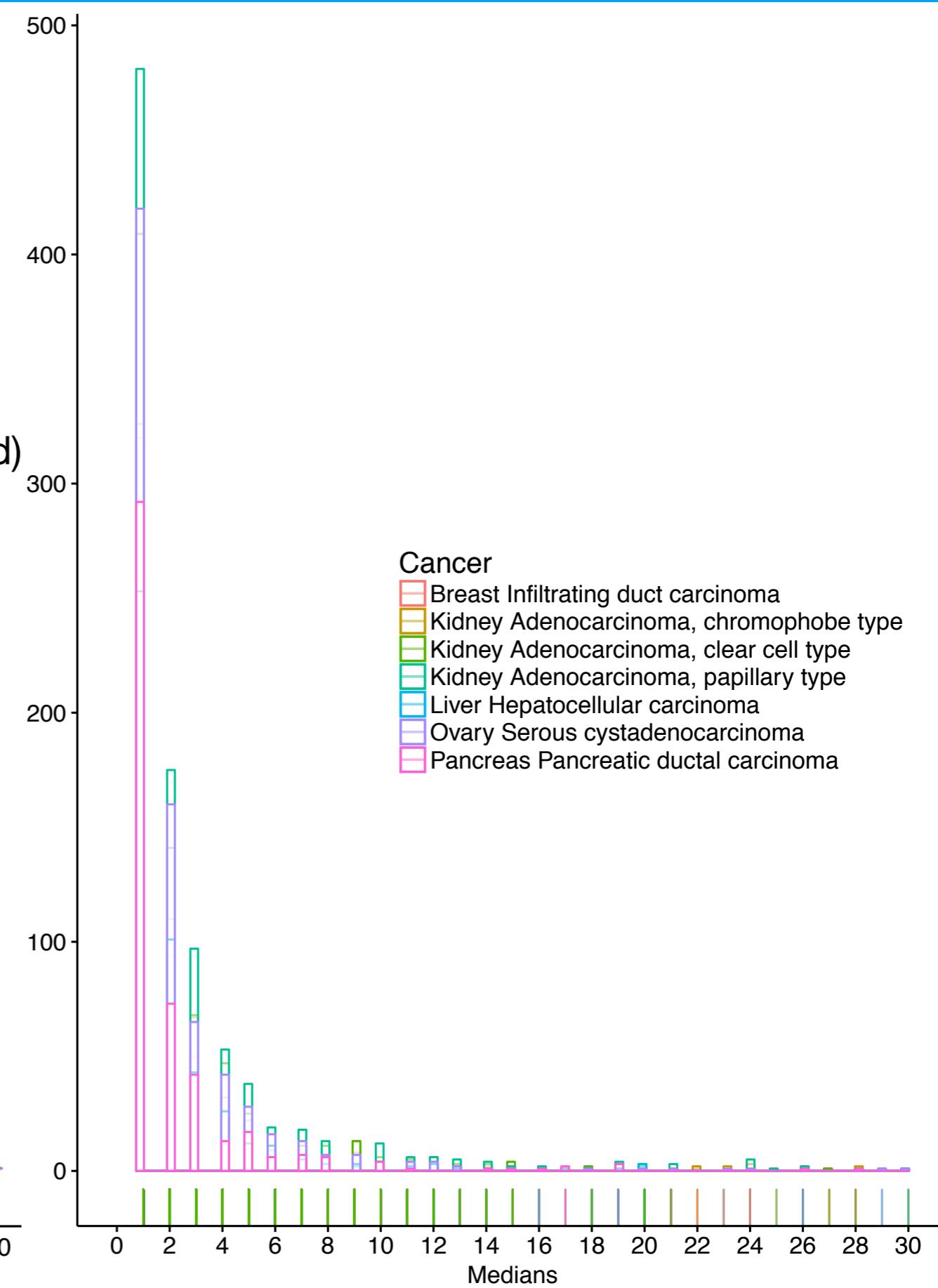
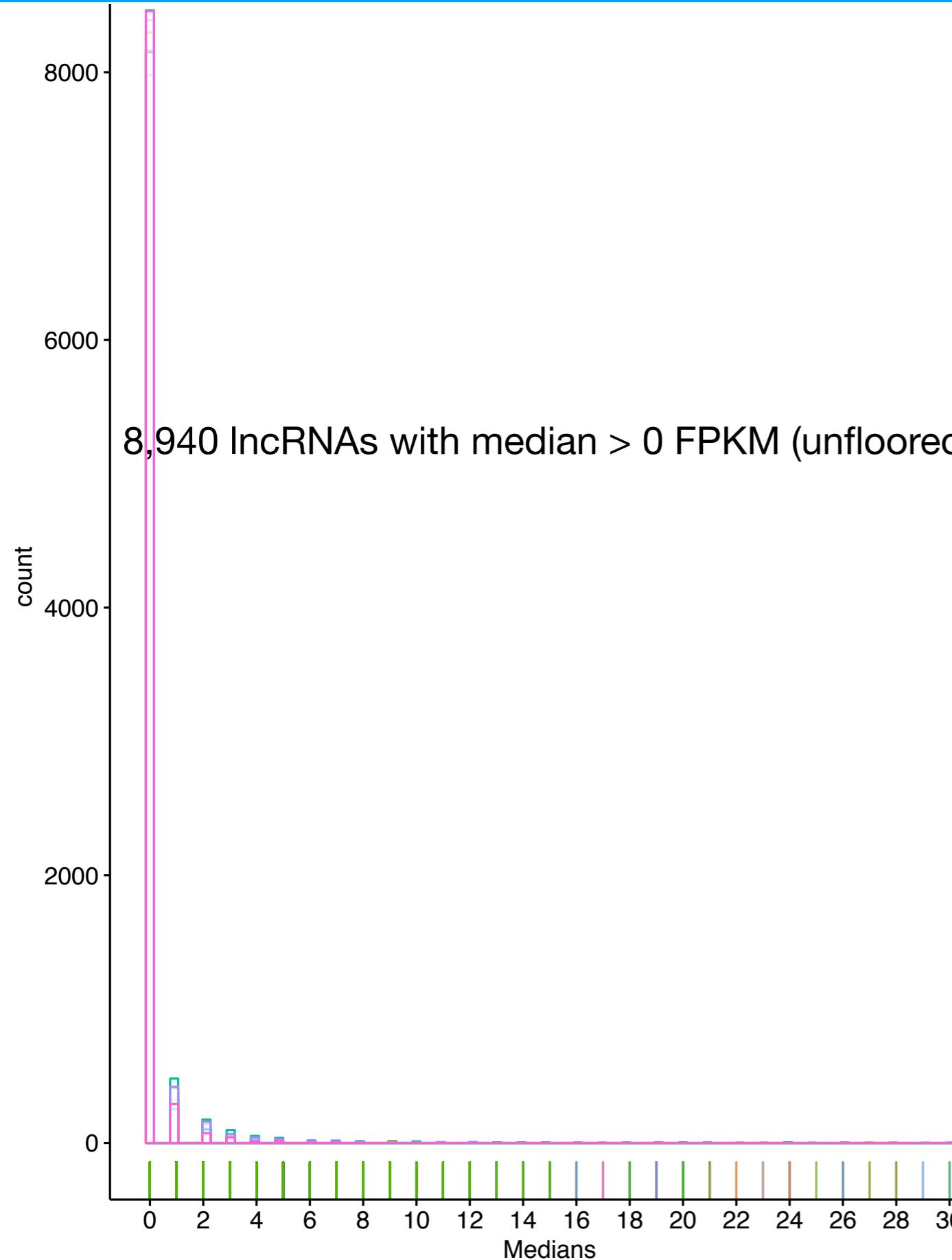
Distribution of median expression values of 8,940 lncRNAs



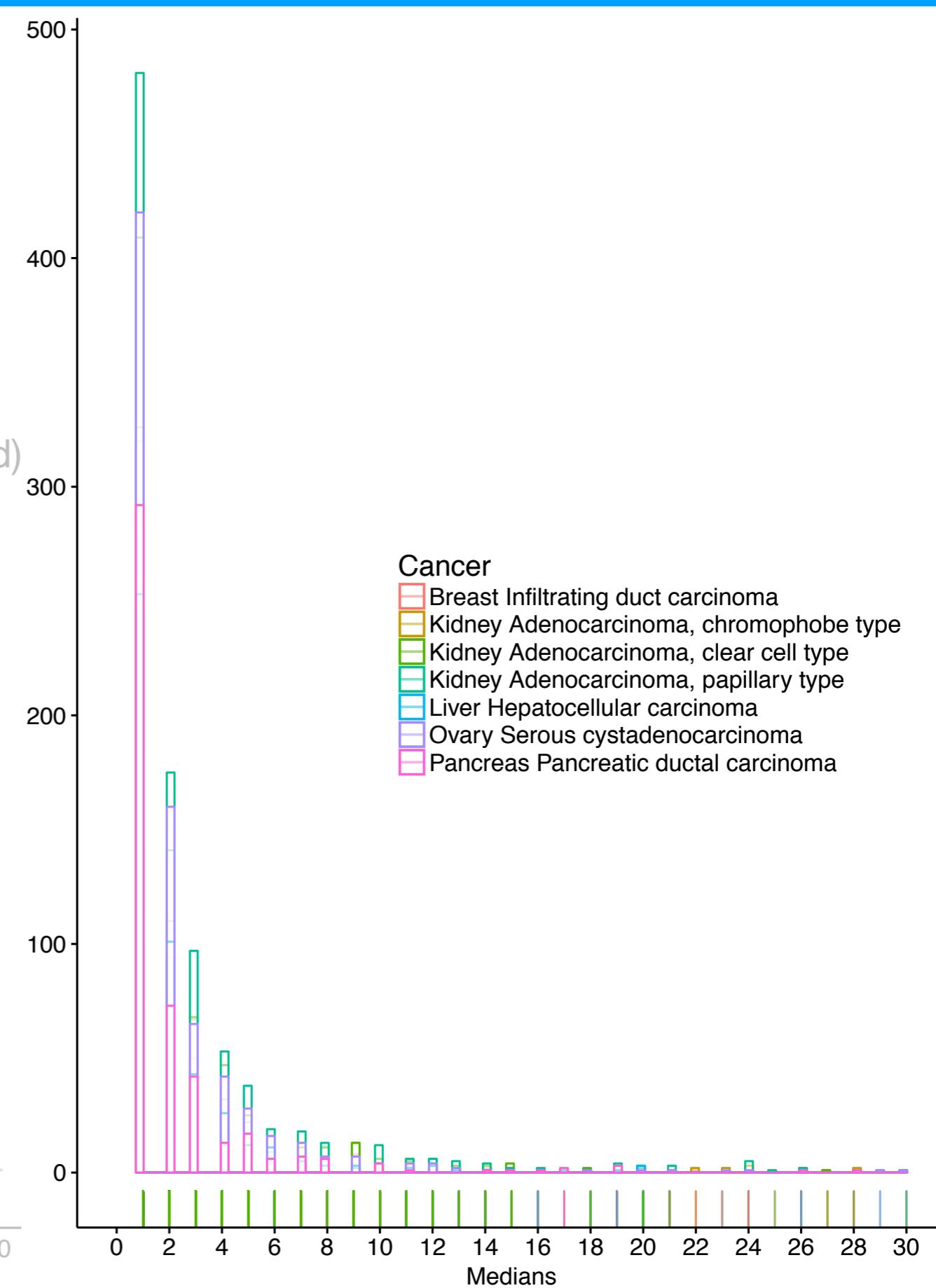
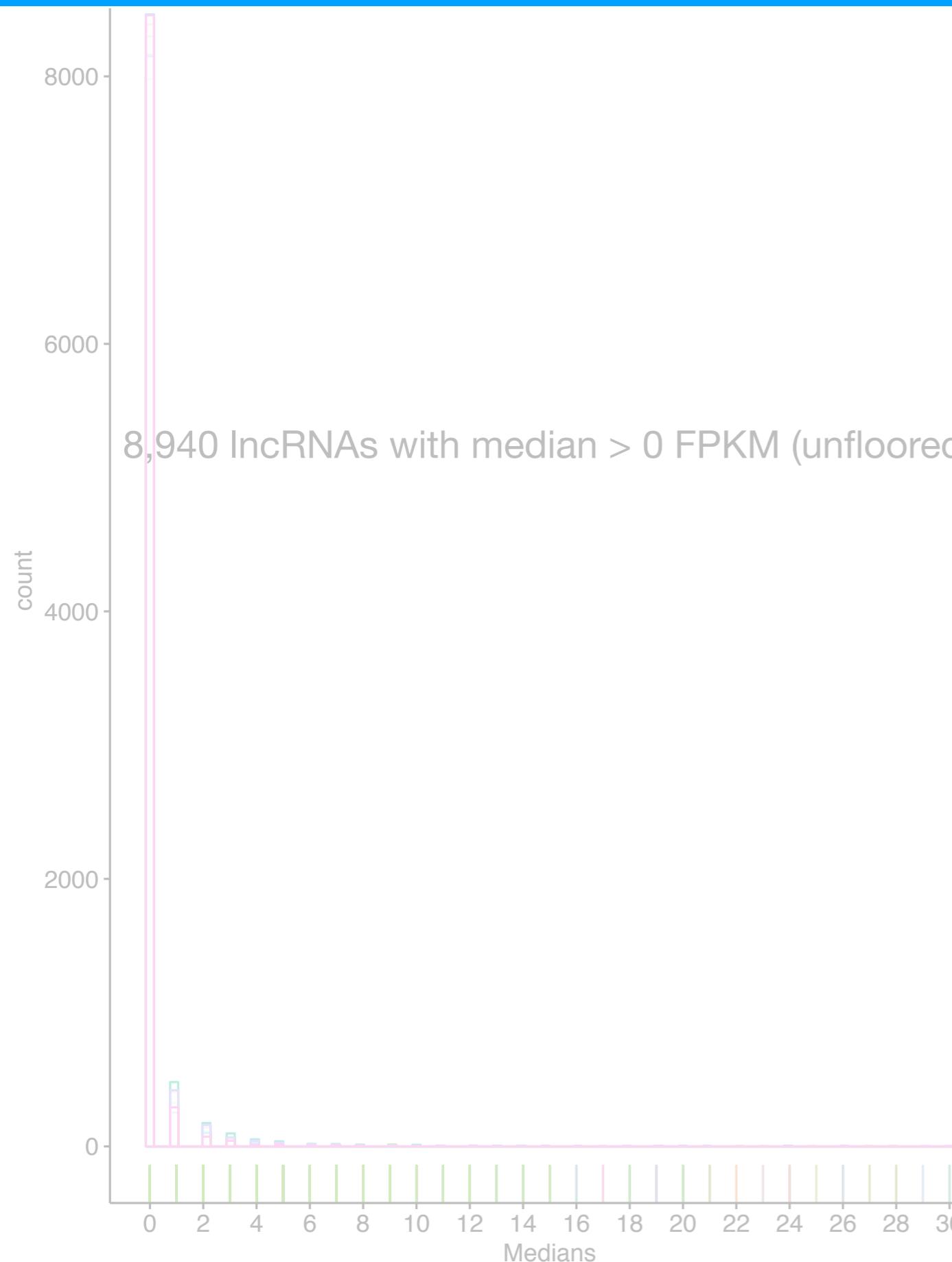
Distribution of median expression values of 8,940 lncRNAs



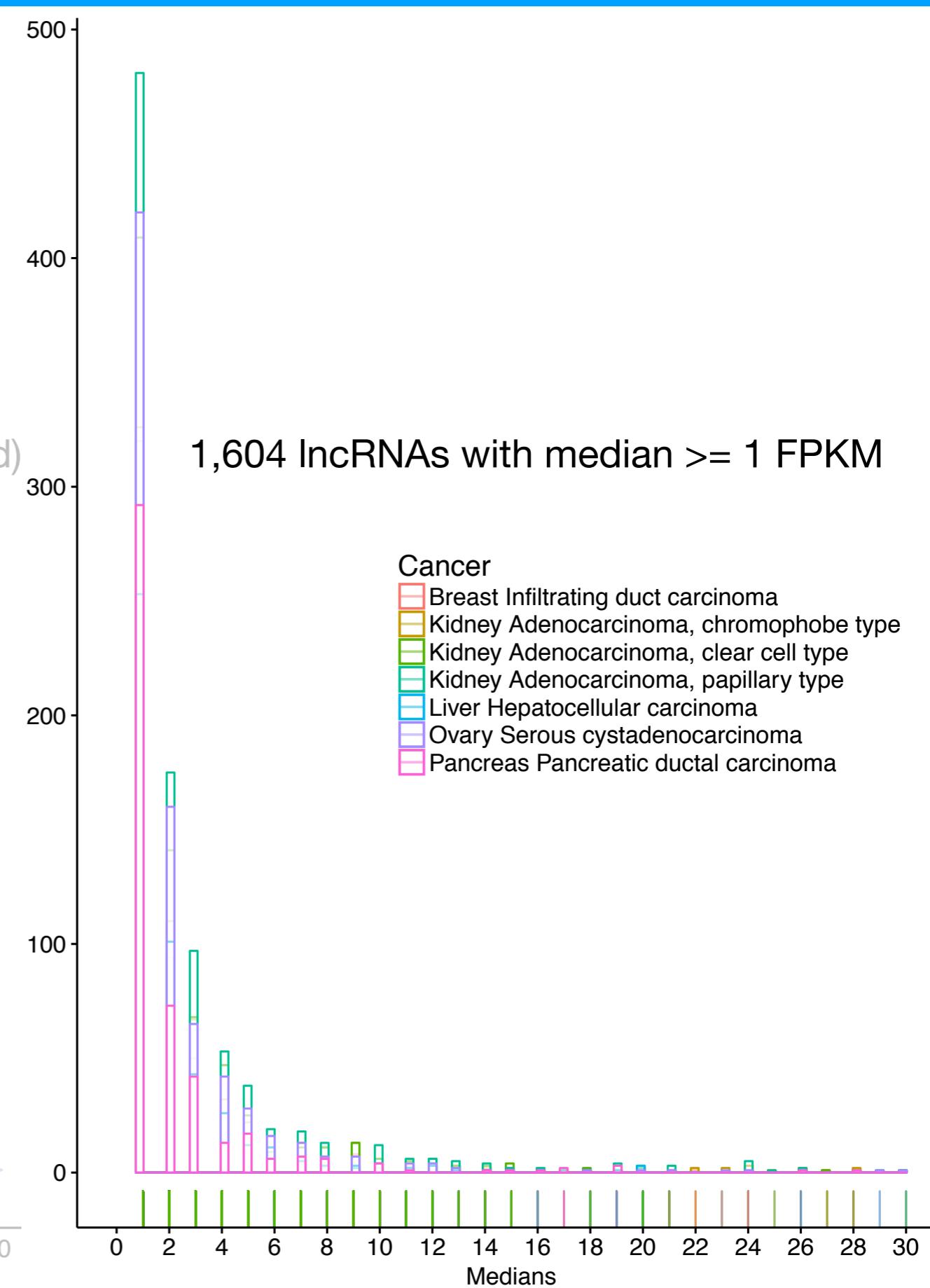
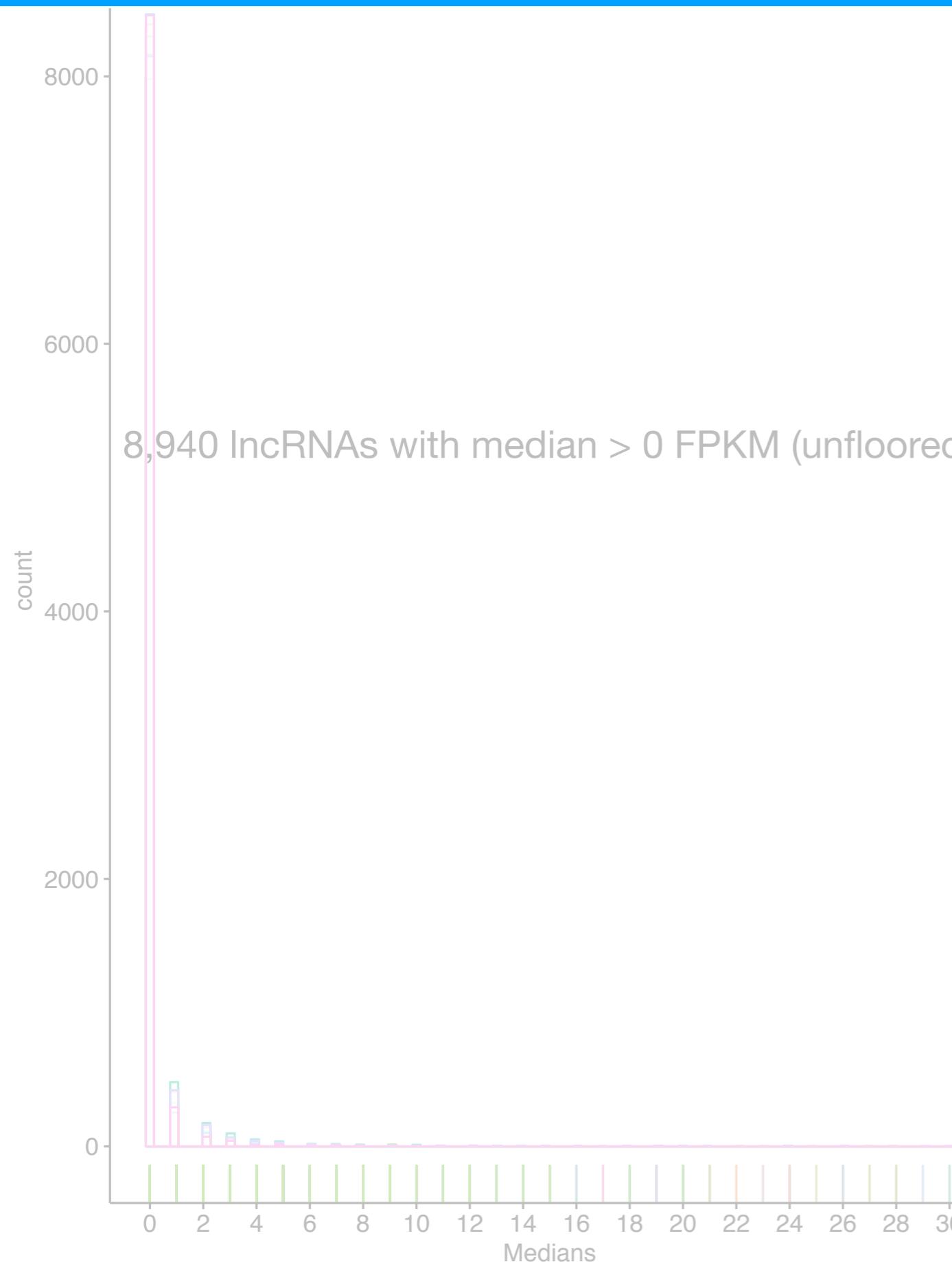
Distribution of median expression values of 8,940 lncRNAs



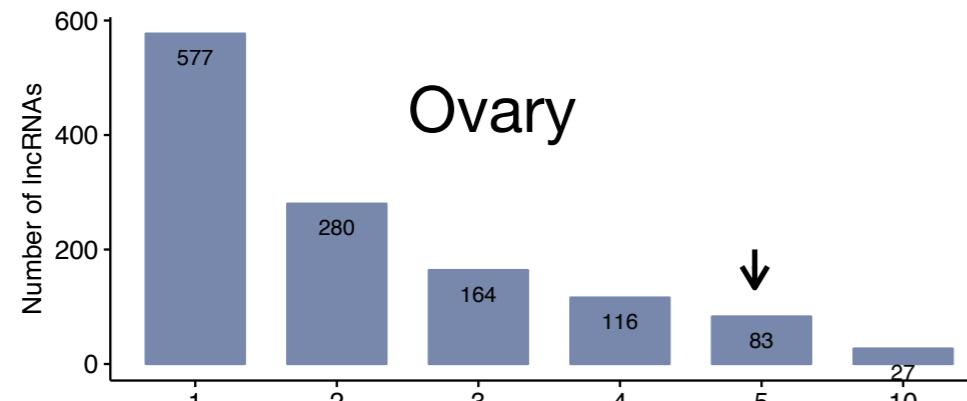
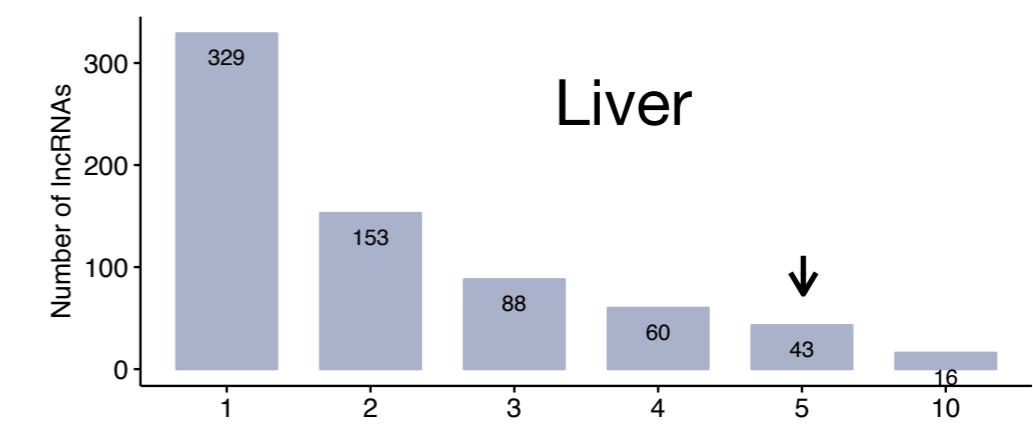
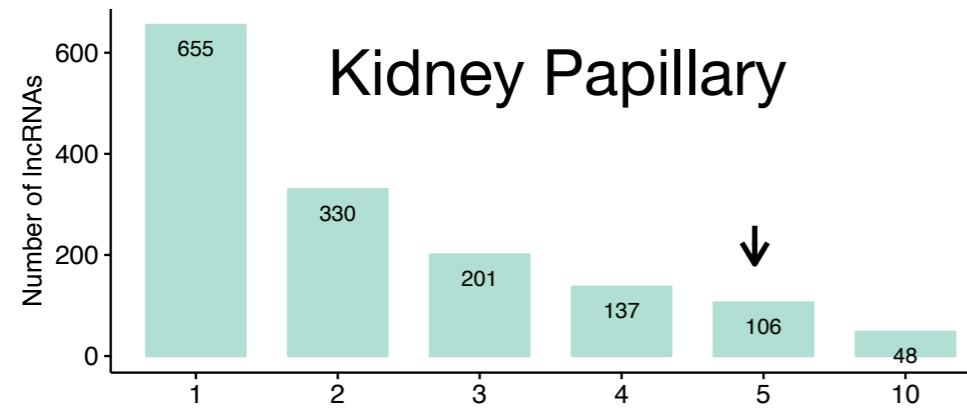
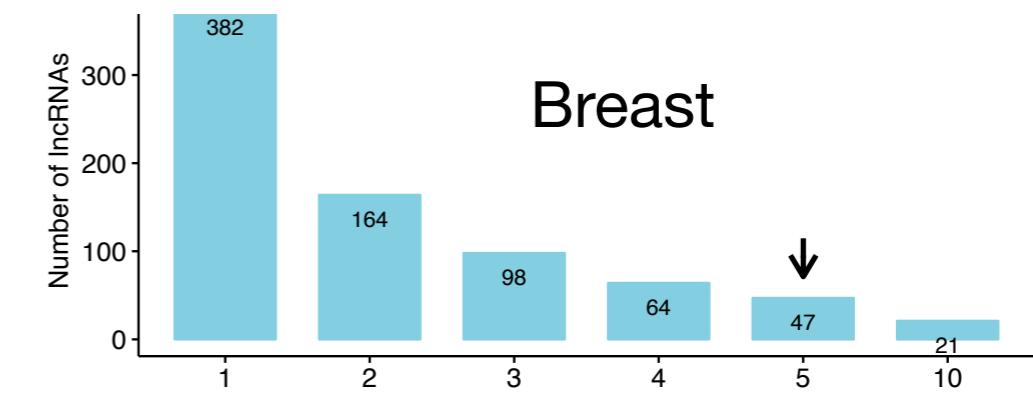
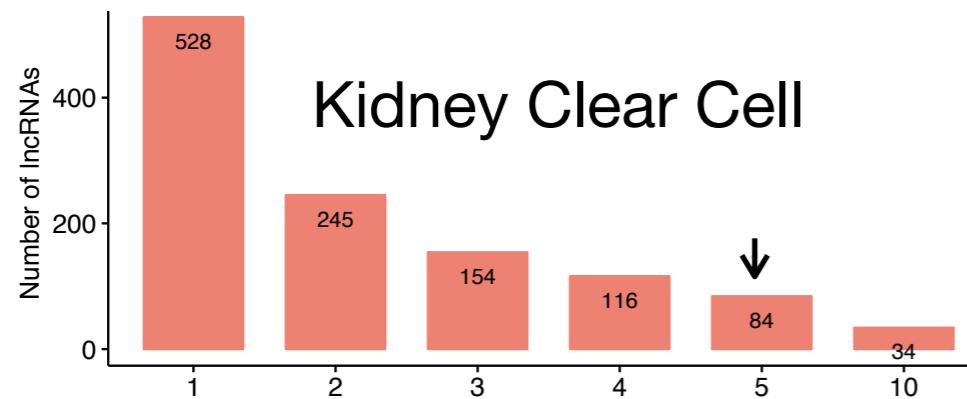
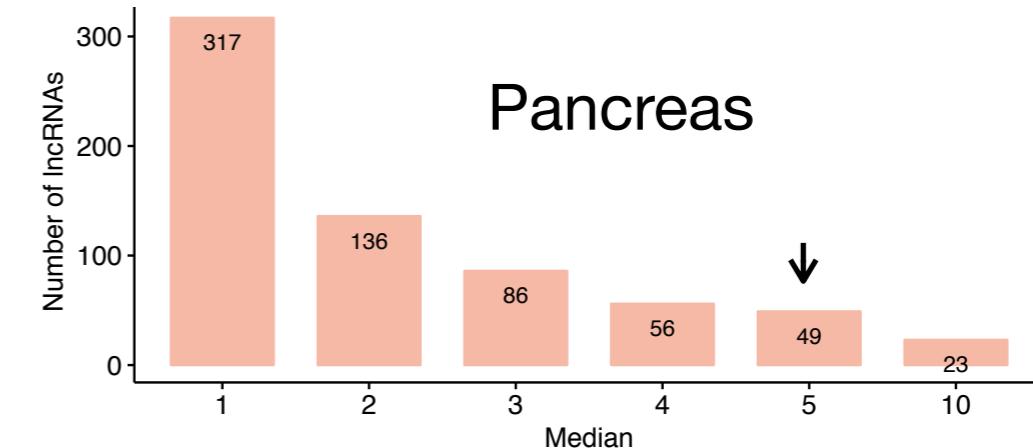
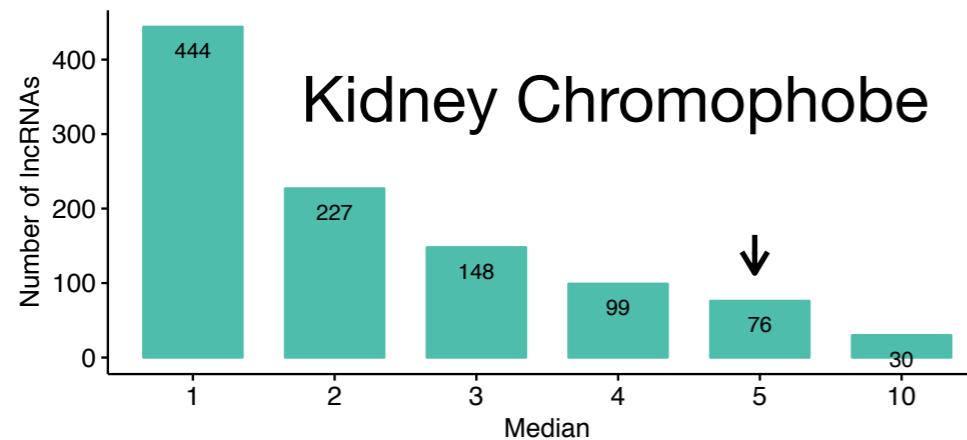
Distribution of median expression values of 8,940 lncRNAs



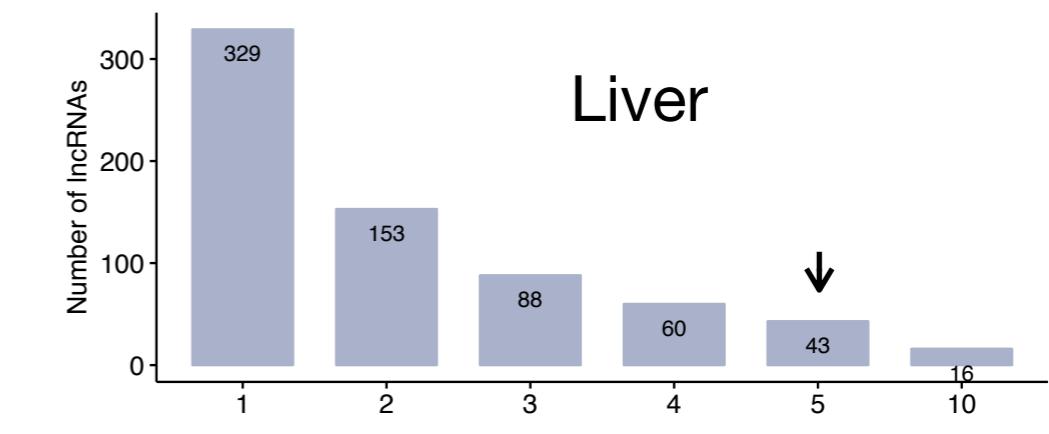
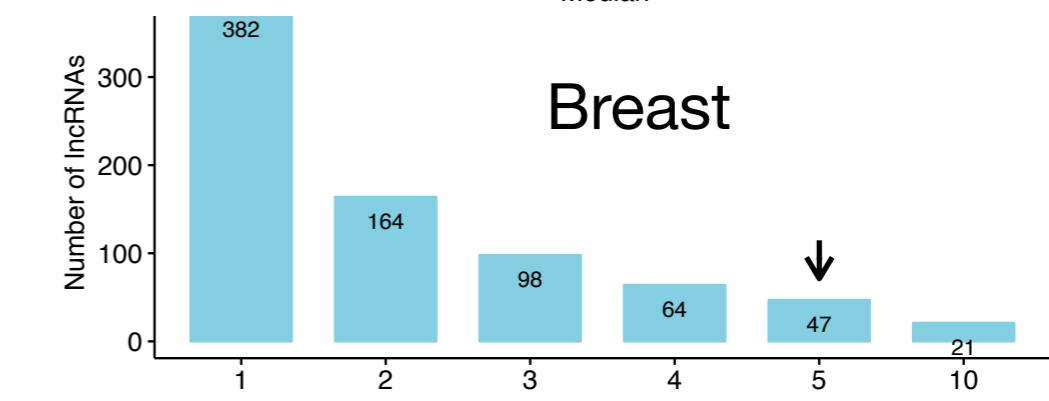
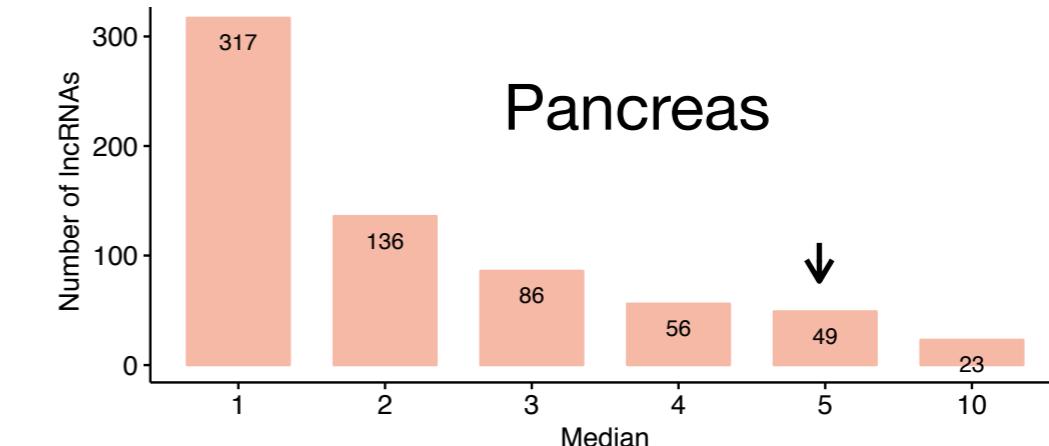
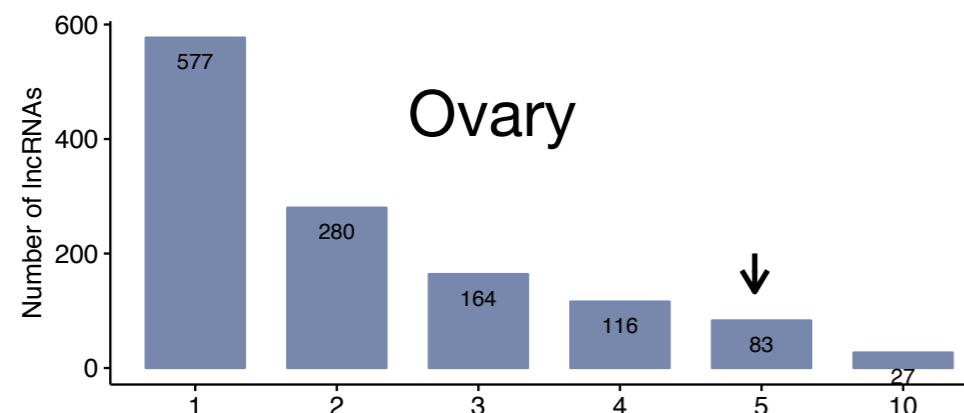
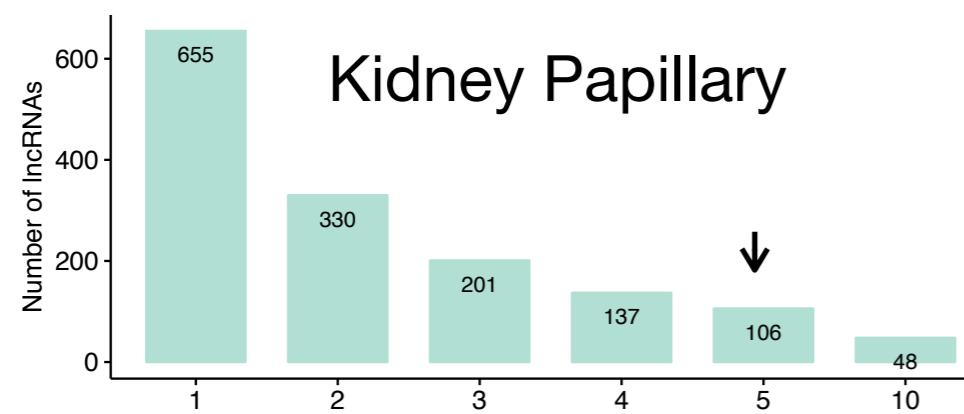
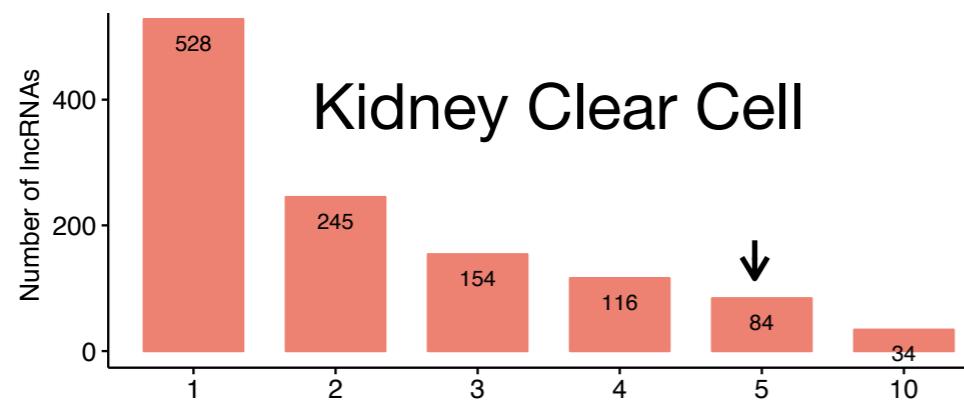
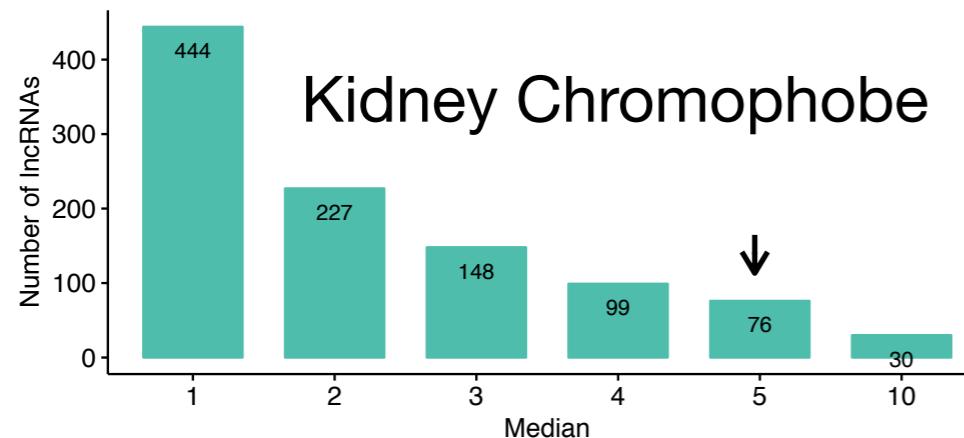
Distribution of median expression values of 8,940 lncRNAs



Identifying median cutoff for filtering low expressing lncRNAs



Identifying median cutoff for filtering low expressing lncRNAs

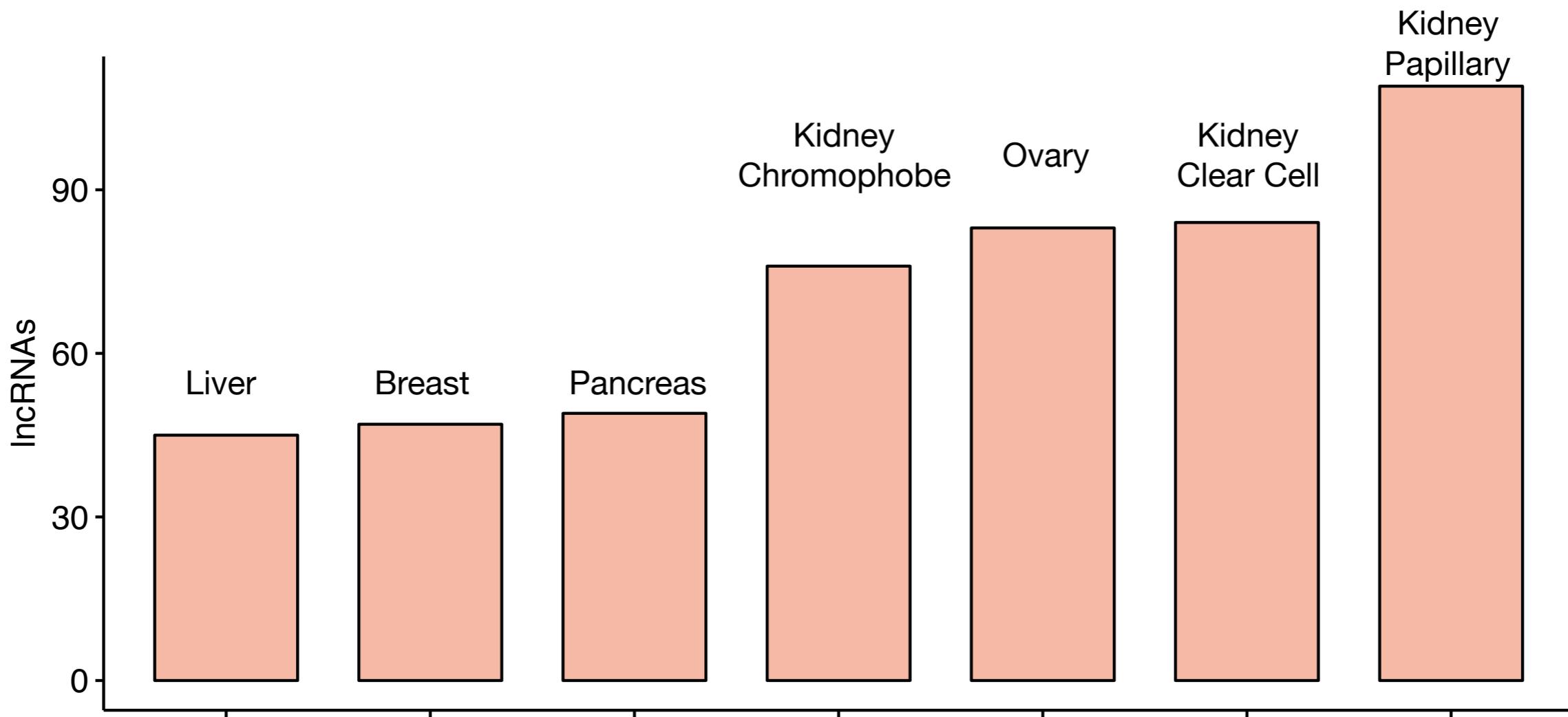


Further:

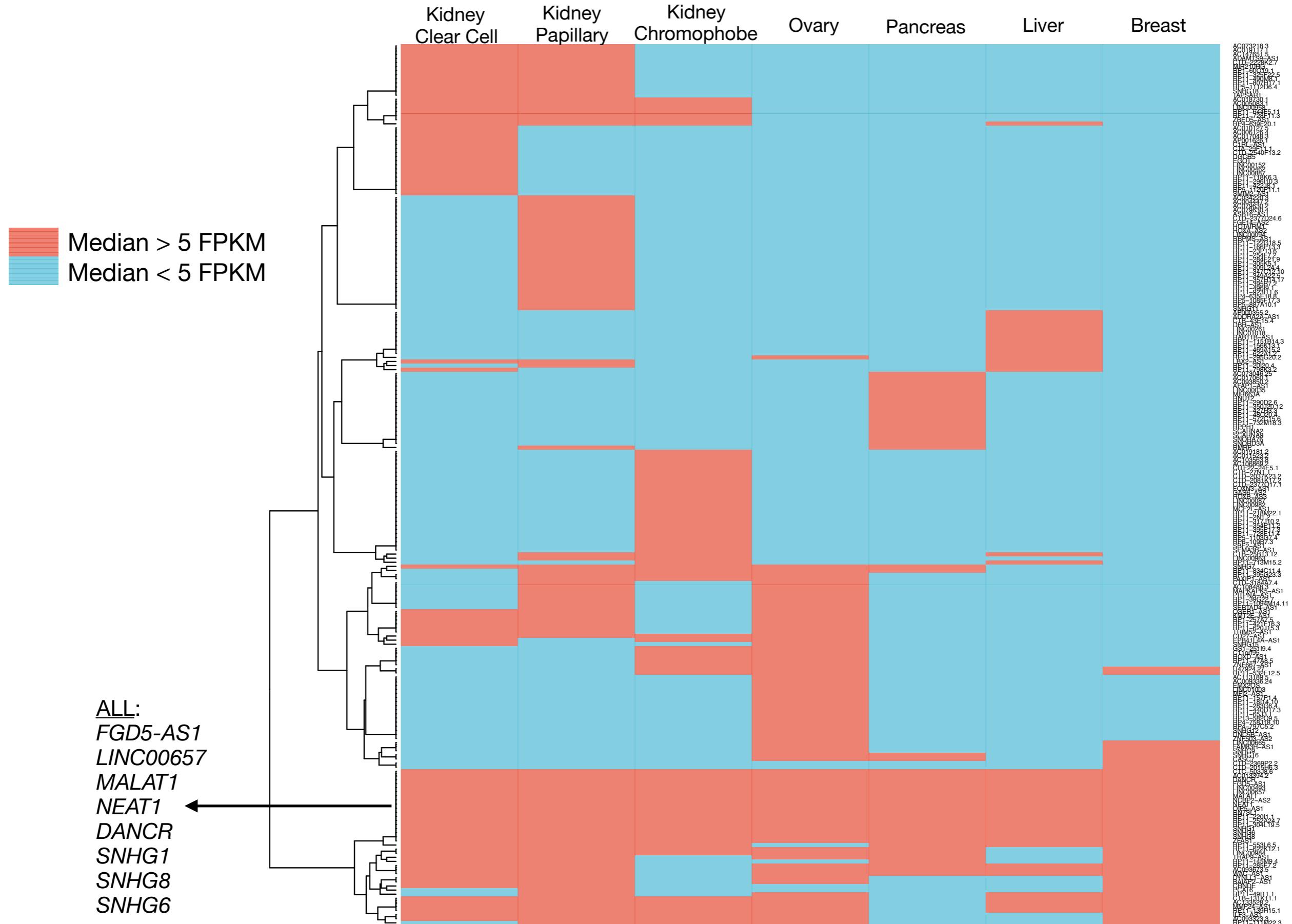
1. Filter lncRNAs not annotated in FANTOM CAT
2. Median cutoff = 5 FPKM

Summary of filtering lowly expressed lncRNAs

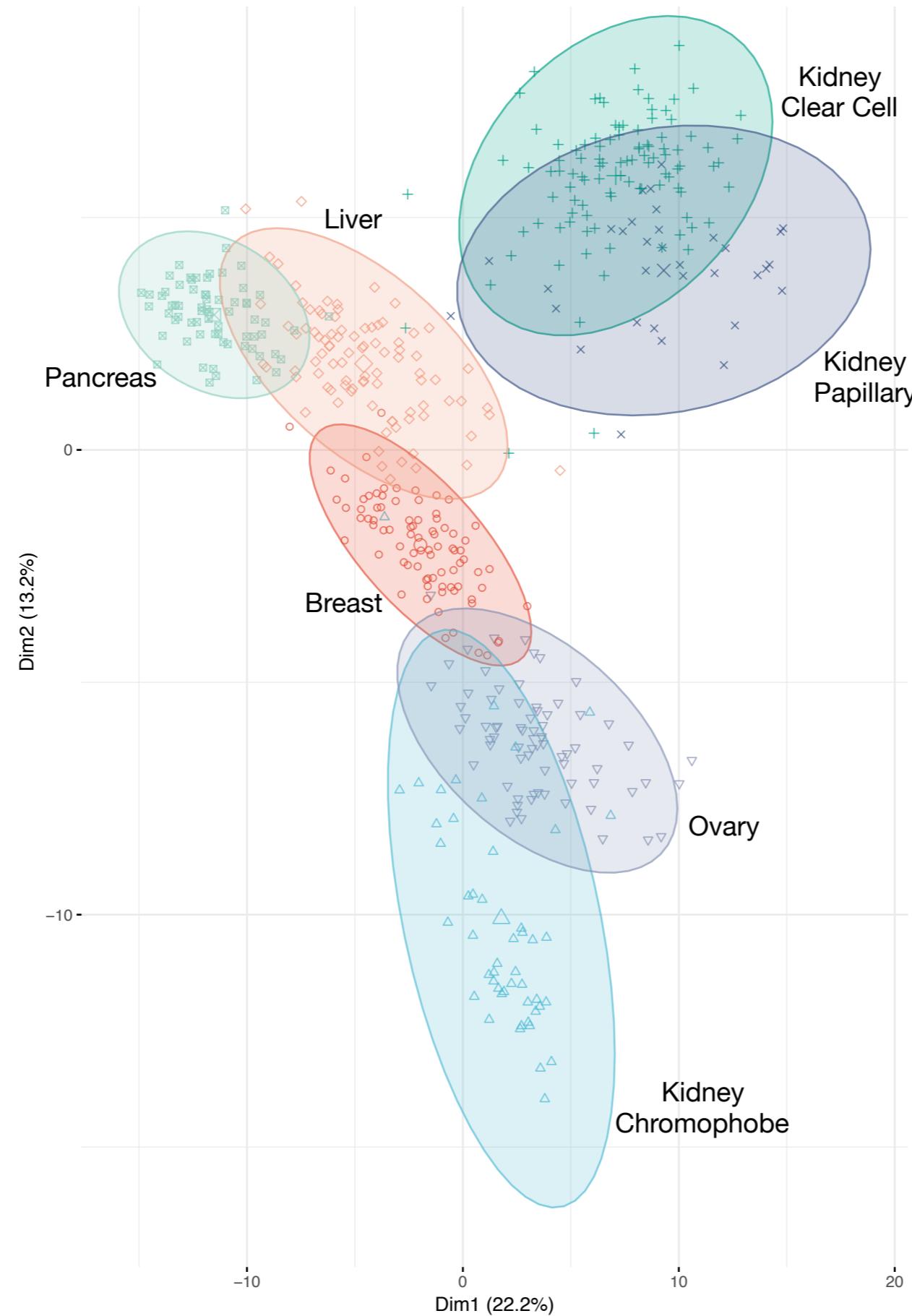
- 12,598 Ensemble lncRNAs in RNA-Seq dataset
- Filter 1 (Annotation in Fantom CAT): 5,607 lncRNAs
- Filter 2 (Median threshold): 493 lncRNAs across cancers
 - 215 unique lncRNAs selected for study with median FPKM > 5 in at least one cancer type



215 highly expressed lncRNAs across cancer types



Principal Component Analysis (PCA) using 215 lncRNAs



Summary of progress

1. Identified high expressing lncRNA candidates in multiple cancer types
2. Survival analysis of lncRNAs using the Cox Proportional Hazards model
3. Differential expression analysis of candidate lncRNAs between tumours (PCAWG) and normal tissues (GTEx)
4. In depth analysis of two lncRNA candidates in Ovarian cancer

Summary of progress

1. Identified high expressing lncRNA candidates in multiple cancer types
2. **Survival analysis of lncRNAs using the Cox Proportional Hazards model**
3. Differential expression analysis of candidate lncRNAs between tumours (PCAWG) and normal tissues (GTEx)
4. In depth analysis of two lncRNA candidates in Ovarian cancer

lncRNAs associated with patient survival outcome

lncRNAs associated with patient survival outcome

- 493 lncRNA-cancer associations tested:
 - 7 cancer types
 - 215 unique lncRNAs

lncRNAs associated with patient survival outcome

- 493 lncRNA-cancer associations tested:
 - 7 cancer types
 - 215 unique lncRNAs
- Median expression of lncRNA used to dichotomize patients into high and low expression groups

lncRNAs associated with patient survival outcome

- 493 lncRNA-cancer associations tested:
 - 7 cancer types
 - 215 unique lncRNAs
- Median expression of lncRNA used to dichotomize patients into high and low expression groups

$$\text{coxph}(\text{Surv}(time\ to\ death, status) \sim expression\ tag)$$

lncRNAs associated with patient survival outcome

- 493 lncRNA-cancer associations tested:
 - 7 cancer types
 - 215 unique lncRNAs
- Median expression of lncRNA used to dichotomize patients into high and low expression groups
- Global FDR correction on all 493 associations (conservative)

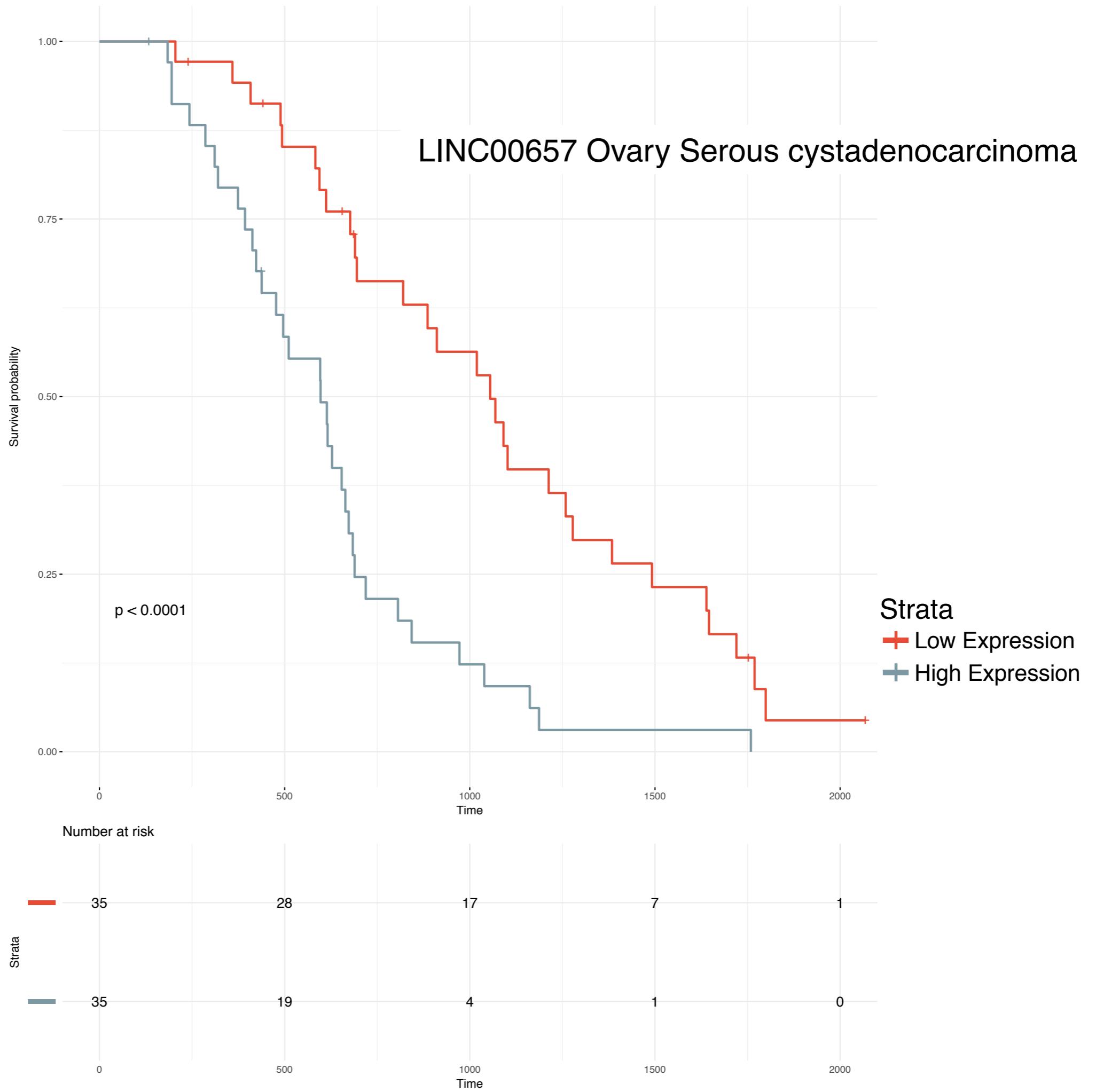
$$\text{coxph}(\text{Surv}(time\ to\ death,\ status) \sim expression\ tag)$$

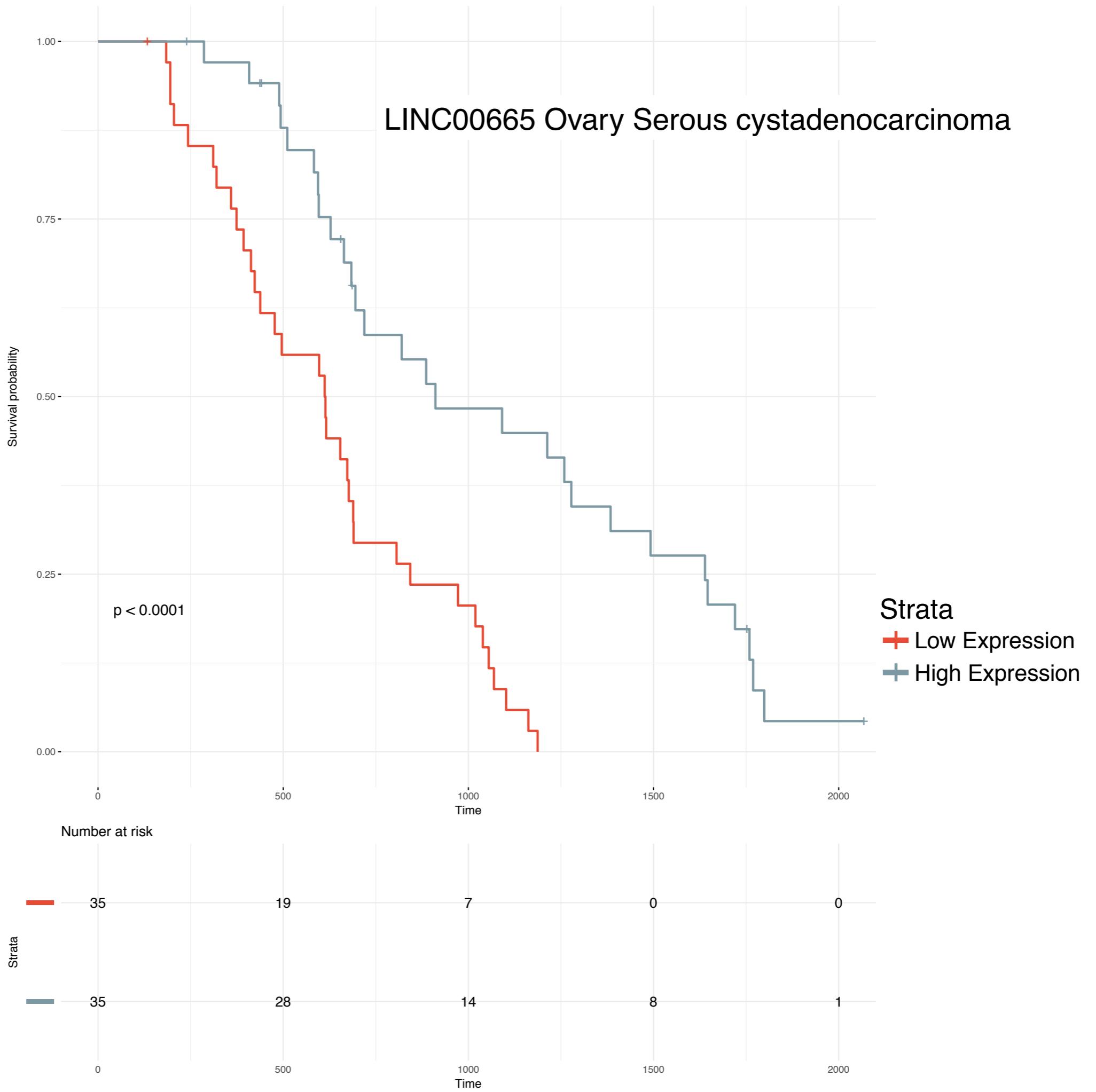
lncRNAs associated with patient survival outcome

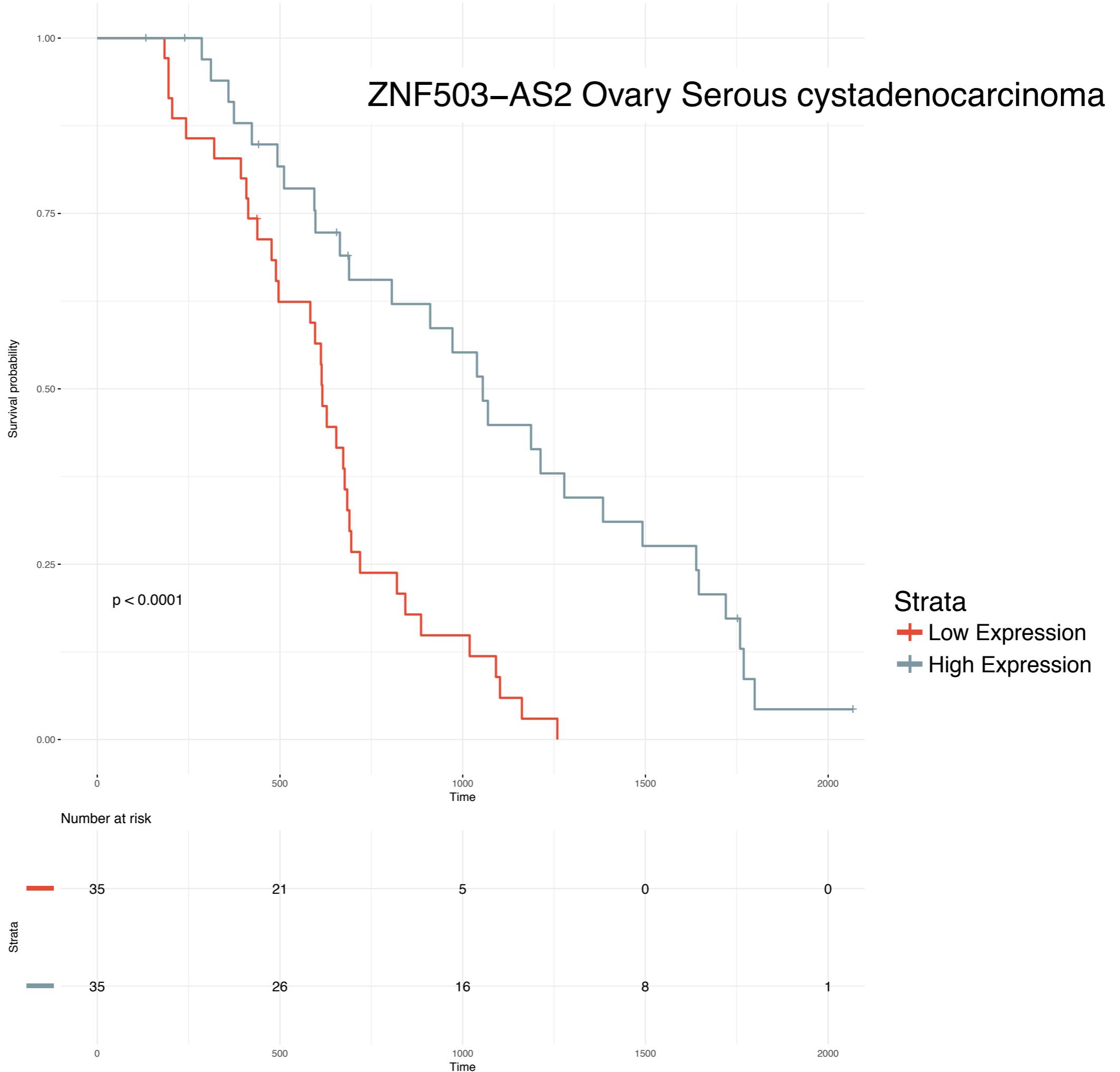
- 493 lncRNA-cancer associations tested:
 - 7 cancer types
 - 215 unique lncRNAs
- Median expression of lncRNA used to dichotomize patients into high and low expression groups
- Global FDR correction on all 493 associations (conservative)
- **42/493 associations with p-value < 0.05 (unadjusted) across four cancer types (Tier 2)**
 - Chosen for further analysis, filtering clinically relevant lncRNAs

lncRNAs associated with patient survival outcome

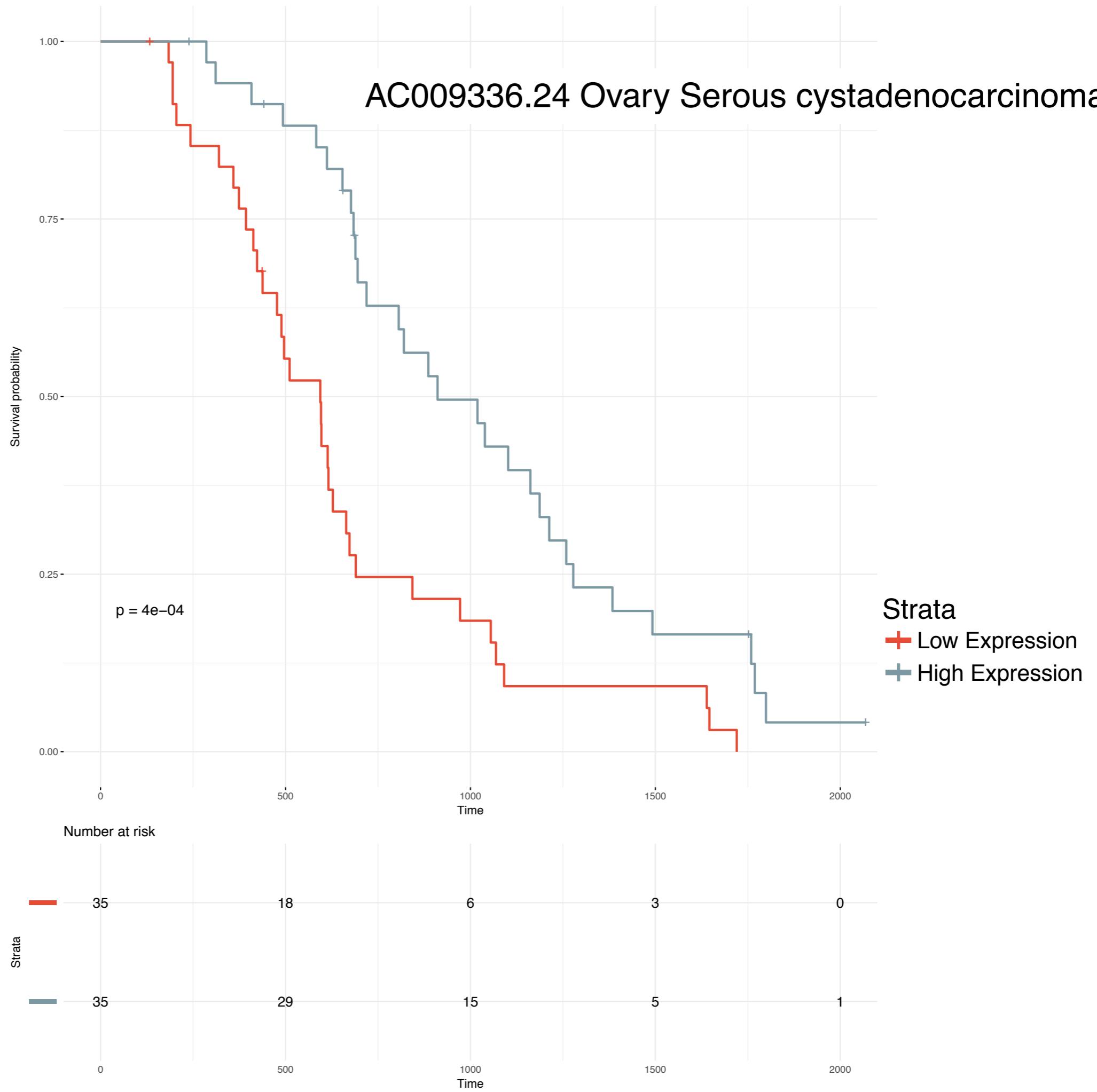
- 493 lncRNA-cancer associations tested:
 - 7 cancer types
 - 215 unique lncRNAs
- Median expression of lncRNA used to dichotomize patients into high and low expression groups
- Global FDR correction on all 493 associations (conservative)
- **42/493 associations with p-value < 0.05 (unadjusted) across four cancer types (Tier 2)**
 - Chosen for further analysis, filtering clinically relevant lncRNAs
- 7/42 associations with adjusted FDR p-value < 0.1 across two cancer types (Tier 1)



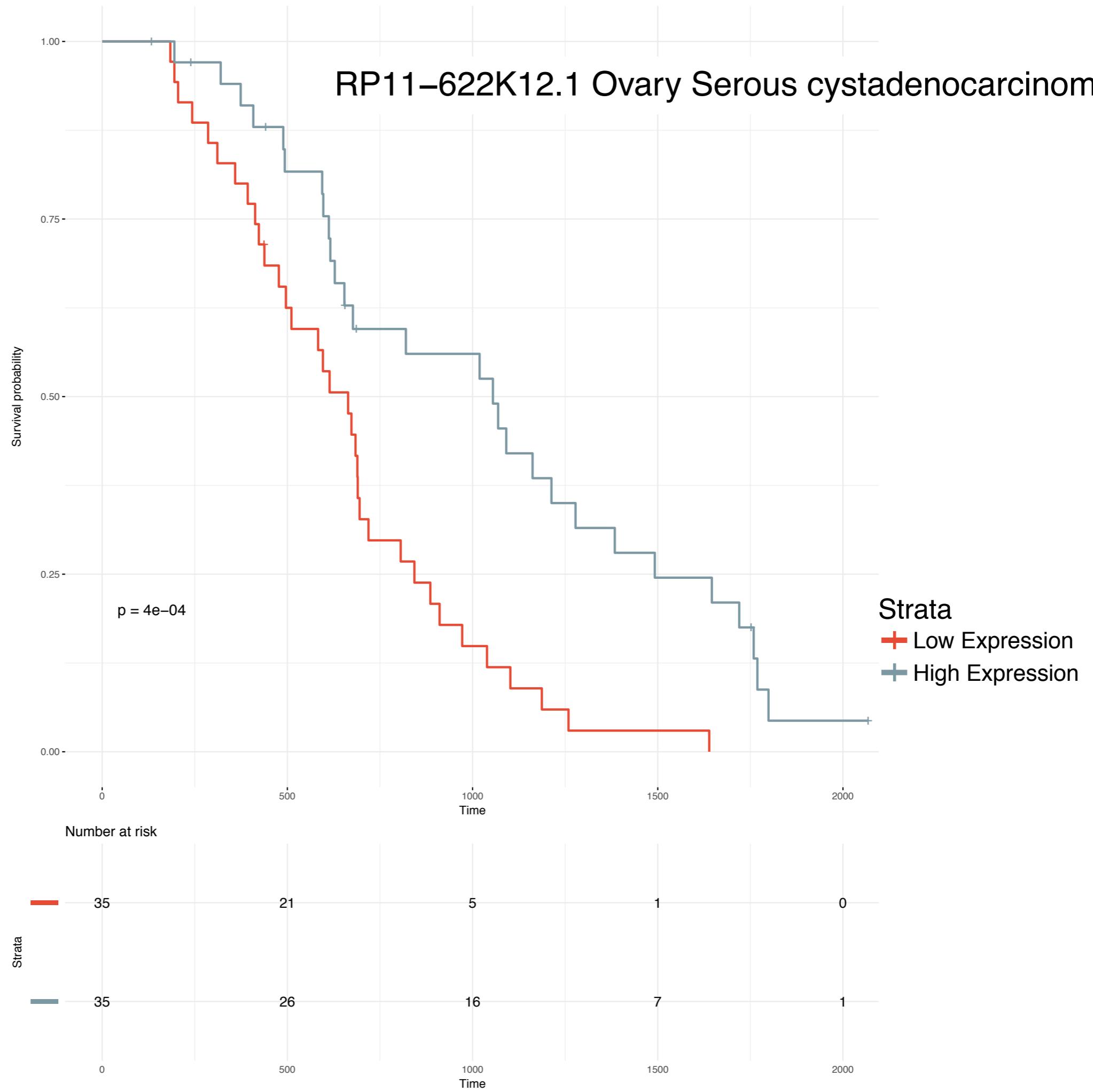


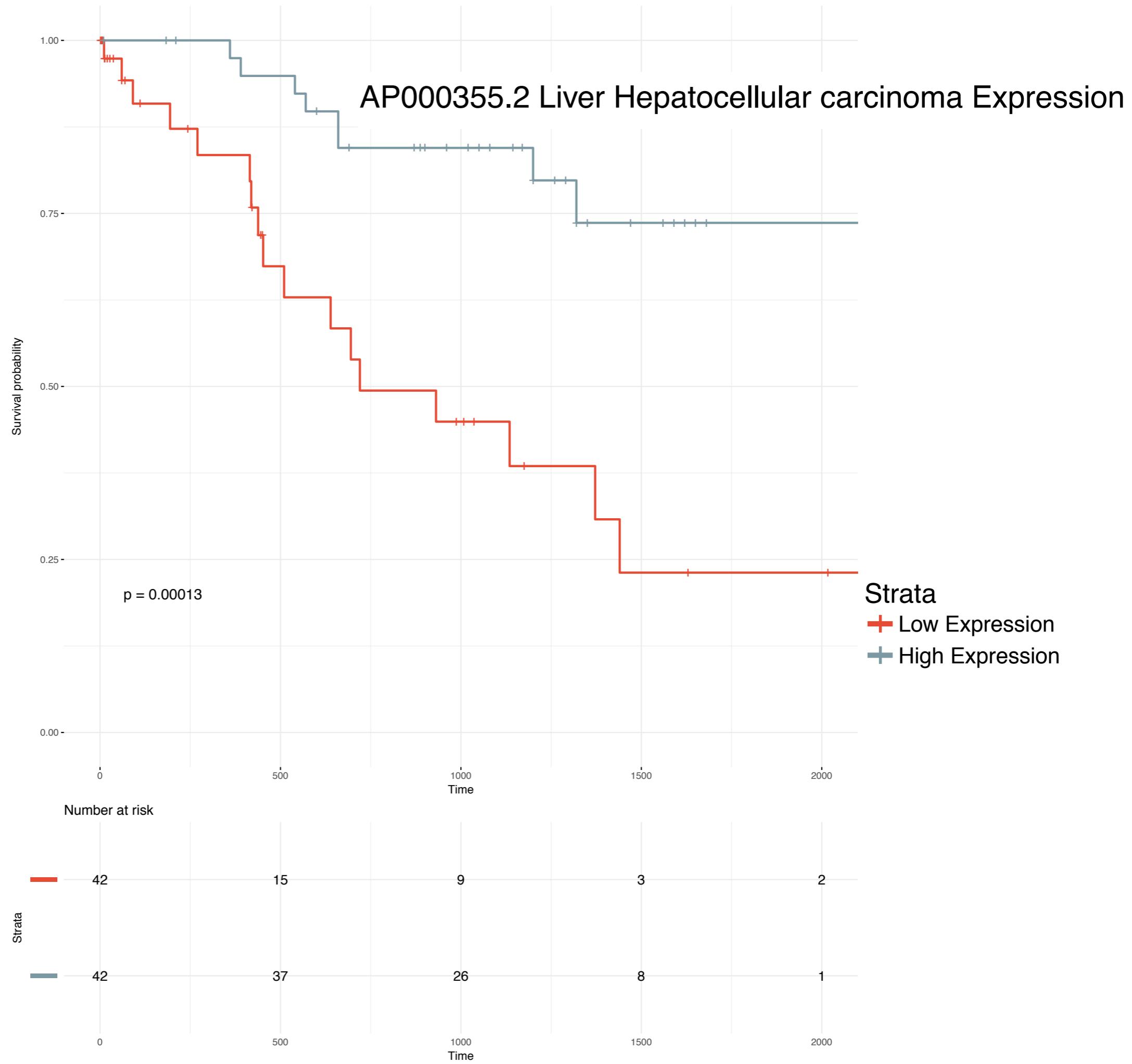


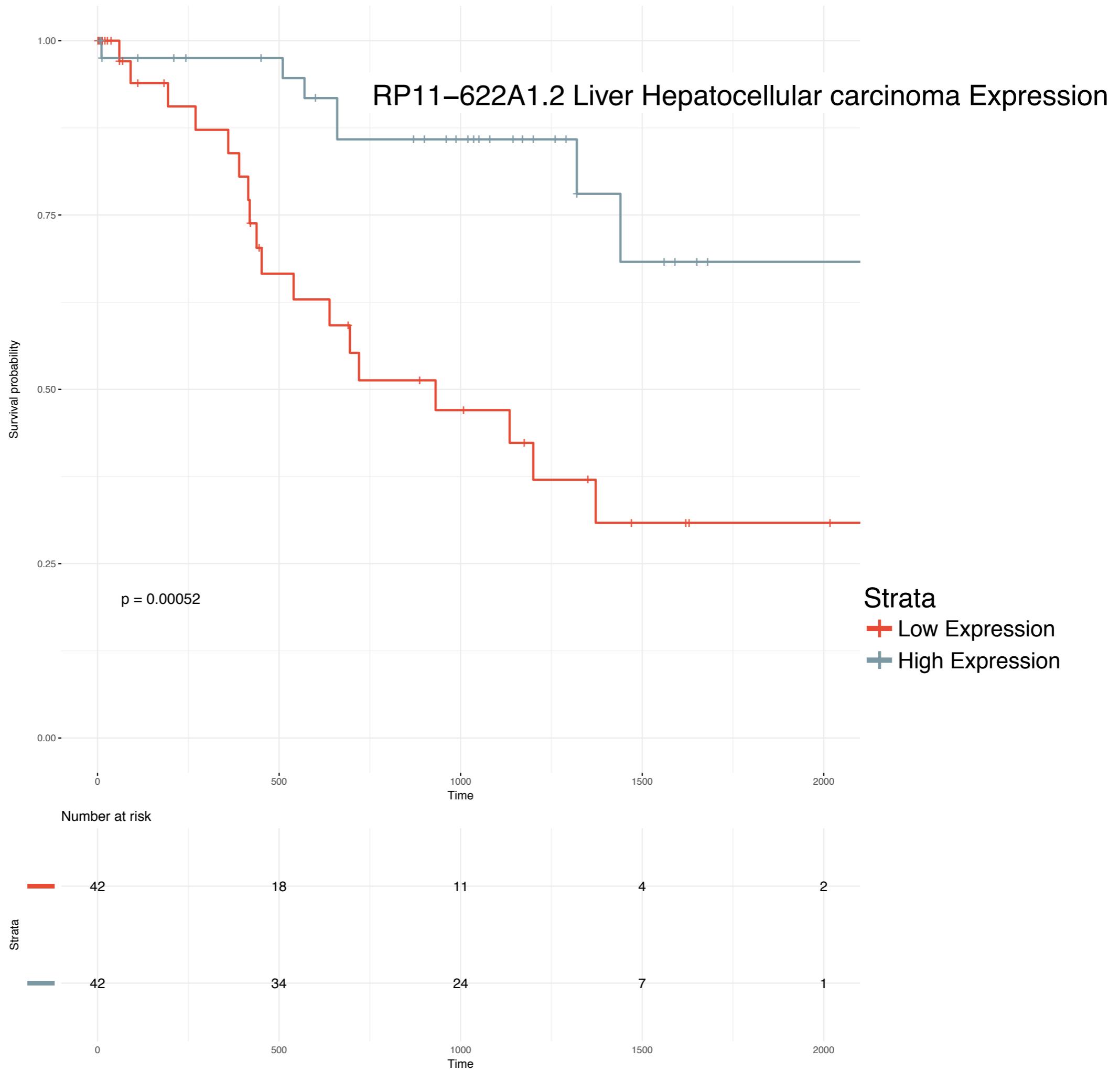
AC009336.24 Ovary Serous cystadenocarcinoma



RP11–622K12.1 Ovary Serous cystadenocarcinoma







Summary of progress

1. Identified high expressing lncRNA candidates in multiple cancer types
2. Survival analysis of lncRNAs using the Cox Proportional Hazards model
3. Differential expression analysis of candidate lncRNAs between tumours (PCAWG) and normal tissues (GTEx)
4. In depth analysis of two lncRNA candidates in Ovarian cancer

Summary of progress

1. Identified high expressing lncRNA candidates in multiple cancer types
2. Survival analysis of lncRNAs using the Cox Proportional Hazards model
3. **Differential expression analysis of candidate lncRNAs between tumours (PCAWG) and normal tissues (GTEx)**
4. In depth analysis of two lncRNA candidates in Ovarian cancer

Rationale:

- To be a meaningful prognostic marker, lncRNA expression should vary between normal and tumour tissues
- To aid in predicting regulatory role in tumour progression
- Hypothesis: lncRNAs that are significantly more abundant in tumour relative to normal tissues are oncogenic
 - Match tumour type to healthy tissue eg. pancreatic PCAWG tumour versus healthy GTEx pancreatic sample

lncRNA expression in tumours versus normal tissues (GTEx)

Within PCAWG (FPKM) and GTEx (RPKM) samples:

Within each patient sample, order all genes based on expression to get rank

1. Least expressed
2. Least expressed
- ⋮
- 25,772. Most expressed
- 25,773. Most expressed

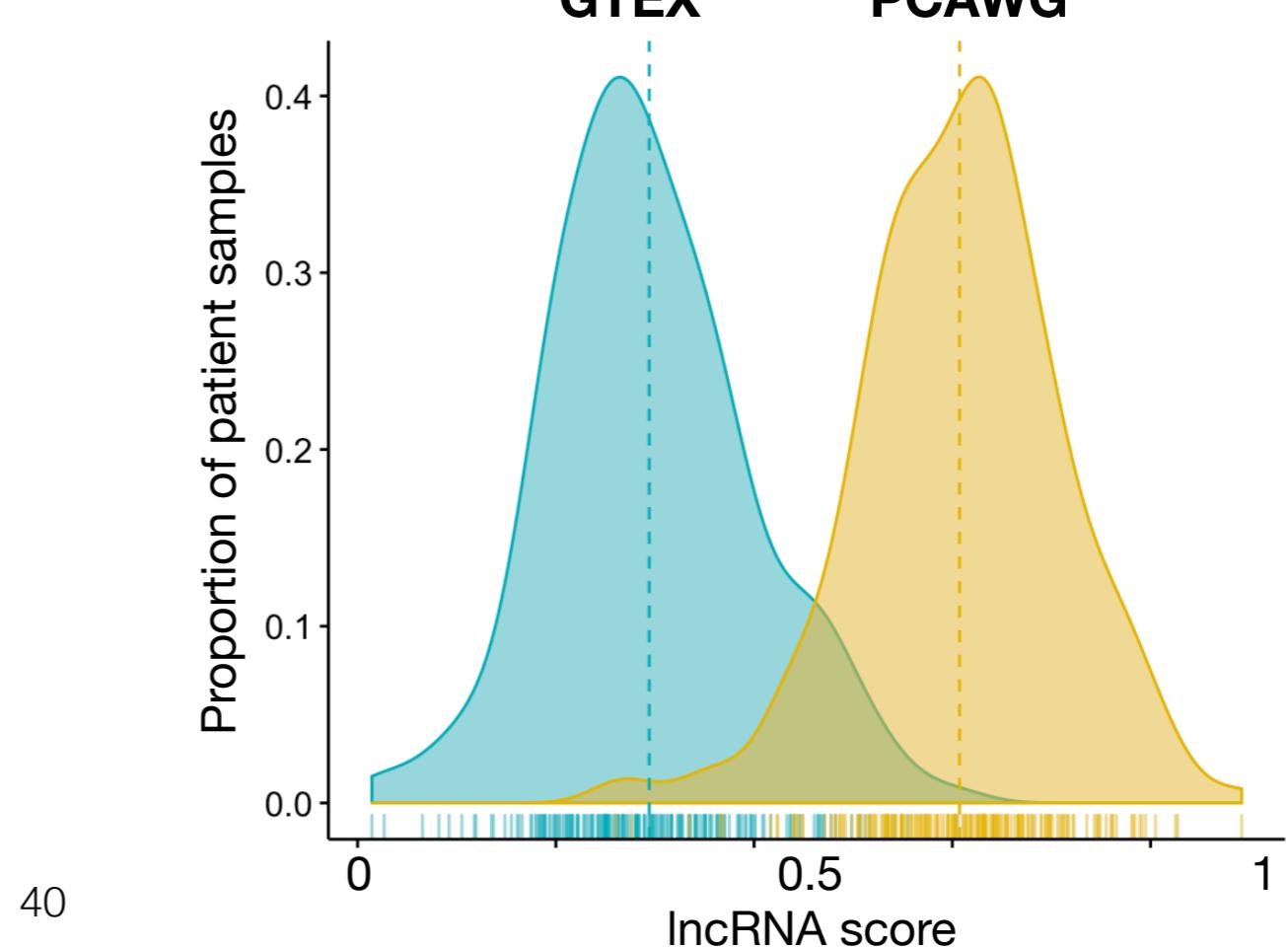
All genes = 20,166 PCGs + 5,607 lncRNAs



Divide rank of gene by length of total genes in list to get score

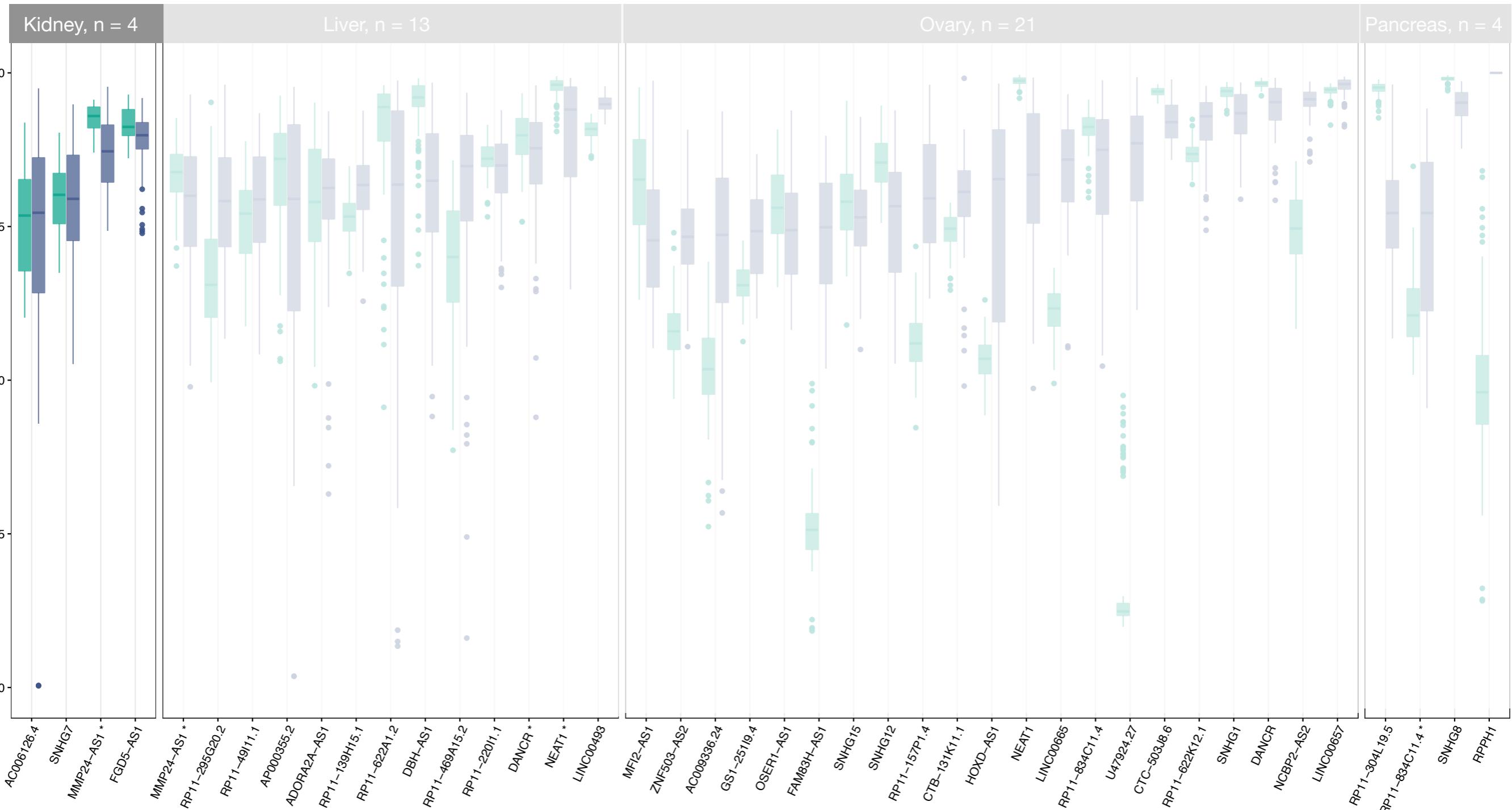


Compare distribution of lncRNA scores between the two datasets



lncRNA expression in tumours versus normal tissues (GTEx)

data  GTEX  PCAWG

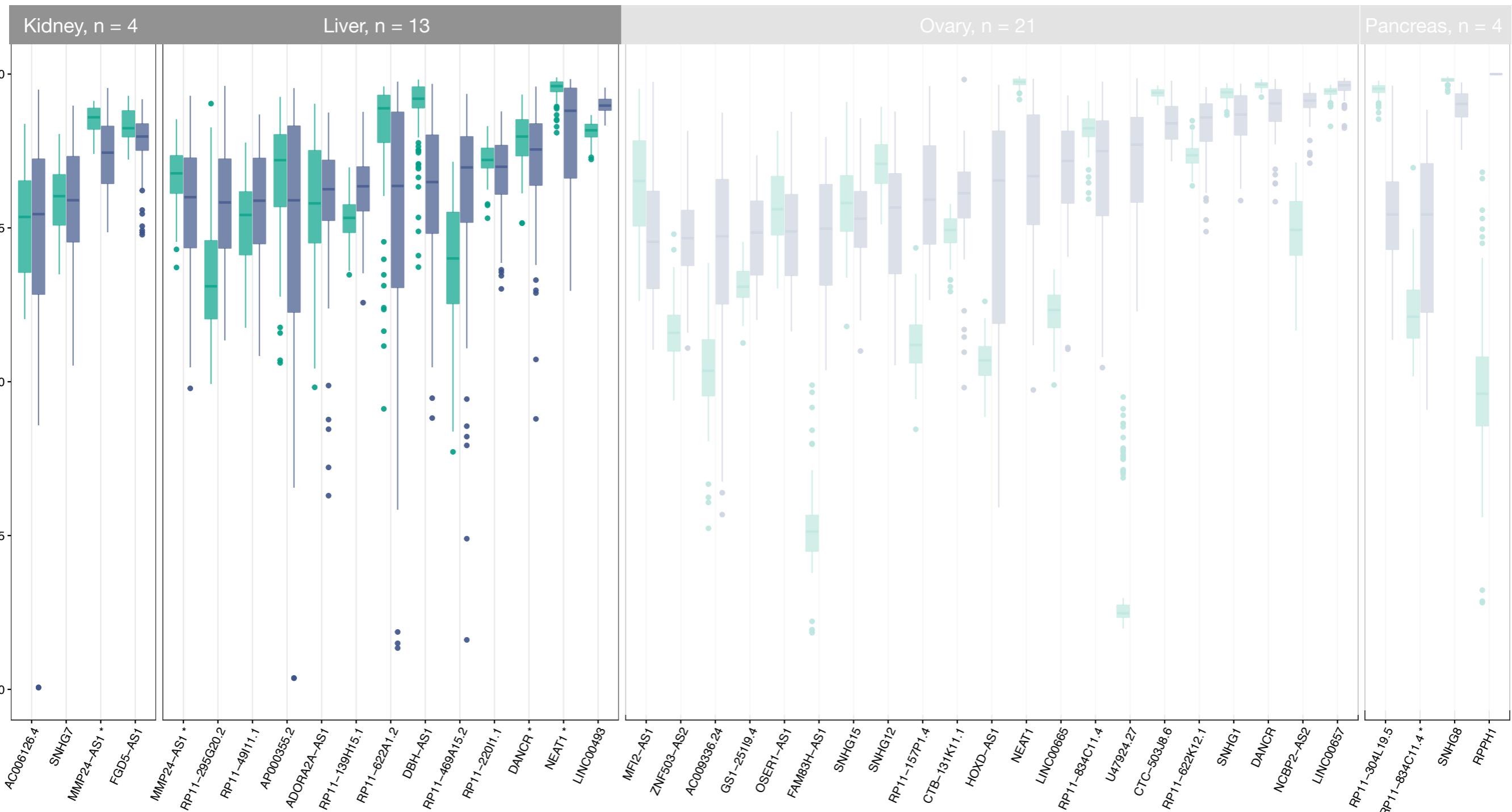


42 lncRNA-cancer associations that had a Cox HR p-val < 0.05

39/42 lncRNAs are differentially expressed, FDR adjusted p-value < 0.05

lncRNA expression in tumours versus normal tissues (GTEx)

data  GTEX  PCAWG

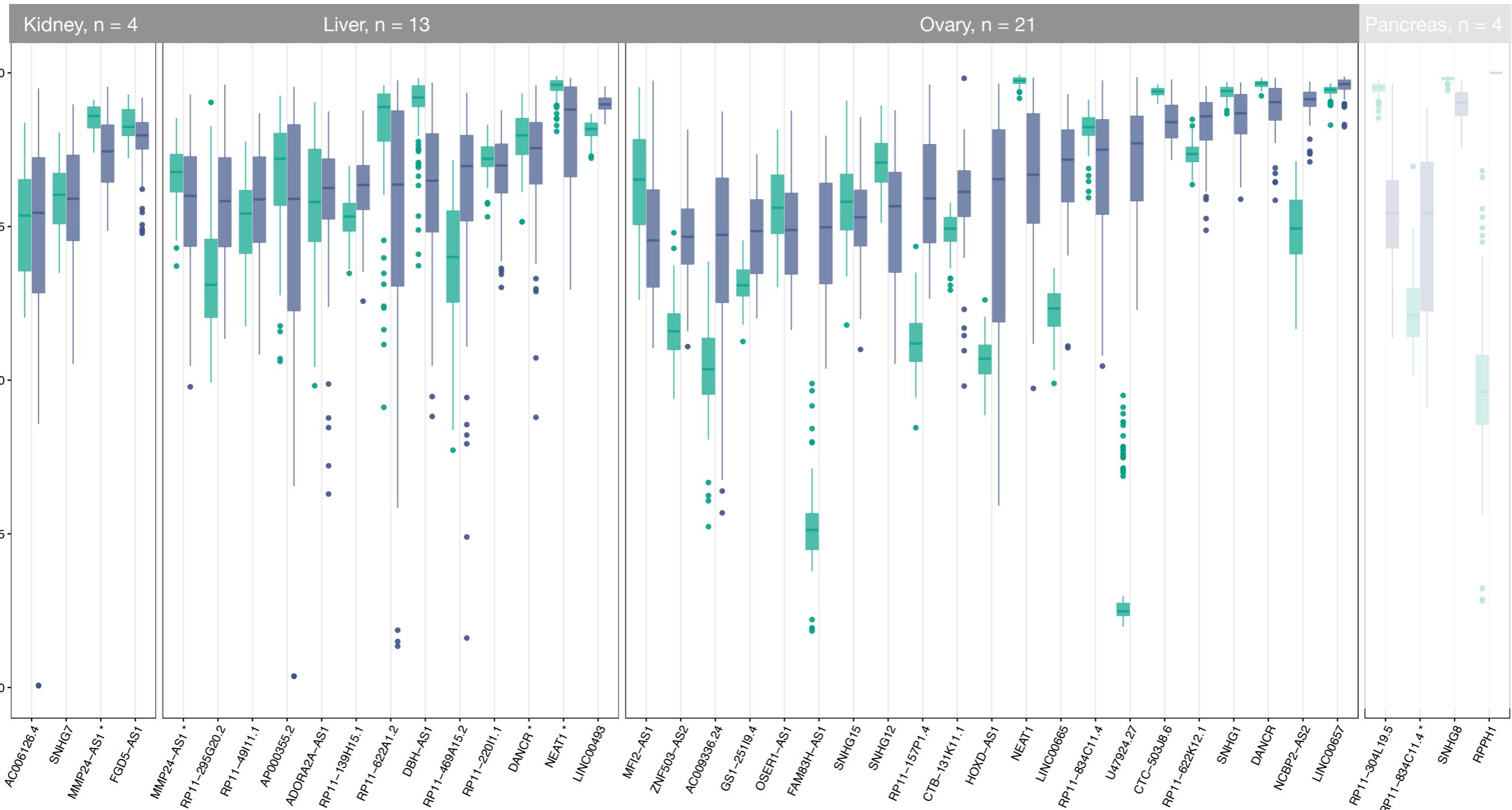


42 lncRNA-cancer associations that had a Cox HR p-val < 0.05

39/42 lncRNAs are differentially expressed, FDR adjusted p-value < 0.05

lncRNA expression in tumours versus normal tissues (GTEx)

data  GTEX  PCAWG

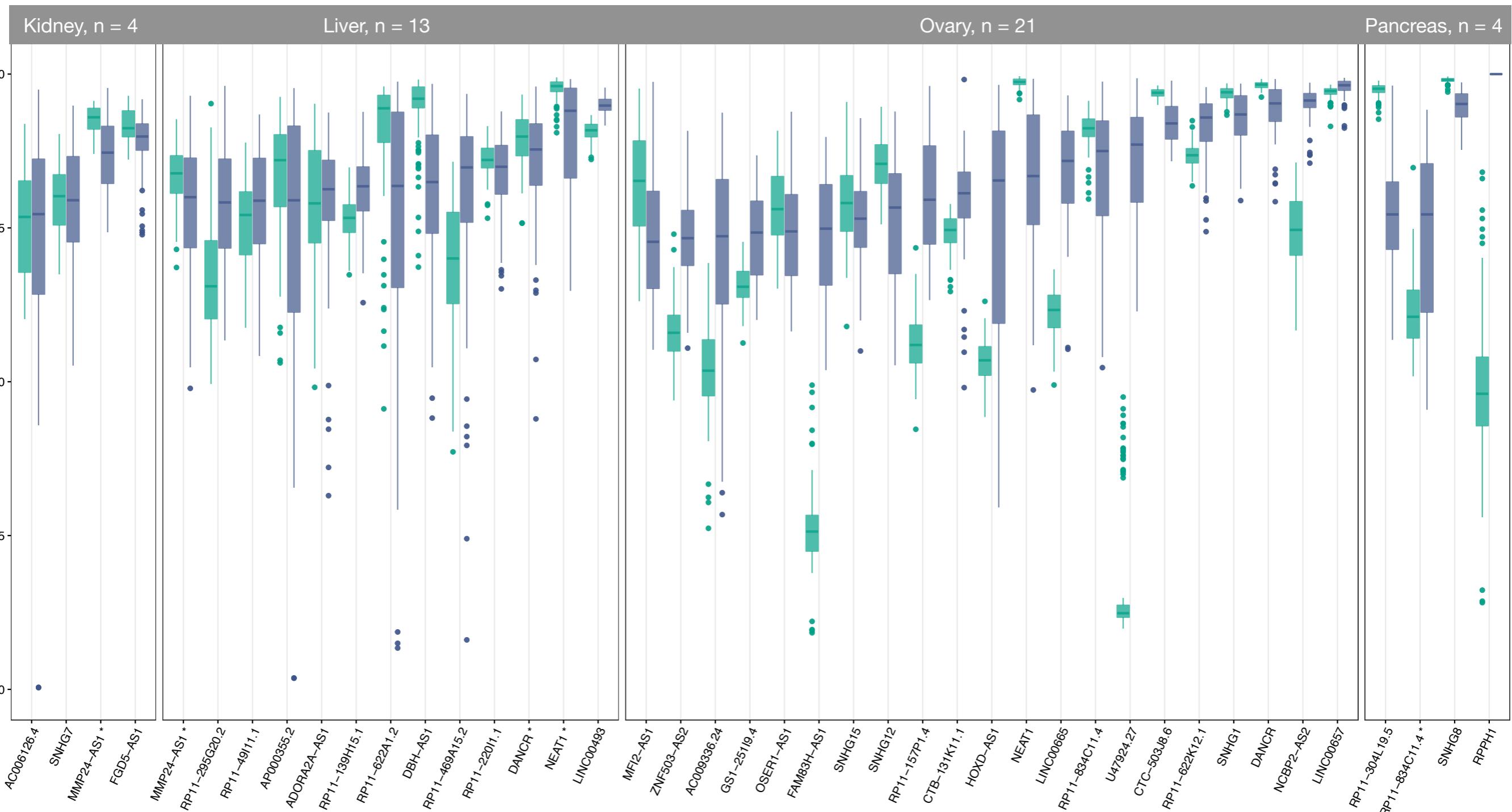


42 lncRNA-cancer associations that had a Cox HR p-val < 0.05

39/42 lncRNAs are differentially expressed, FDR adjusted p-value < 0.05

lncRNA expression in tumours versus normal tissues (GTEx)

data  GTEX  PCAWG

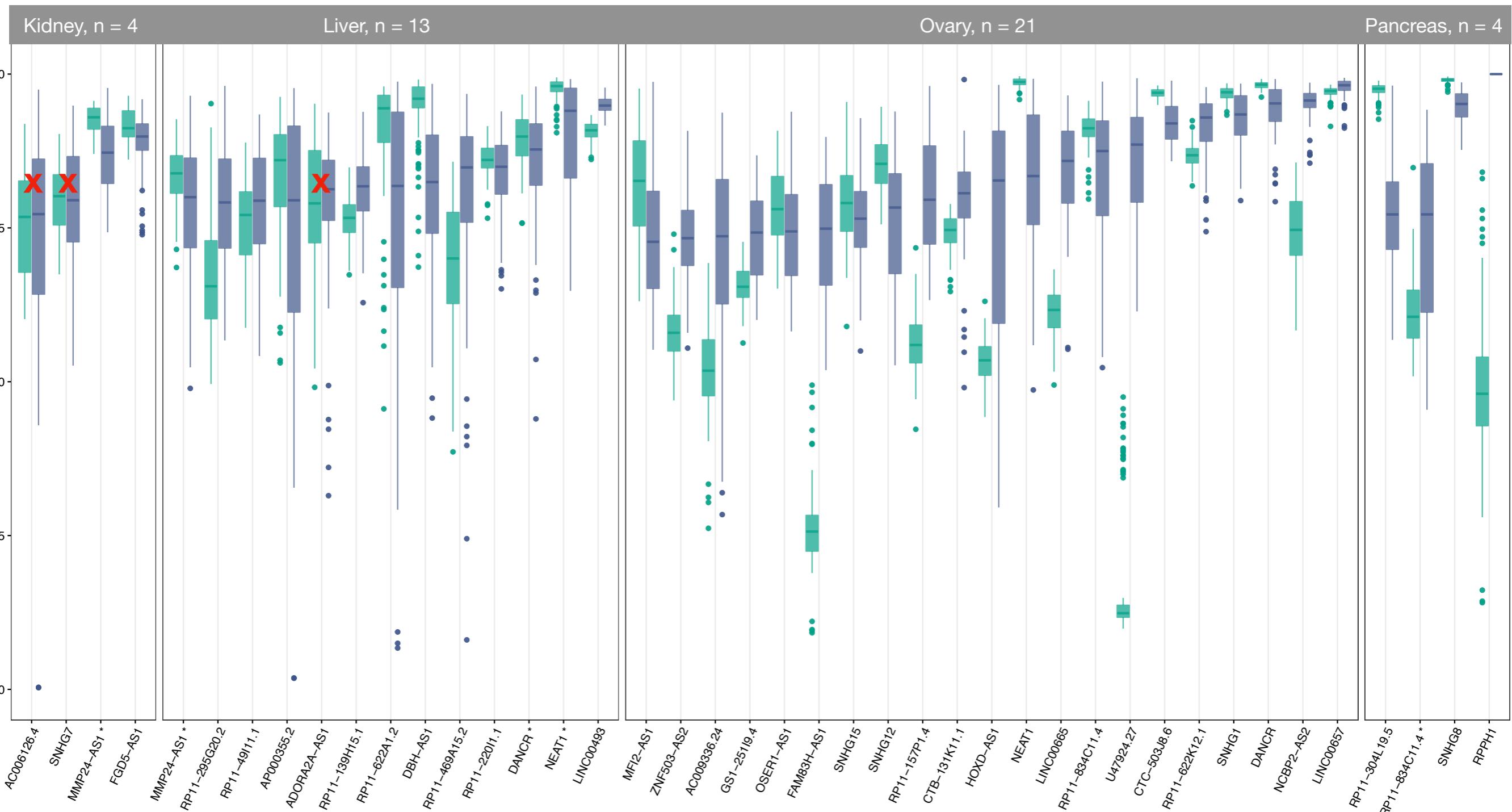


42 lncRNA-cancer associations that had a Cox HR p-val < 0.05

39/42 lncRNAs are differentially expressed, FDR adjusted p-value < 0.05

lncRNA expression in tumours versus normal tissues (GTEx)

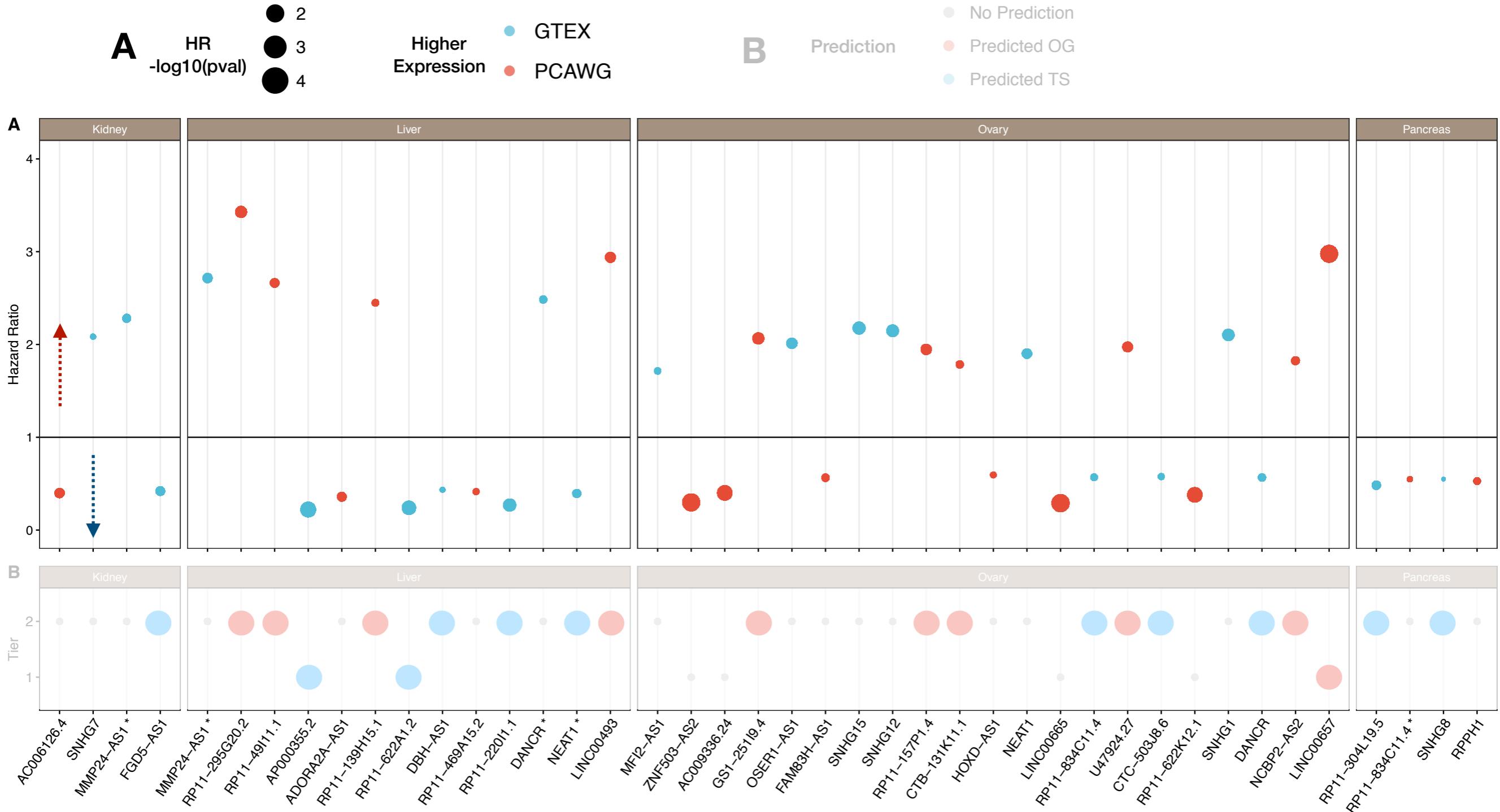
data  GTEX  PCAWG



42 lncRNA-cancer associations that had a Cox HR p-val < 0.05

39/42 lncRNAs are differentially expressed, FDR adjusted p-value < 0.05

Integrating differential expression and survival analysis results

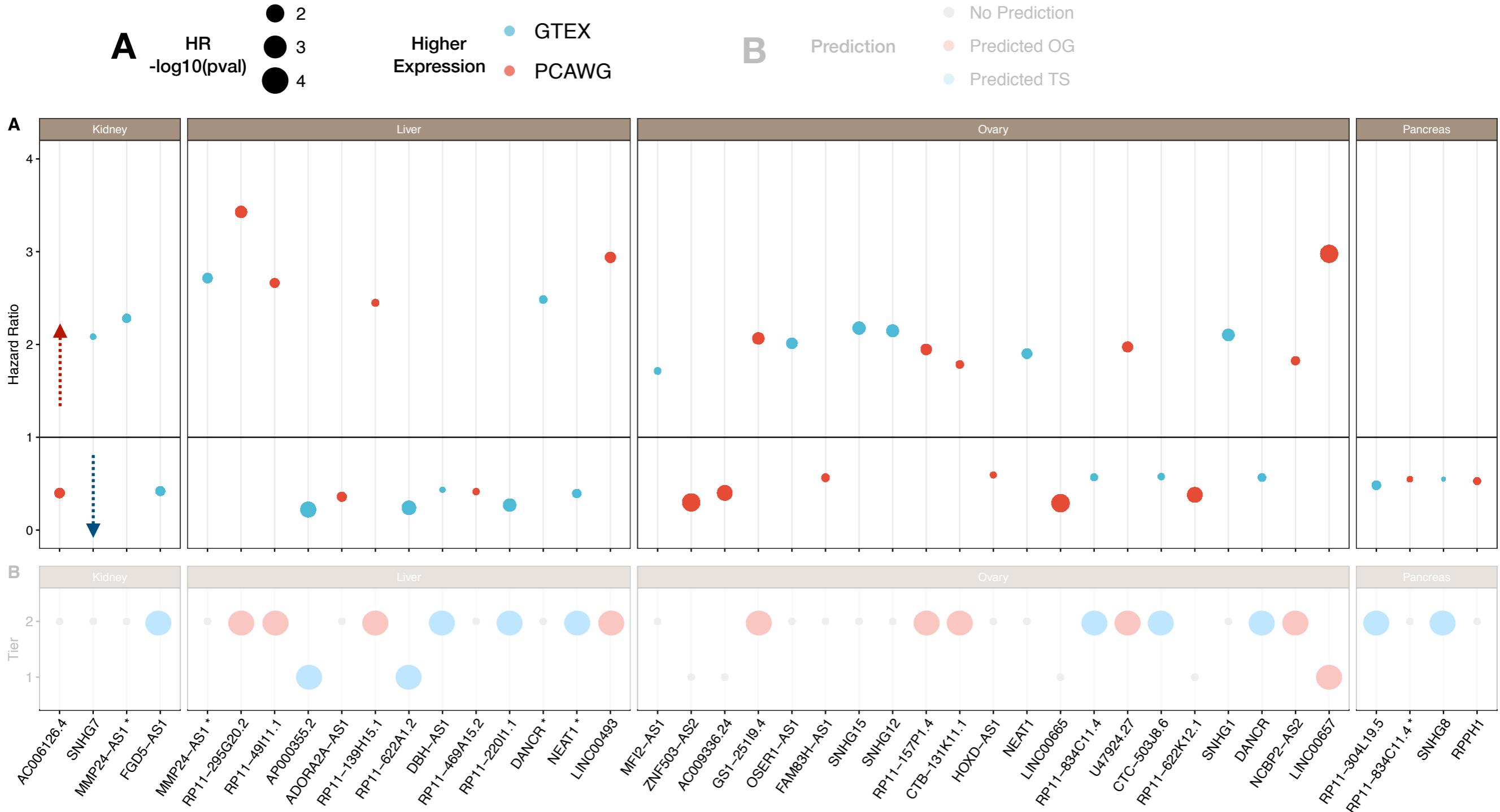


If red in top panel and HR > 1 , predicted oncogene (bottom panel)

If blue in top panel and HR <1, predicted tumour suppressor (bottom panel)

21/42 lncRNA-cancer associations match predicted pattern

Integrating differential expression and survival analysis results

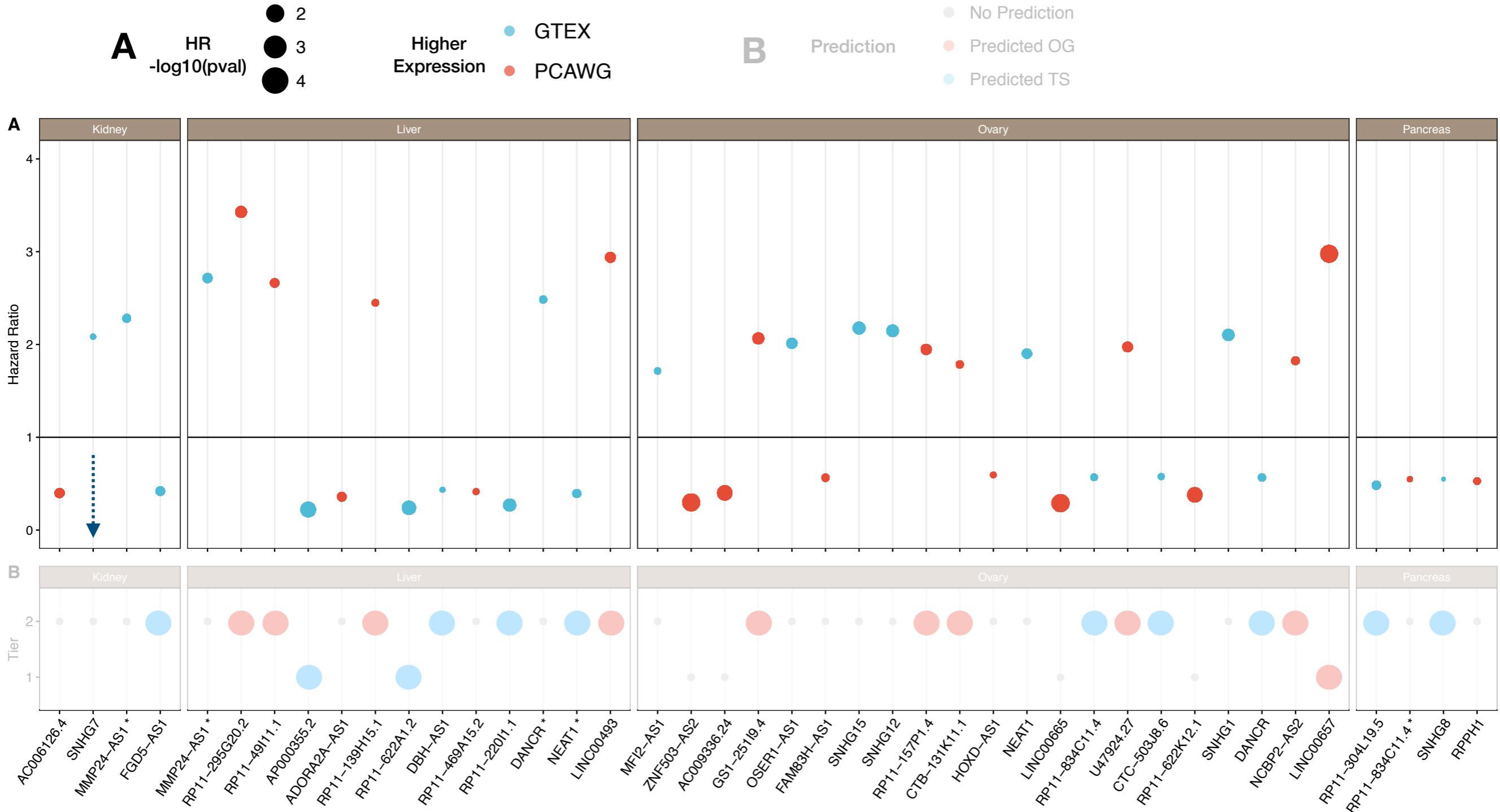


If red in top panel and HR > 1 , predicted oncogene (bottom panel)

If blue in top panel and HR <1, predicted tumour suppressor (bottom panel)

21/42 lncRNA-cancer associations match predicted pattern

Integrating differential expression and survival analysis results

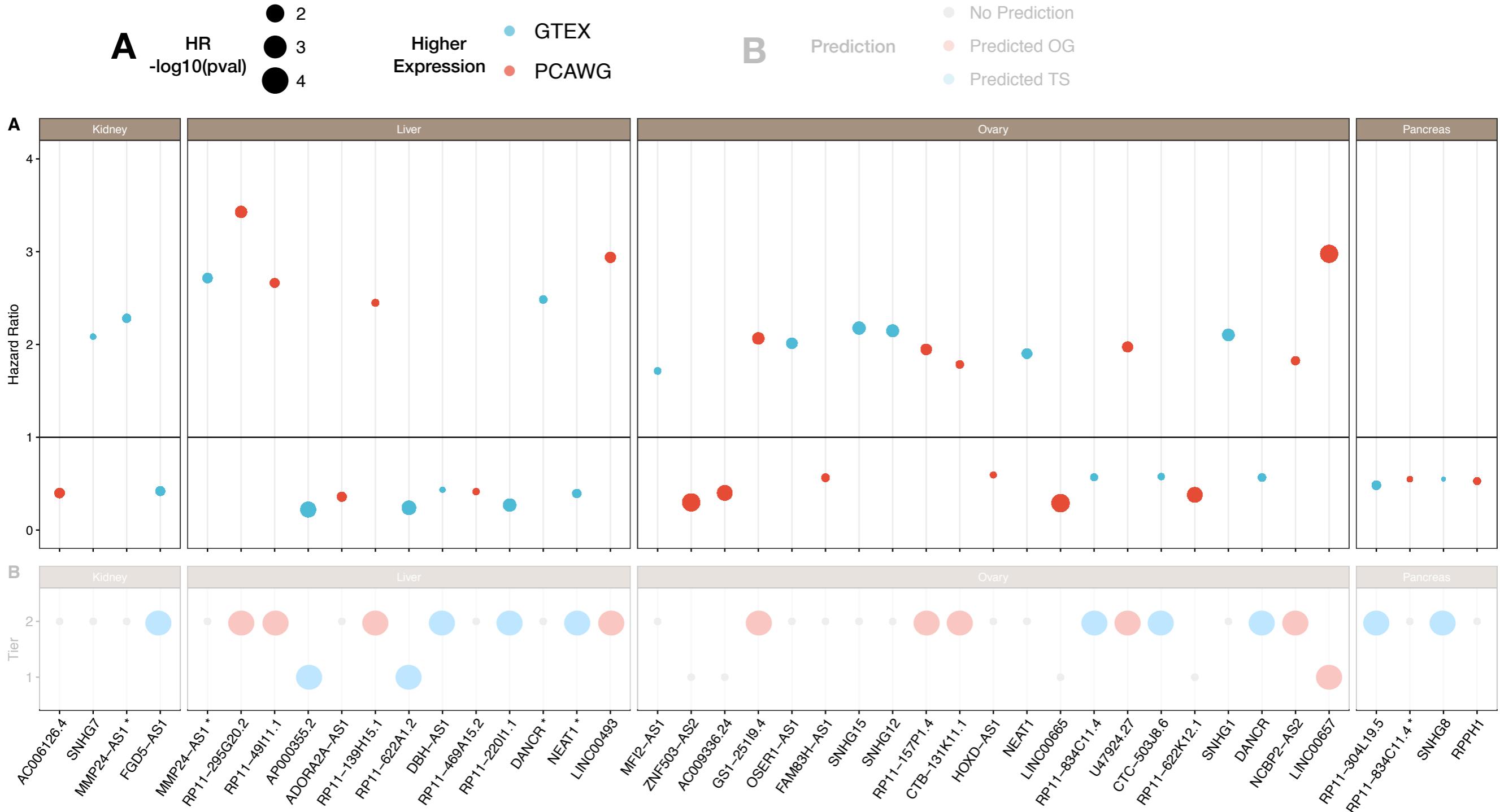


If red in top panel and HR > 1 , predicted oncogene (bottom panel)

If blue in top panel and HR <1, predicted tumour suppressor (bottom panel)

21/42 lncRNA-cancer associations match predicted pattern

Integrating differential expression and survival analysis results

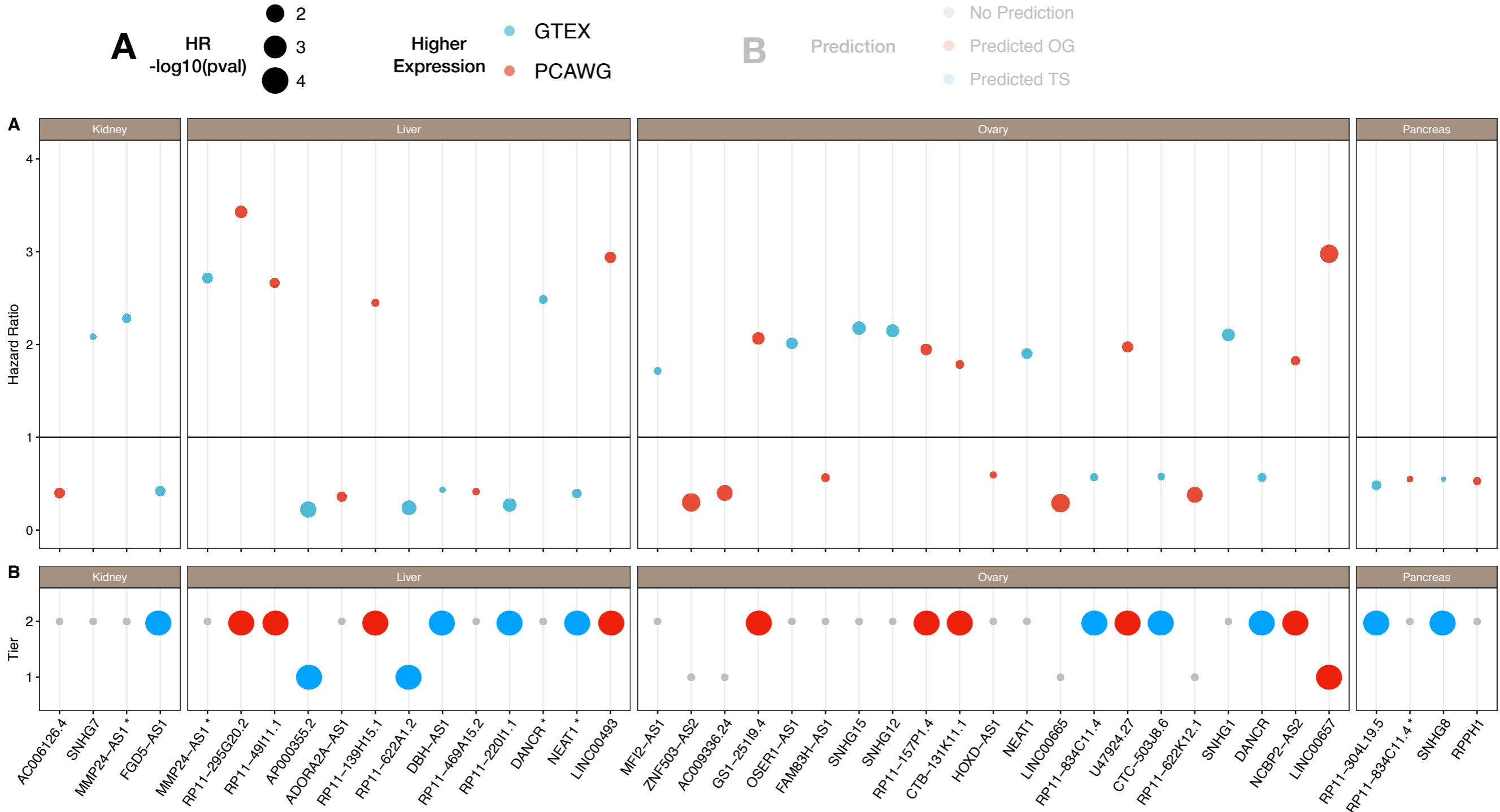


If red in top panel and HR > 1 , predicted oncogene (bottom panel)

If blue in top panel and HR <1, predicted tumour suppressor (bottom panel)

21/42 lncRNA-cancer associations match predicted pattern

Integrating differential expression and survival analysis results

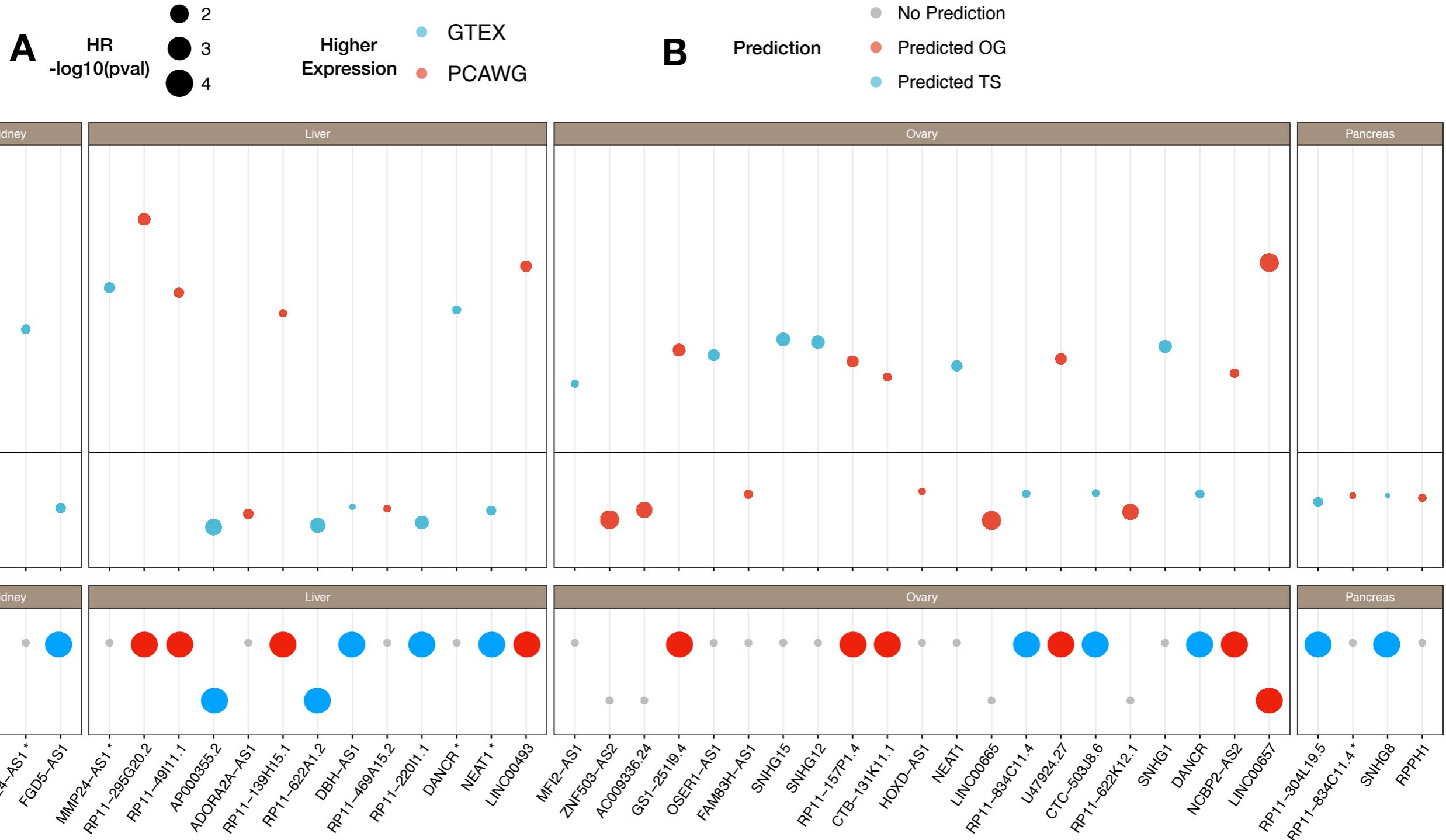


If red in top panel and HR > 1 , predicted oncogene (bottom panel)

If blue in top panel and HR <1, predicted tumour suppressor (bottom panel)

21/42 lncRNA-cancer associations match predicted pattern

Integrating differential expression and survival analysis results



If red in top panel and HR > 1 , predicted oncogene (bottom panel)

If blue in top panel and HR <1, predicted tumour suppressor (bottom panel)

21/42 IncRNA-cancer associations match predicted pattern

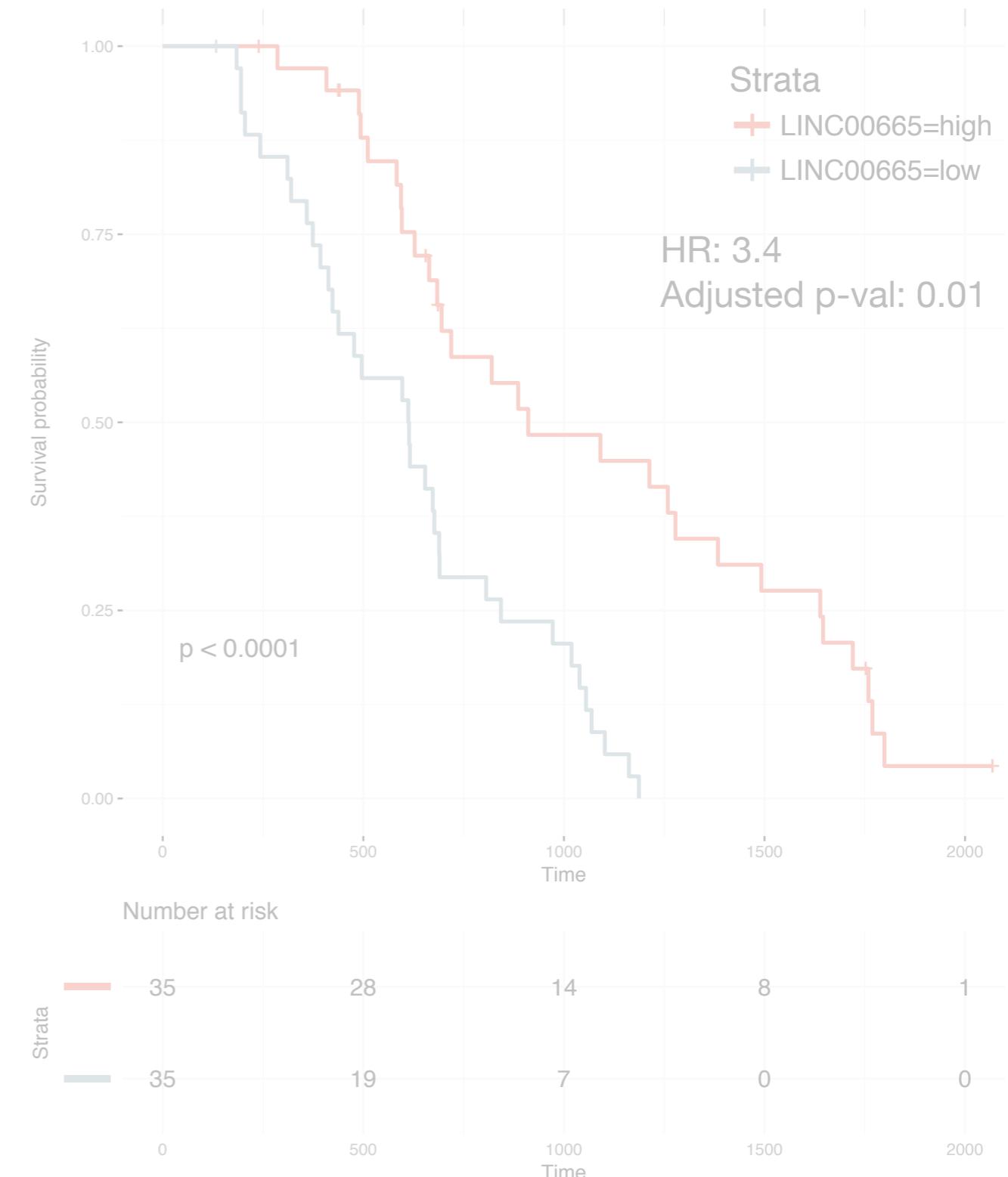
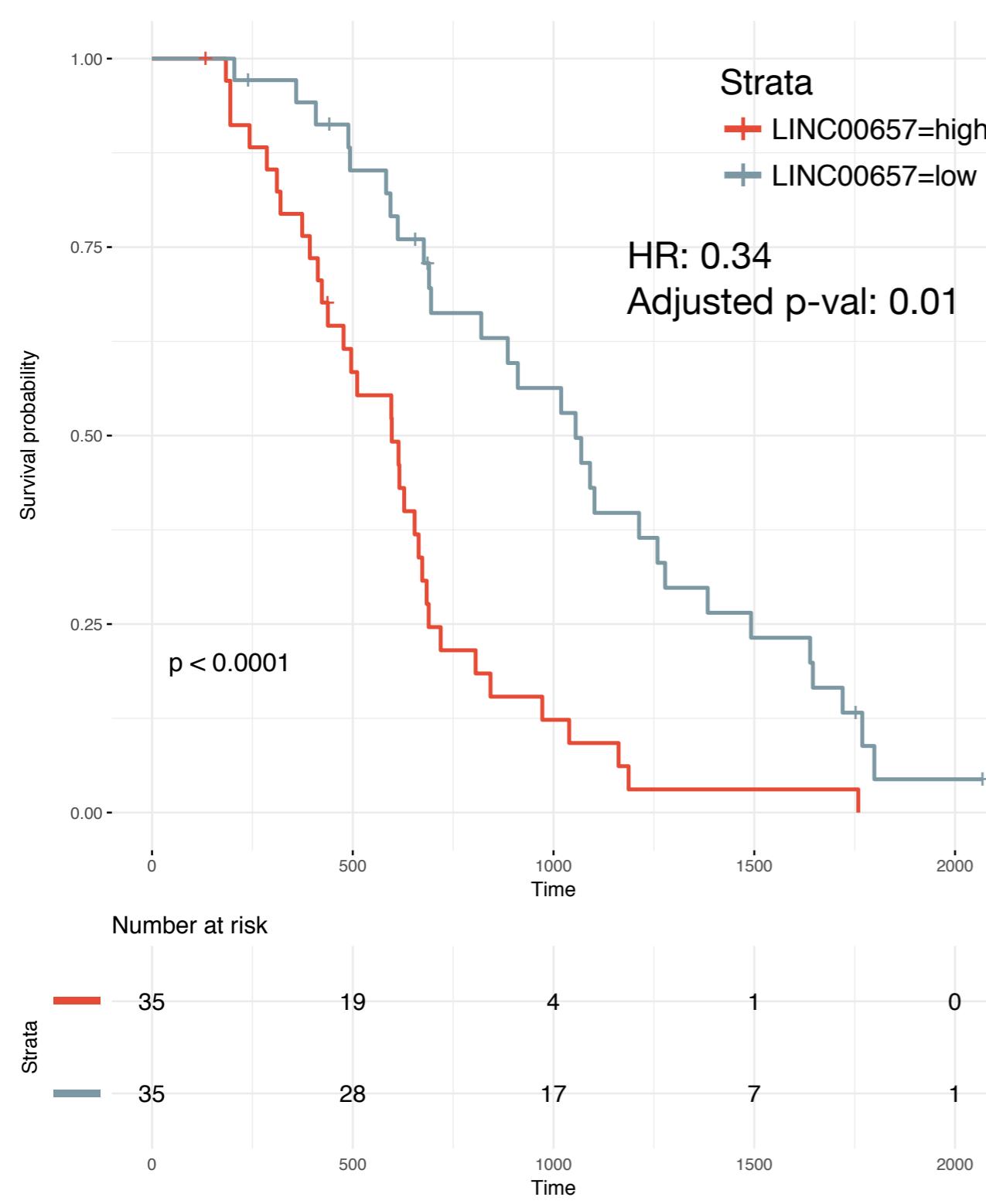
Summary of progress

1. Identified high expressing lncRNA candidates in multiple cancer types
2. Survival analysis of lncRNAs using the Cox Proportional Hazards model
3. Differential expression analysis of candidate lncRNAs between tumours (PCAWG) and normal tissues (GTEx)
4. In depth analysis of two lncRNA candidates in Ovarian cancer

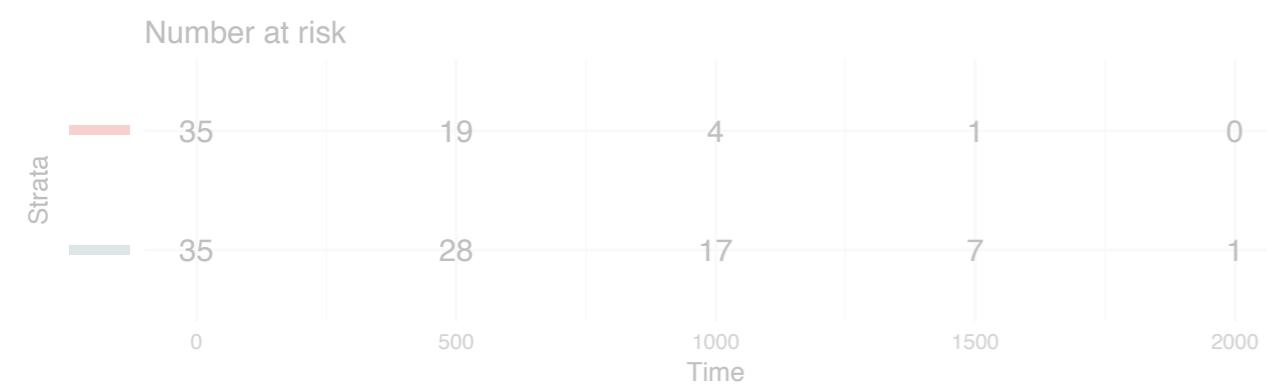
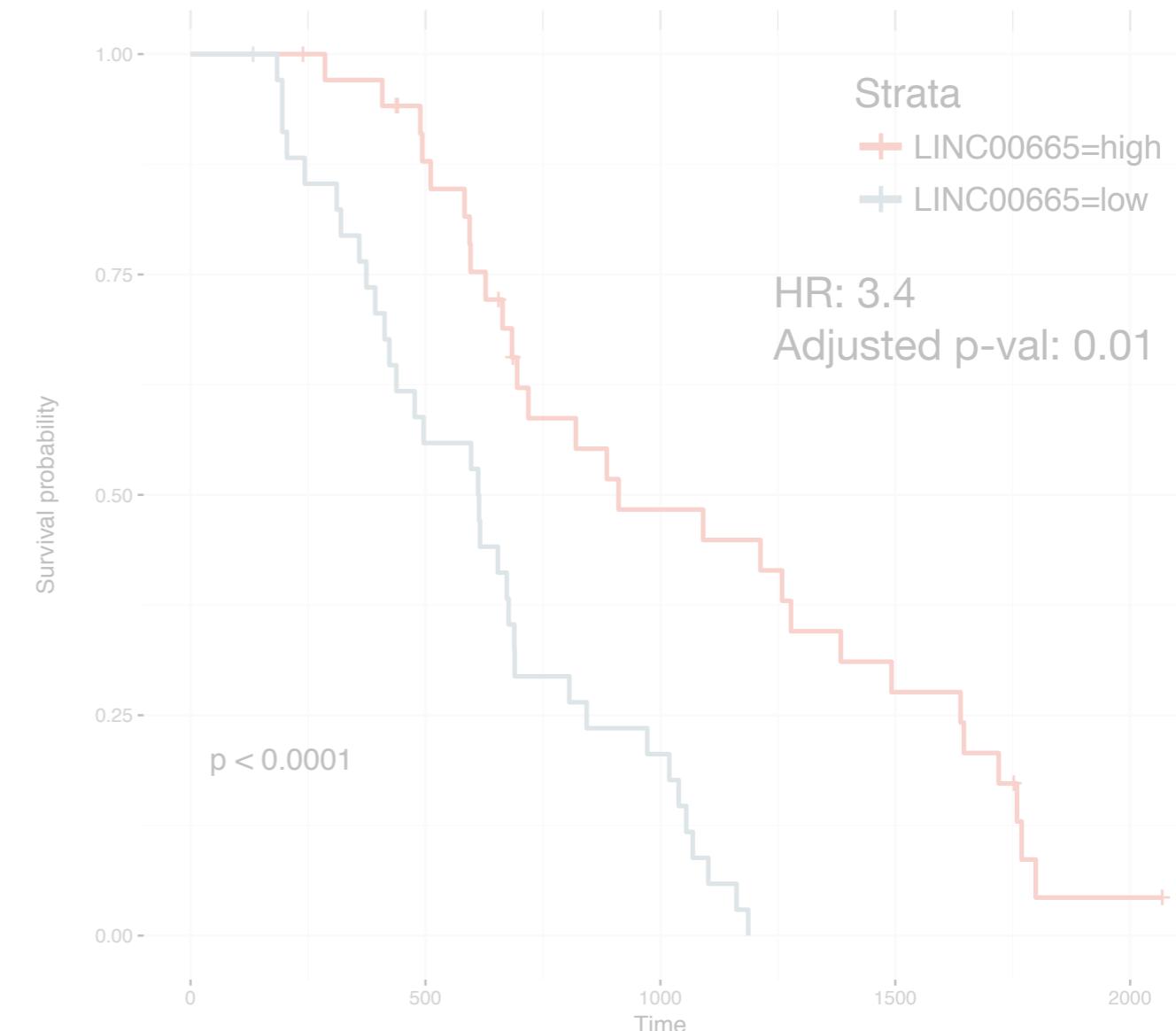
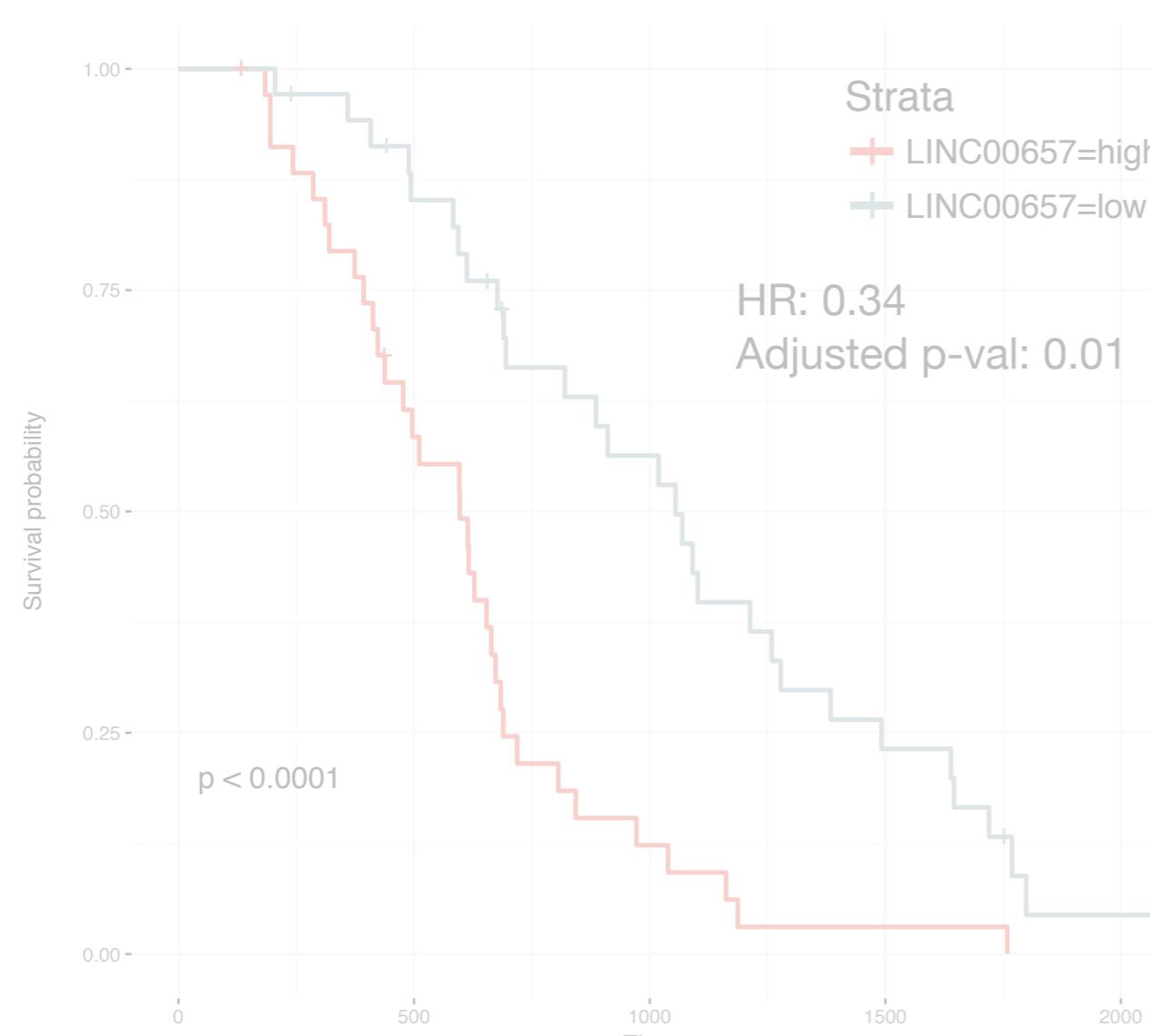
Summary of progress

1. Identified high expressing lncRNA candidates in multiple cancer types
2. Survival analysis of lncRNAs using the Cox Proportional Hazards model
3. Differential expression analysis of candidate lncRNAs between tumours (PCAWG) and normal tissues (GTEx)
4. **In depth analysis of two lncRNA candidates in Ovarian cancer**

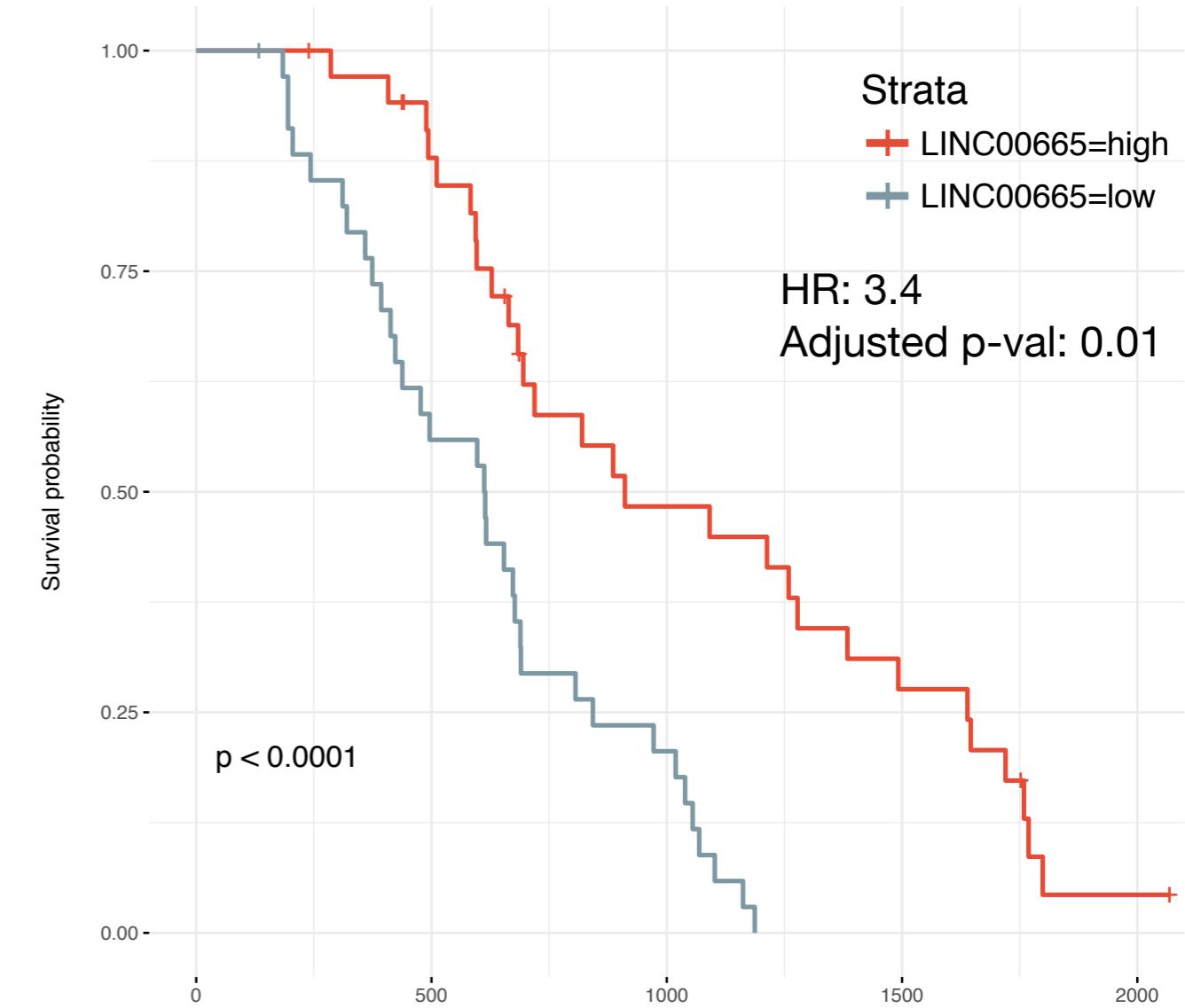
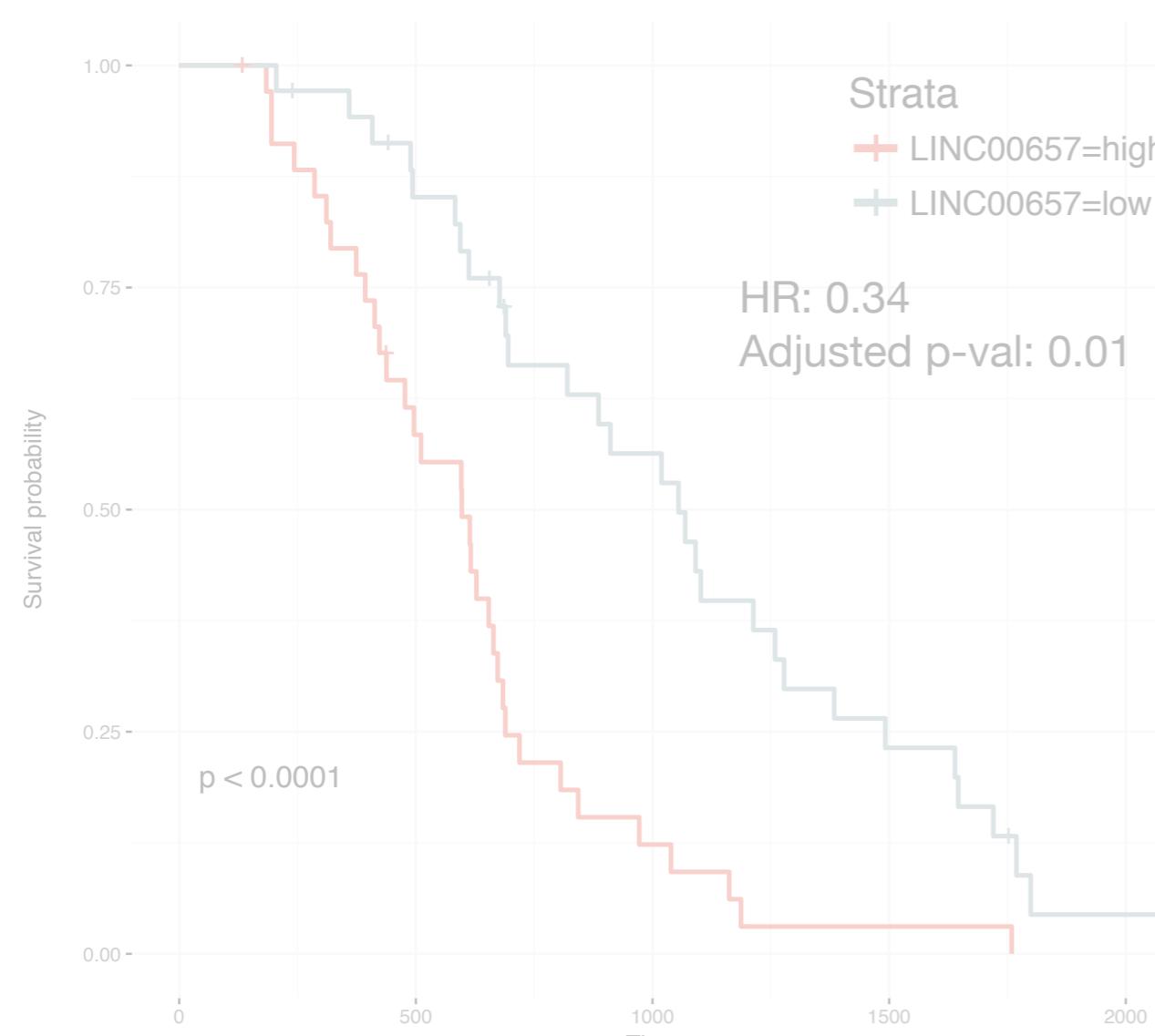
Survival curves for top two lncRNA candidates in Ovarian cancer



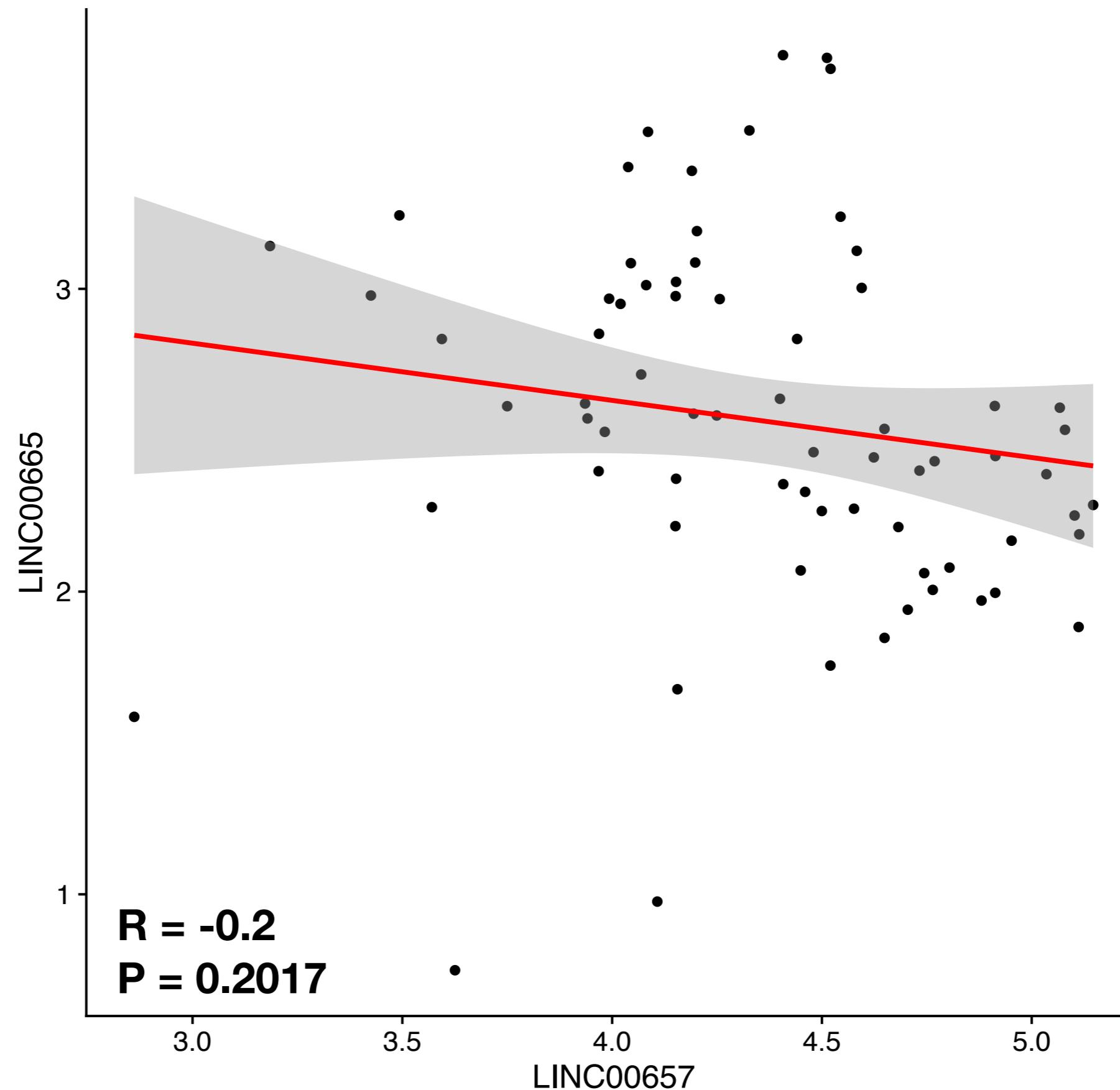
Survival curves for top two lncRNA candidates in Ovarian cancer



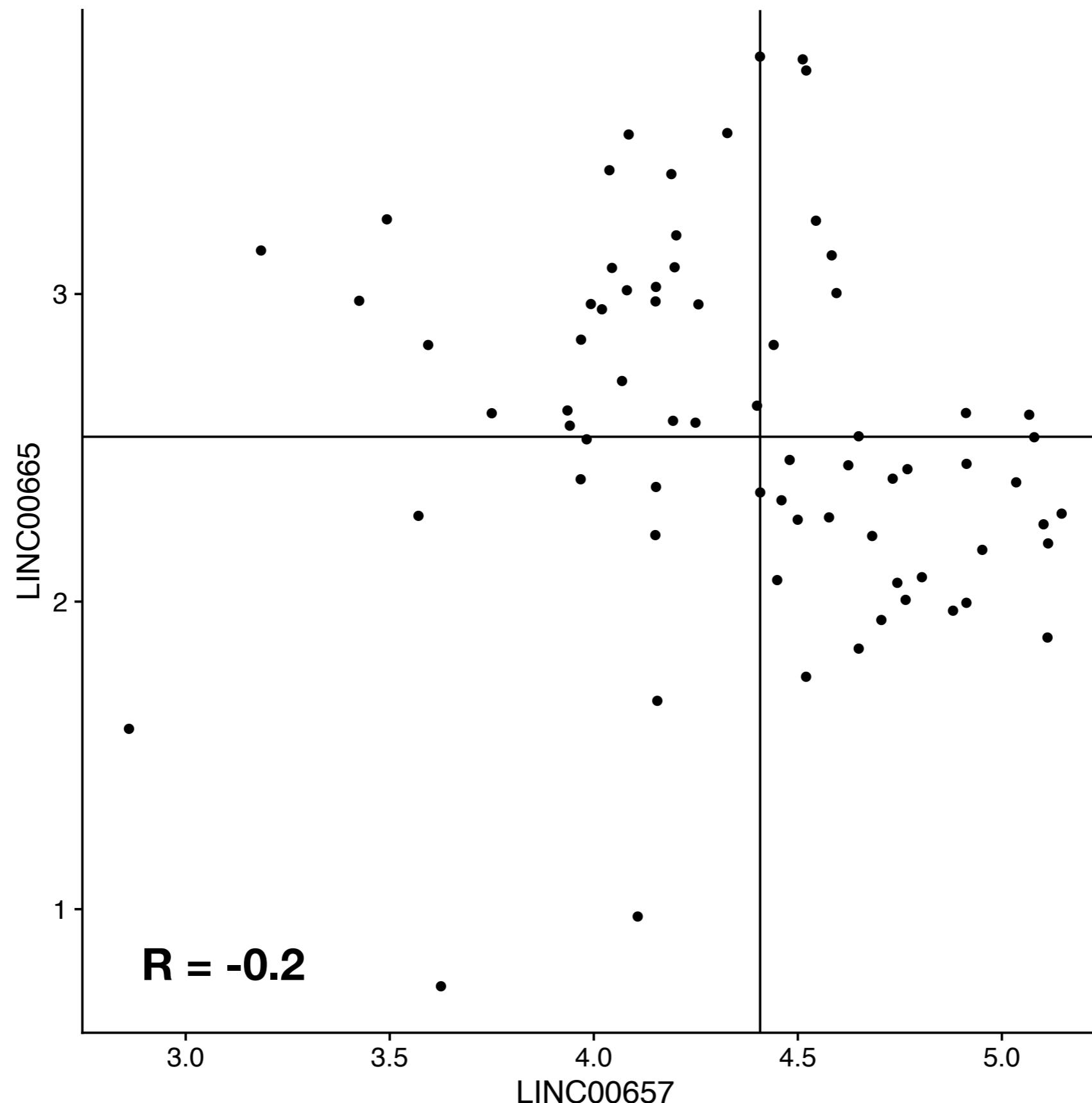
Survival curves for top two lncRNA candidates in Ovarian cancer



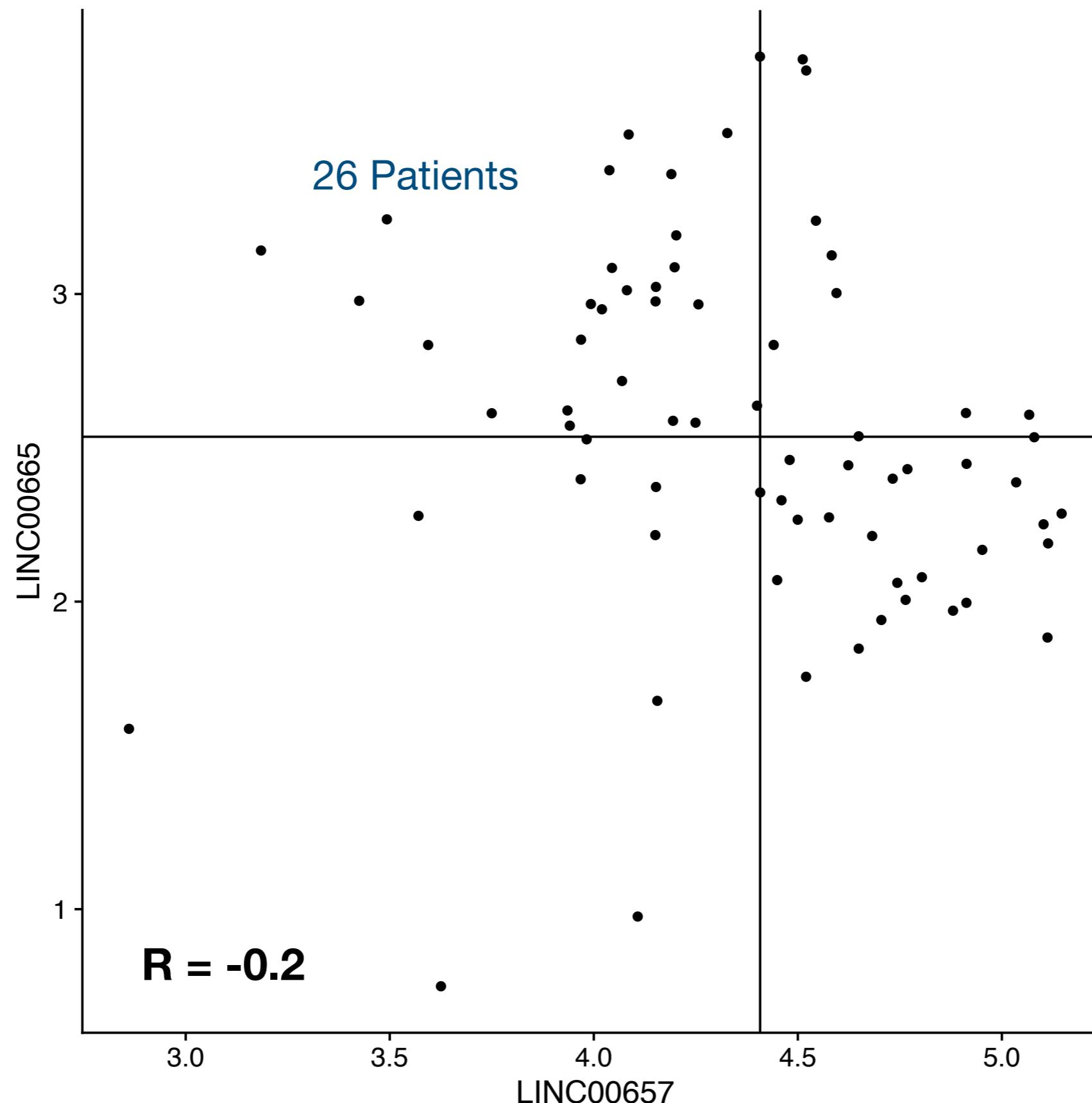
Correlation of LINC00657 and LINC00665 in ovarian cancer



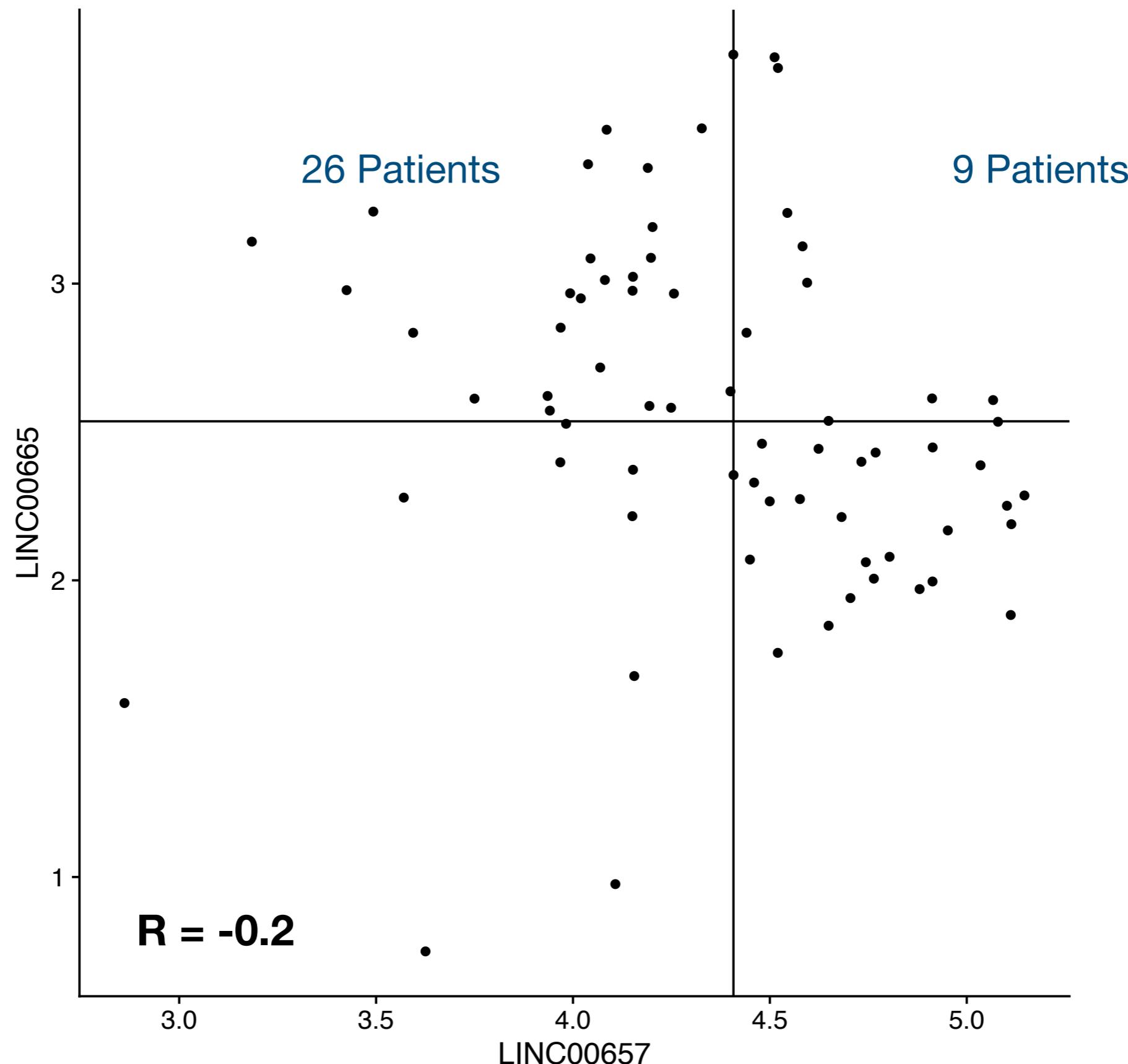
Correlation of LINC00657 and LINC00665 in ovarian cancer



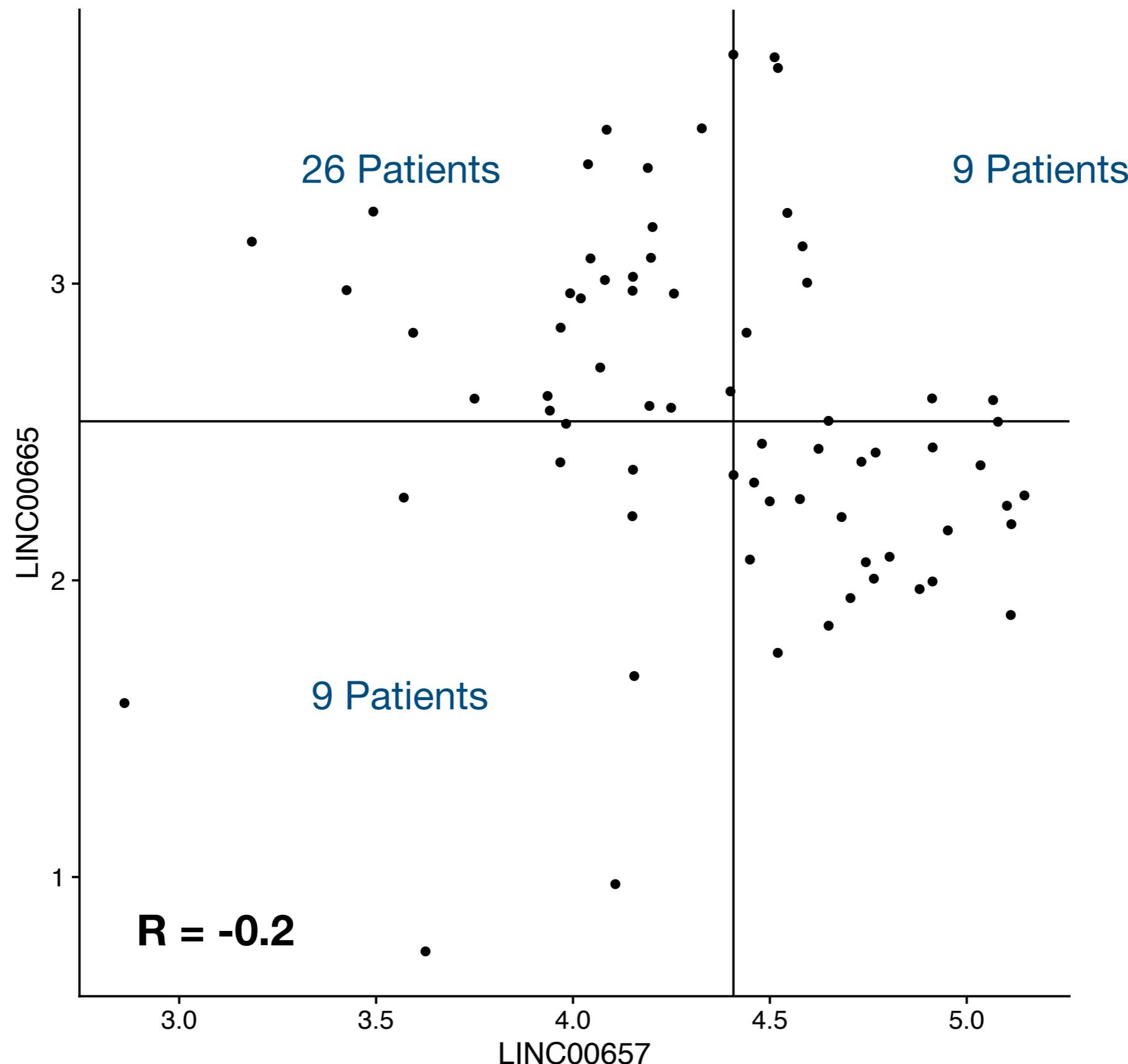
Correlation of LINC00657 and LINC00665 in ovarian cancer



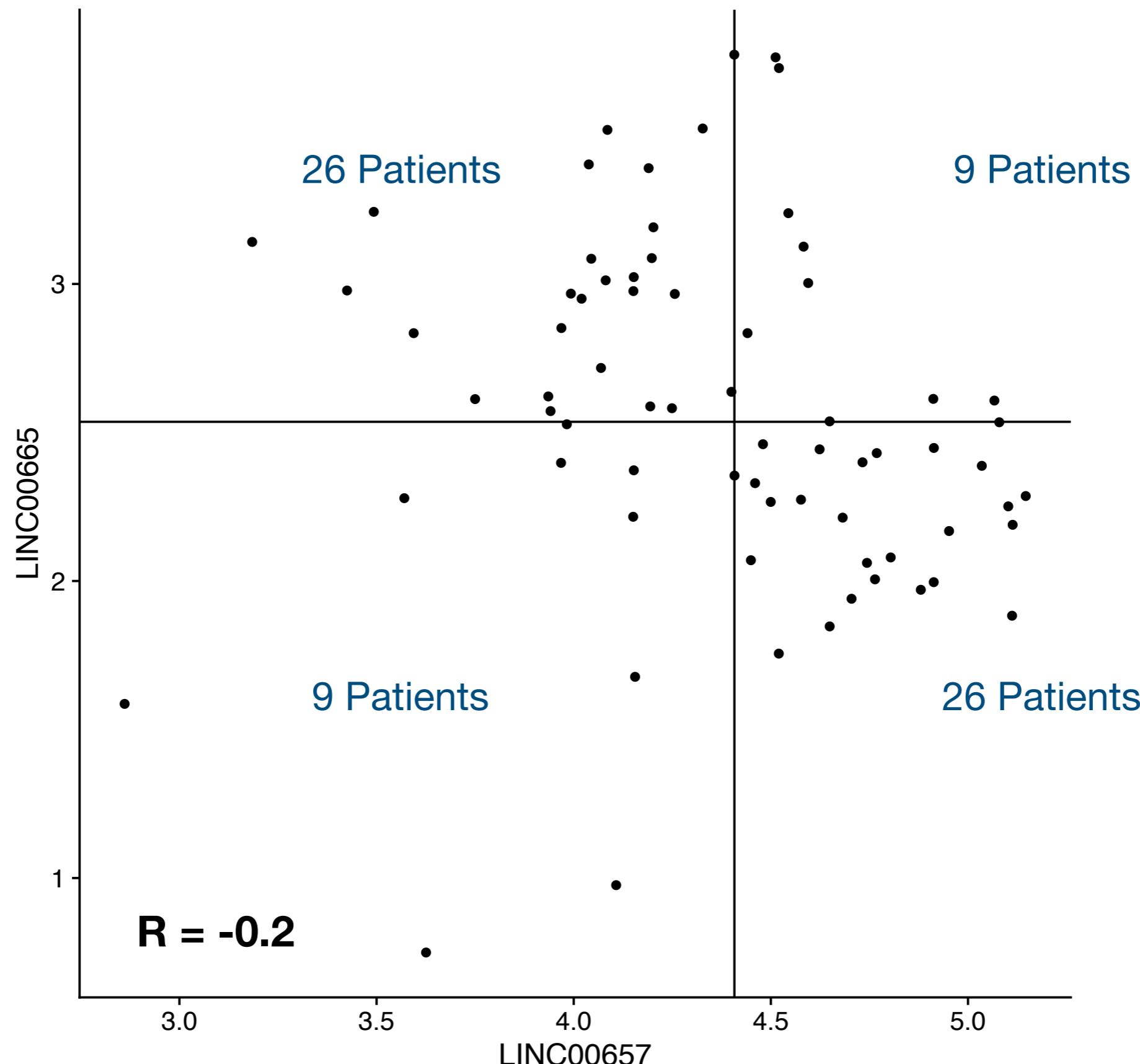
Correlation of LINC00657 and LINC00665 in ovarian cancer



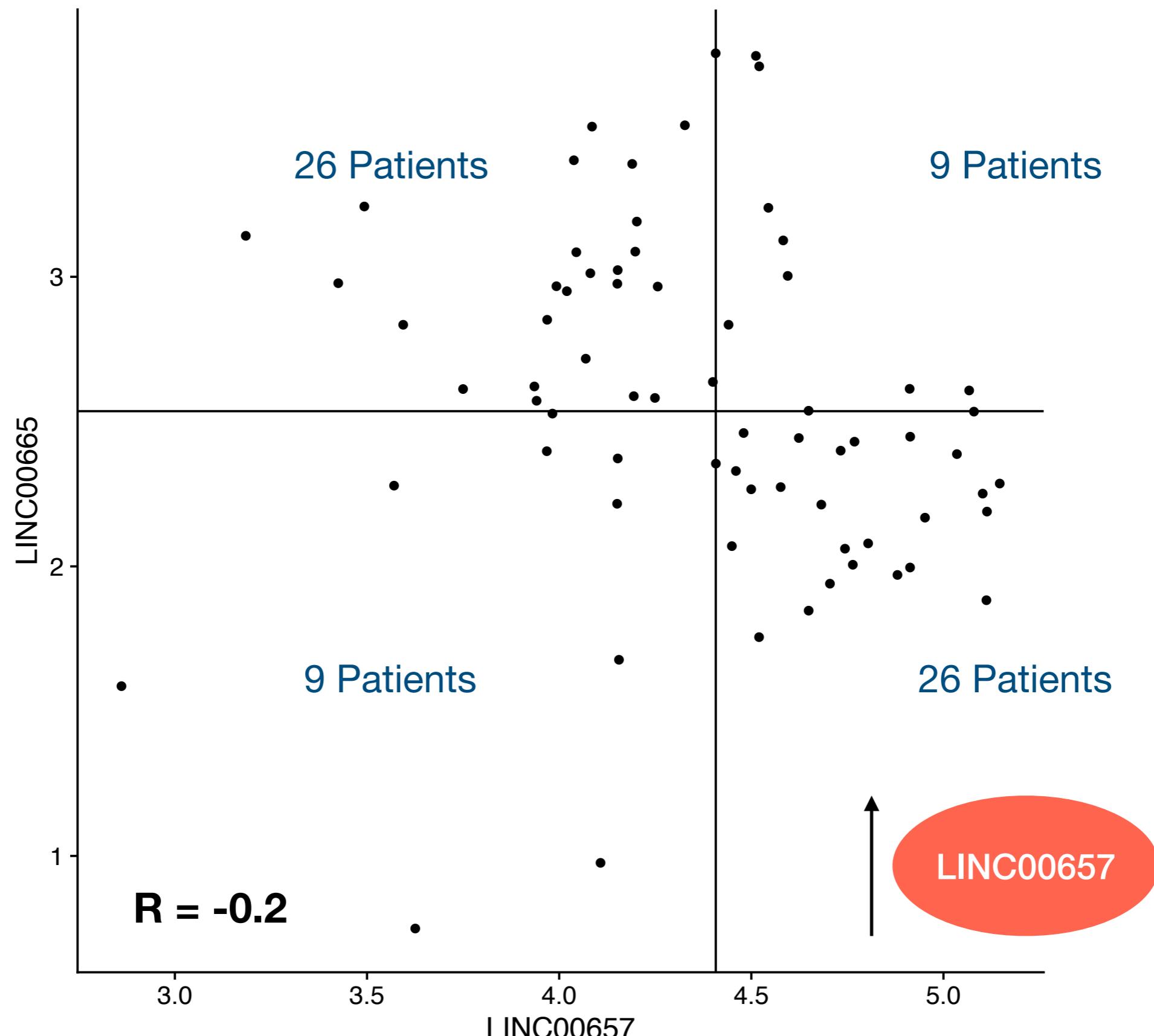
Correlation of LINC00657 and LINC00665 in ovarian cancer



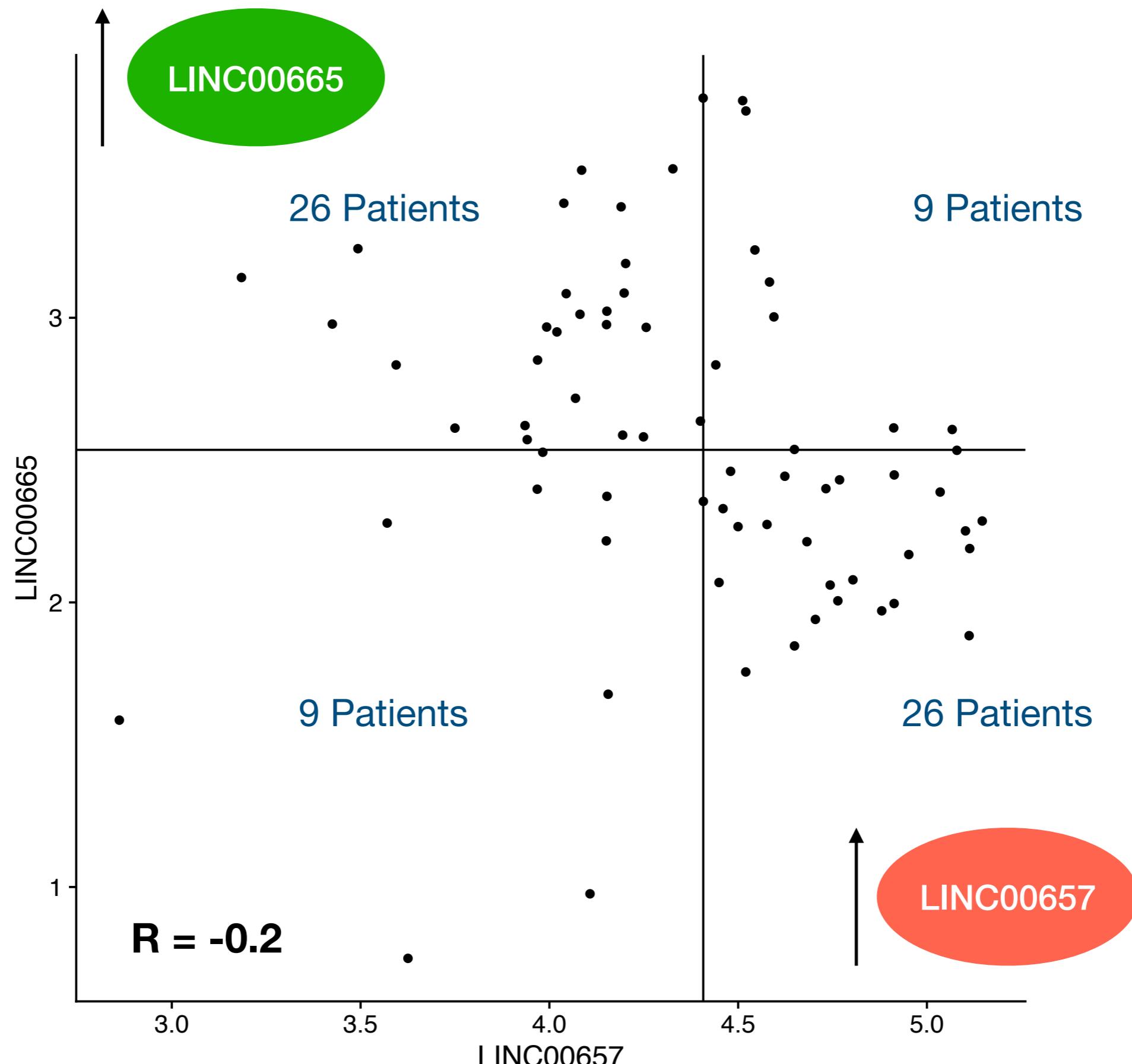
Correlation of LINC00657 and LINC00665 in ovarian cancer



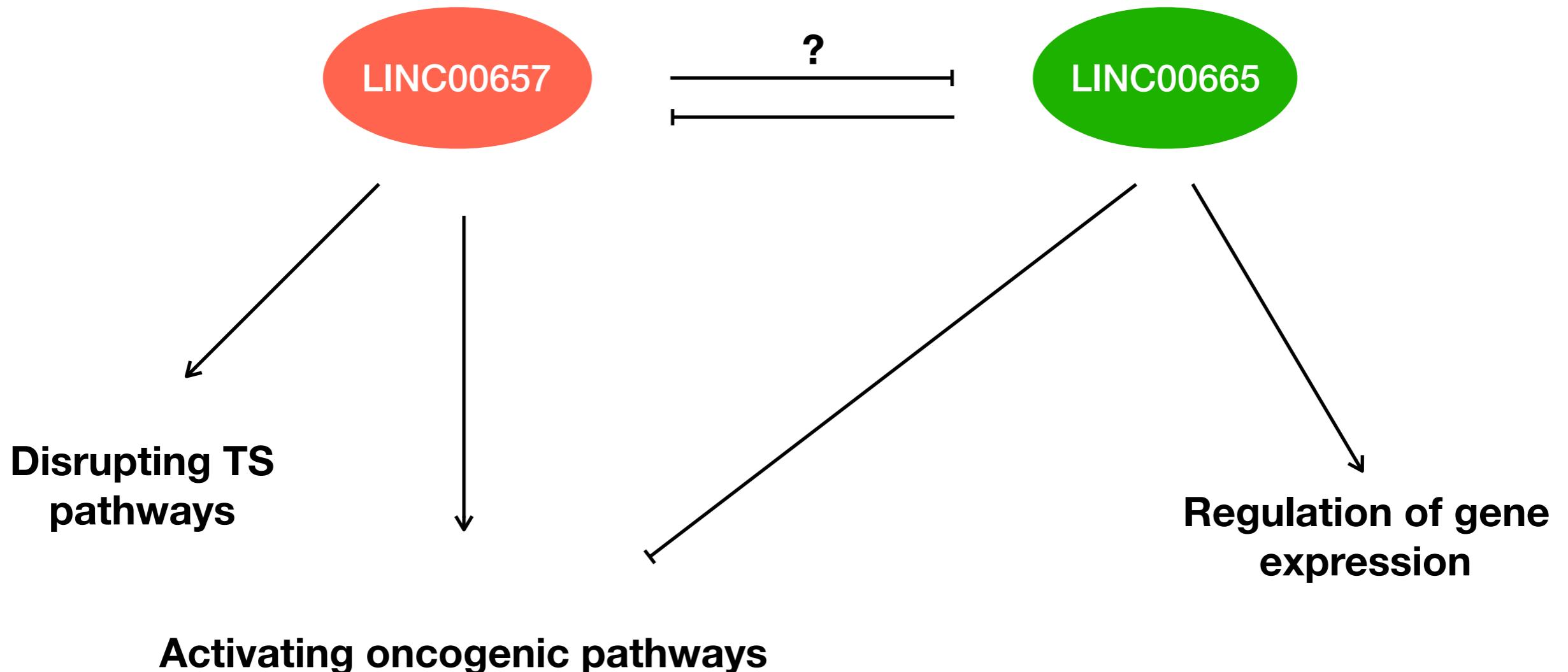
Correlation of LINC00657 and LINC00665 in ovarian cancer



Correlation of LINC00657 and LINC00665 in ovarian cancer



Proposed roles of top lncRNAs' influences on survival

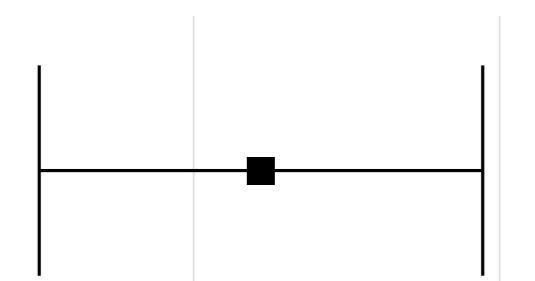


Multivariate Cox Regression Model

```
coxph(Surv(time to death, status) ~ LINC00665 + LINC00657)
```



HR = 0.49
HR, univariate = (0.34)



HR, p-val = 0.021

1 2 5

0.5

1

2

5

Multivariate Cox Regression Model

Model 1: `coxph(Surv(time to death, status) ~ LINC00657)`

`anova(model1, model3),
p-value = 0.0066`

Model 3: `coxph(Surv(time to death, status) ~ LINC00665 + LINC00657)`

Model 2: `coxph(Surv(time to death, status) ~ LINC00665)`

`anova(model2, model3),
p-value = 0.0204`

Model 3: `coxph(Surv(time to death, status) ~ LINC00665 + LINC00657)`

Multivariate Cox Regression Model

Model 1: `coxph(Surv(time to death, status) ~ LINC00657)`

`anova(model1, model3),
p-value = 0.0066`

Model 3: `coxph(Surv(time to death, status) ~ LINC00665 + LINC00657)`

Model 2: `coxph(Surv(time to death, status) ~ LINC00665)`

`anova(model2, model3),
p-value = 0.0204`

Model 3: `coxph(Surv(time to death, status) ~ LINC00665 + LINC00657)`

Multivariate Cox Regression Model

Model 1: `coxph(Surv(time to death, status) ~ LINC00657)`

`anova(model1, model3),
p-value = 0.0066`

Model 3: `coxph(Surv(time to death, status) ~ LINC00665 + LINC00657)`

Model 2: `coxph(Surv(time to death, status) ~ LINC00665)`

`anova(model2, model3),
p-value = 0.0204`

Model 3: `coxph(Surv(time to death, status) ~ LINC00665 + LINC00657)`

Multivariate Cox Regression Model

Model 1: `coxph(Surv(time to death, status) ~ LINC00657)`

`anova(model1, model3),
p-value = 0.0066`

Model 3: `coxph(Surv(time to death, status) ~ LINC00665 + LINC00657)`

Model 2: `coxph(Surv(time to death, status) ~ LINC00665)`

`anova(model2, model3),
p-value = 0.0204`

Model 3: `coxph(Surv(time to death, status) ~ LINC00665 + LINC00657)`

Multivariate Cox Regression Model

Model 1: `coxph(Surv(time to death, status) ~ LINC00657)`

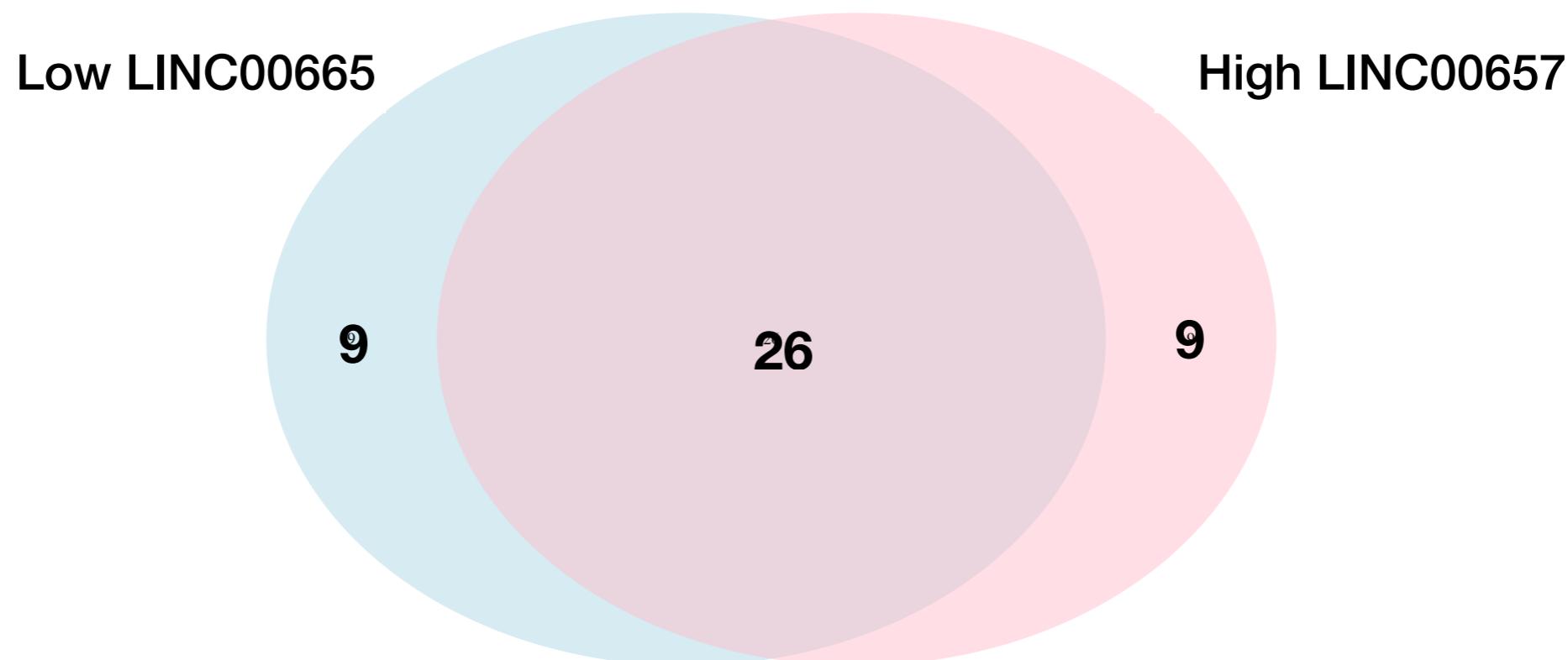
`anova(model1, model3),
p-value = 0.0066`

Model 3: `coxph(Surv(time to death, status) ~ LINC00665 + LINC00657)`

Model 2: `coxph(Surv(time to death, status) ~ LINC00665)`

`anova(model2, model3),
p-value = 0.0204`

Model 3: `coxph(Surv(time to death, status) ~ LINC00665 + LINC00657)`

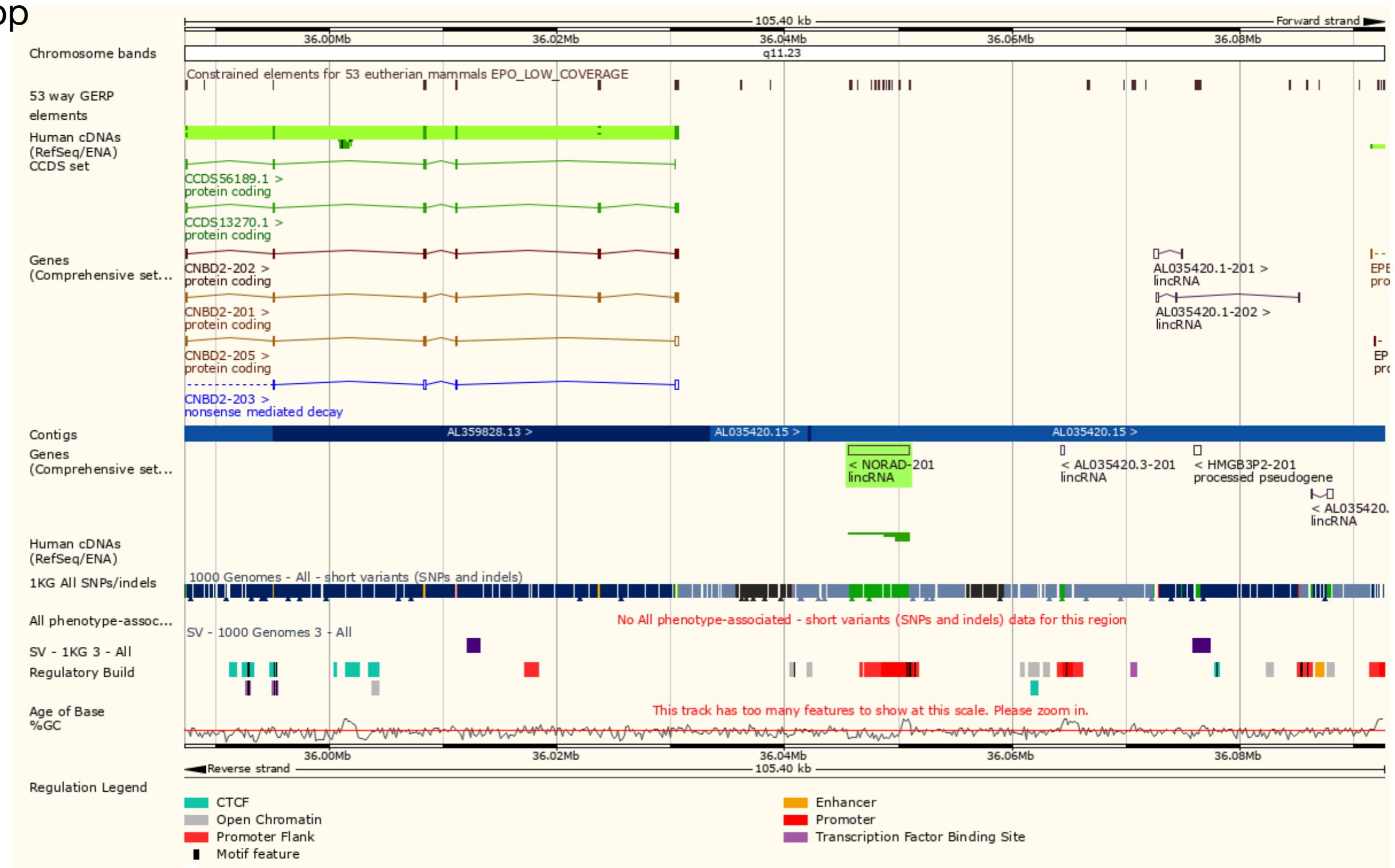


Summary of progress

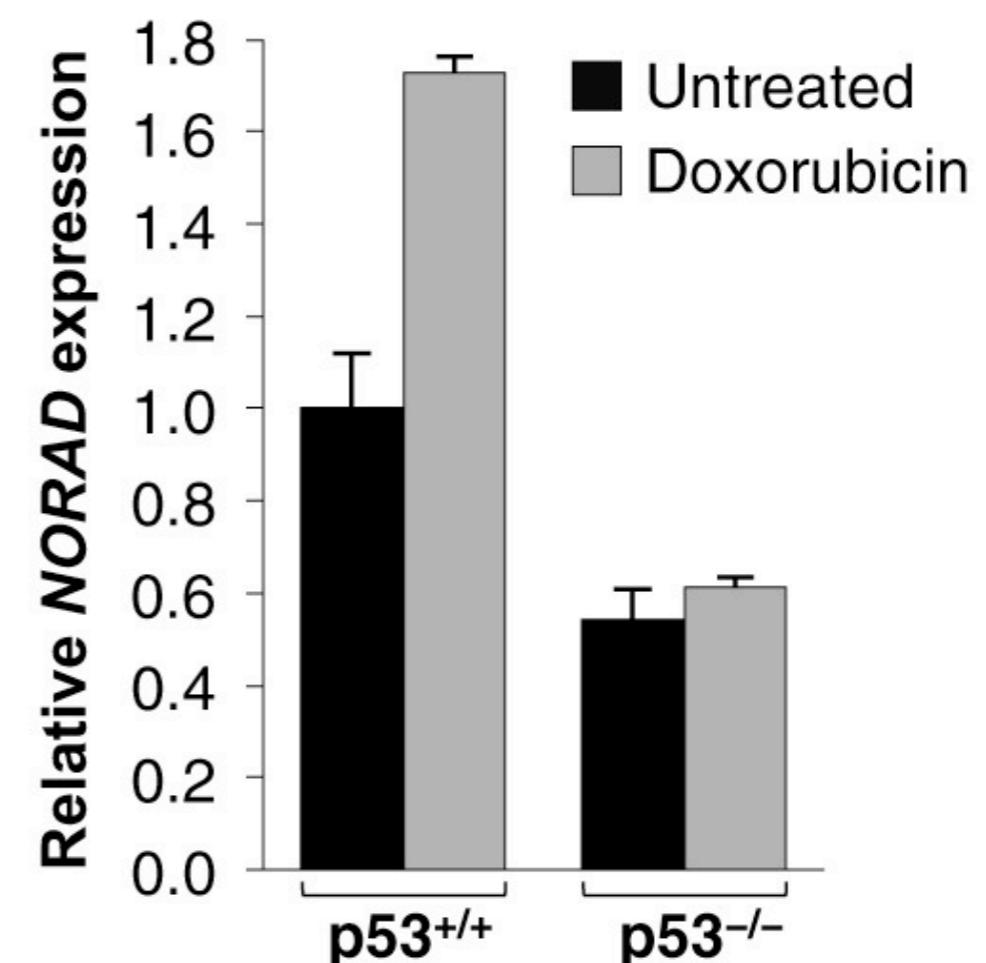
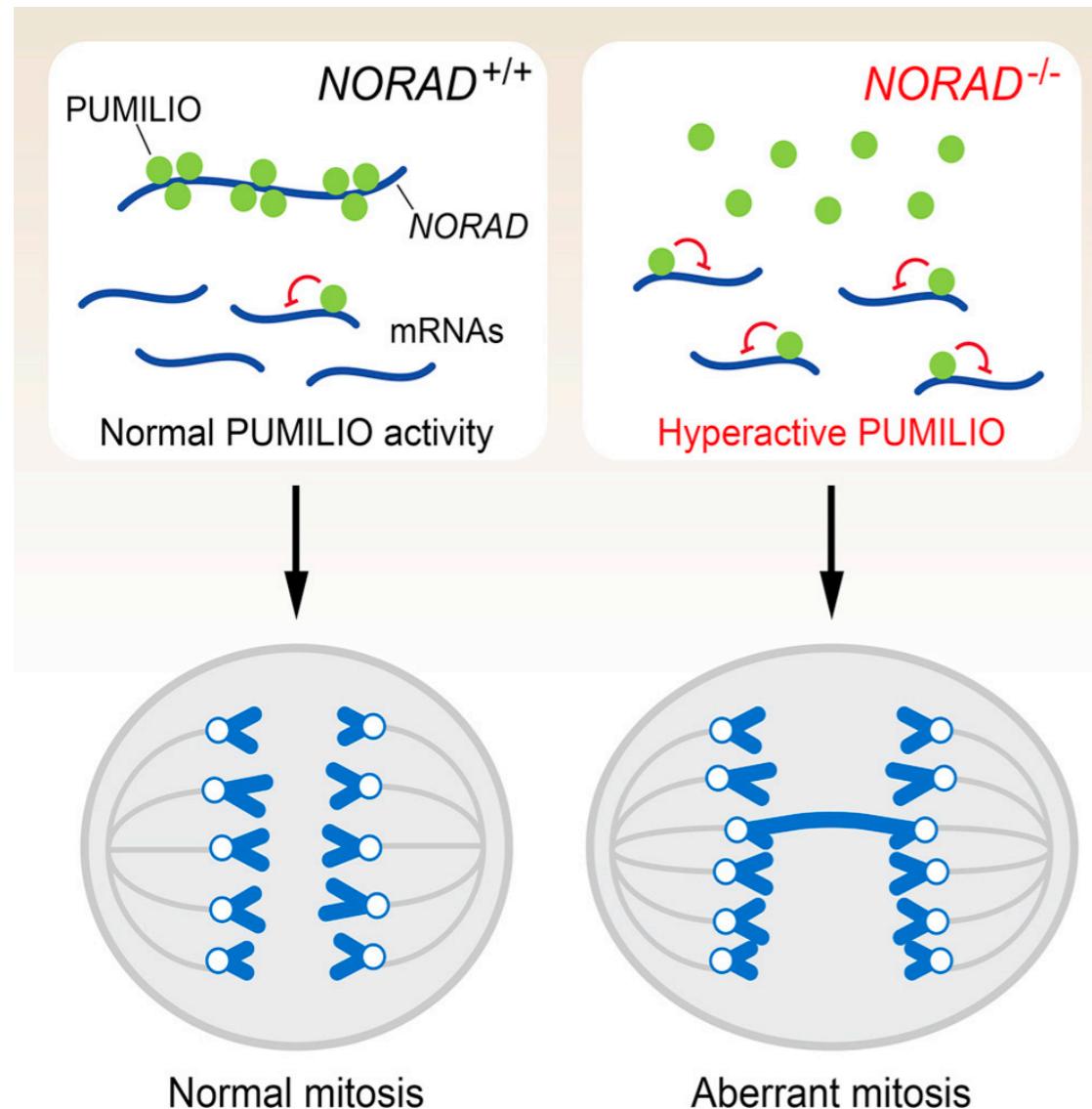
1. Identified high expressing lncRNA candidates in multiple cancer types
2. Survival analysis of lncRNAs using the Cox Proportional Hazards model
3. Differential expression analysis of candidate lncRNAs between tumours (PCAWG) and normal tissues (GTEx)
4. **In depth analysis of two lncRNA candidates in Ovarian cancer**
 - **Predicting mechanism through differential expression and pathway enrichment analysis**

LINC00657 (NORAD) genomic properties

Chr 20
5,339 bp



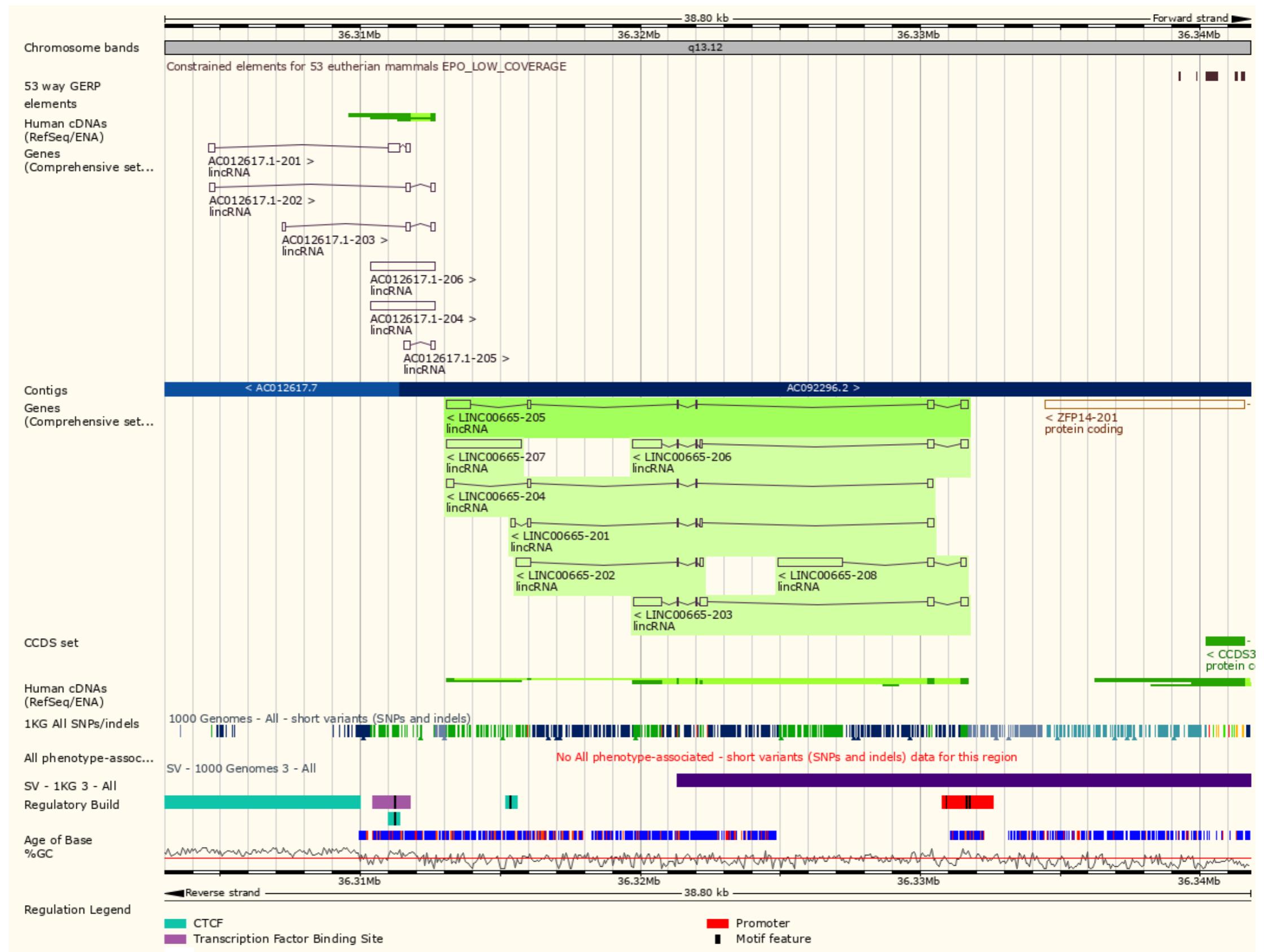
LINC00657 (NORAD) functional annotation



- DNA damage induced NORAD expression in a TP53-dependent manner in HCT116 human colon cancer cells
- Conditional knockout or knockdown of NORAD expression caused multiple mitotic errors, including anaphase bridges, mitotic slippage, and significant aneuploidy

LINC00665 genomic properties

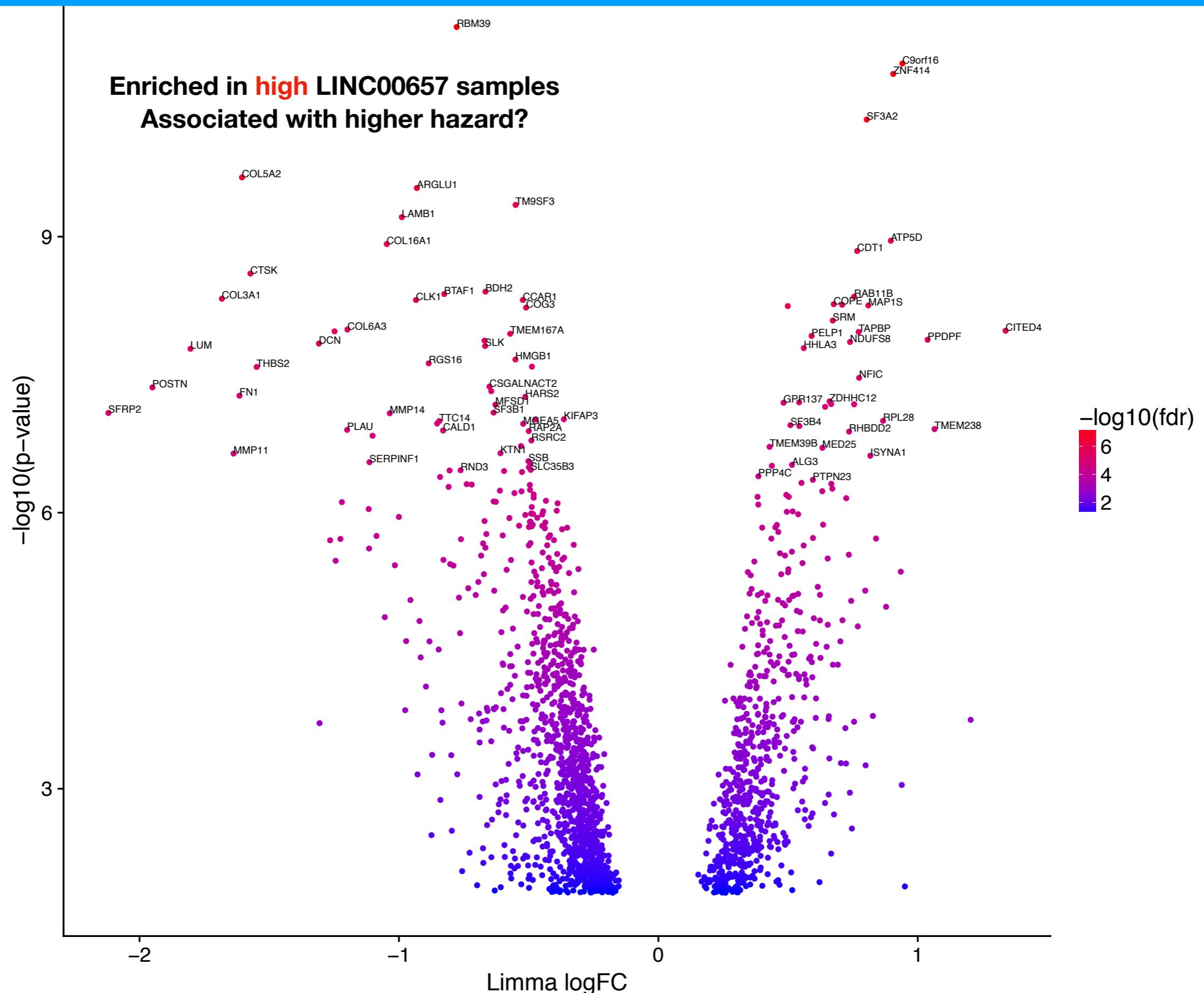
Chr 19
2,732 bp



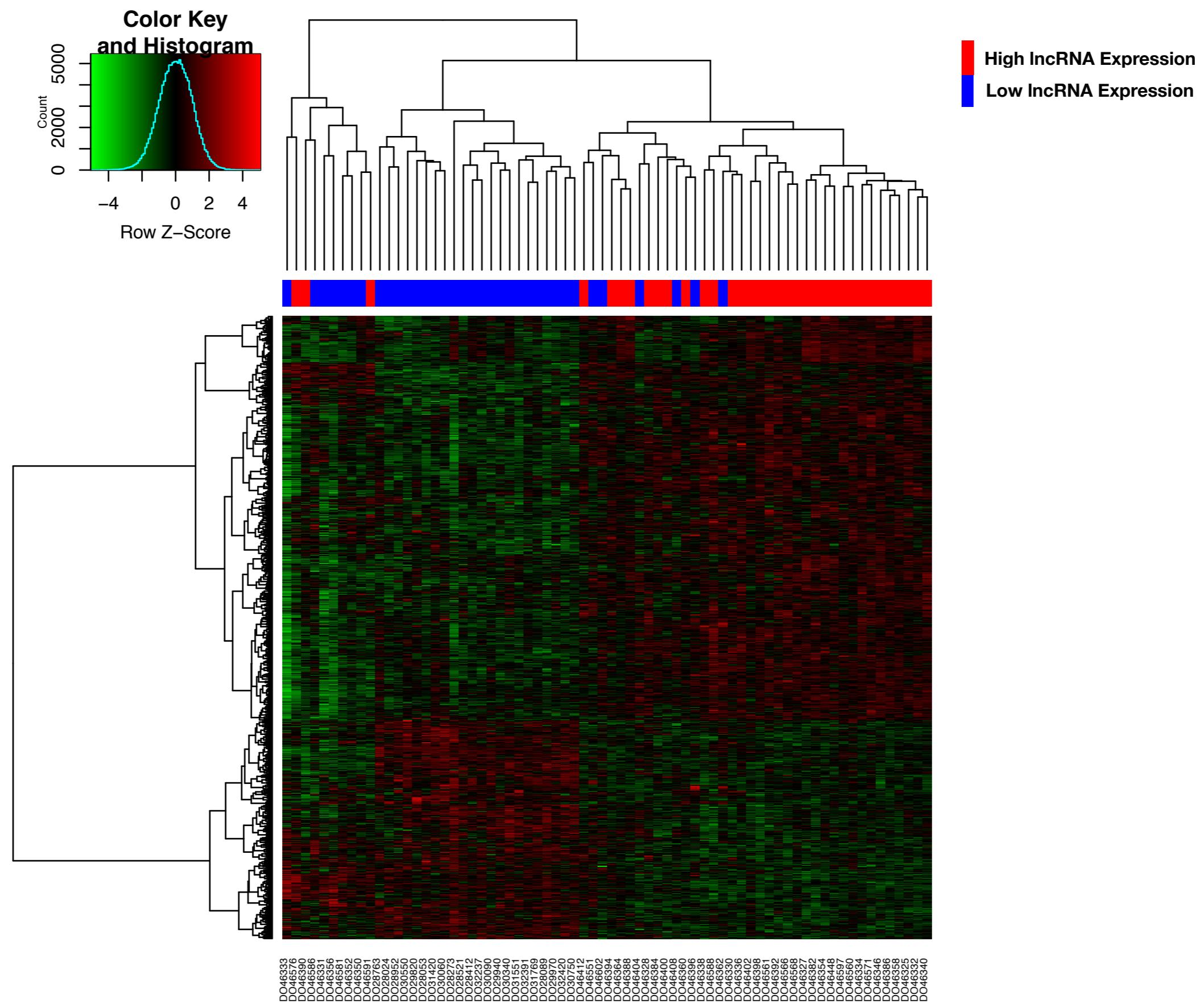
Predicting regulatory mechanisms of LINC00657 and LINC00665

- Differential expression analysis of PCGs between high and low lncRNA expressing patients
 - LINC00657: 1,837 differentially expressed (DE) PCGs
 - LINC00665: 641 differentially expressed (DE) PCGs
 - All PCGs with median expression > 5FPKM
 - All DE genes have an FDR adjusted p-val < 0.05
- **Hypothesis:** Differentially expressed genes that are enriched in **high** LINC00657 and **low** LINC00665 patients are associated with pathways that lead to tumour progression
- Identify pathways differentially expressed between the two groups

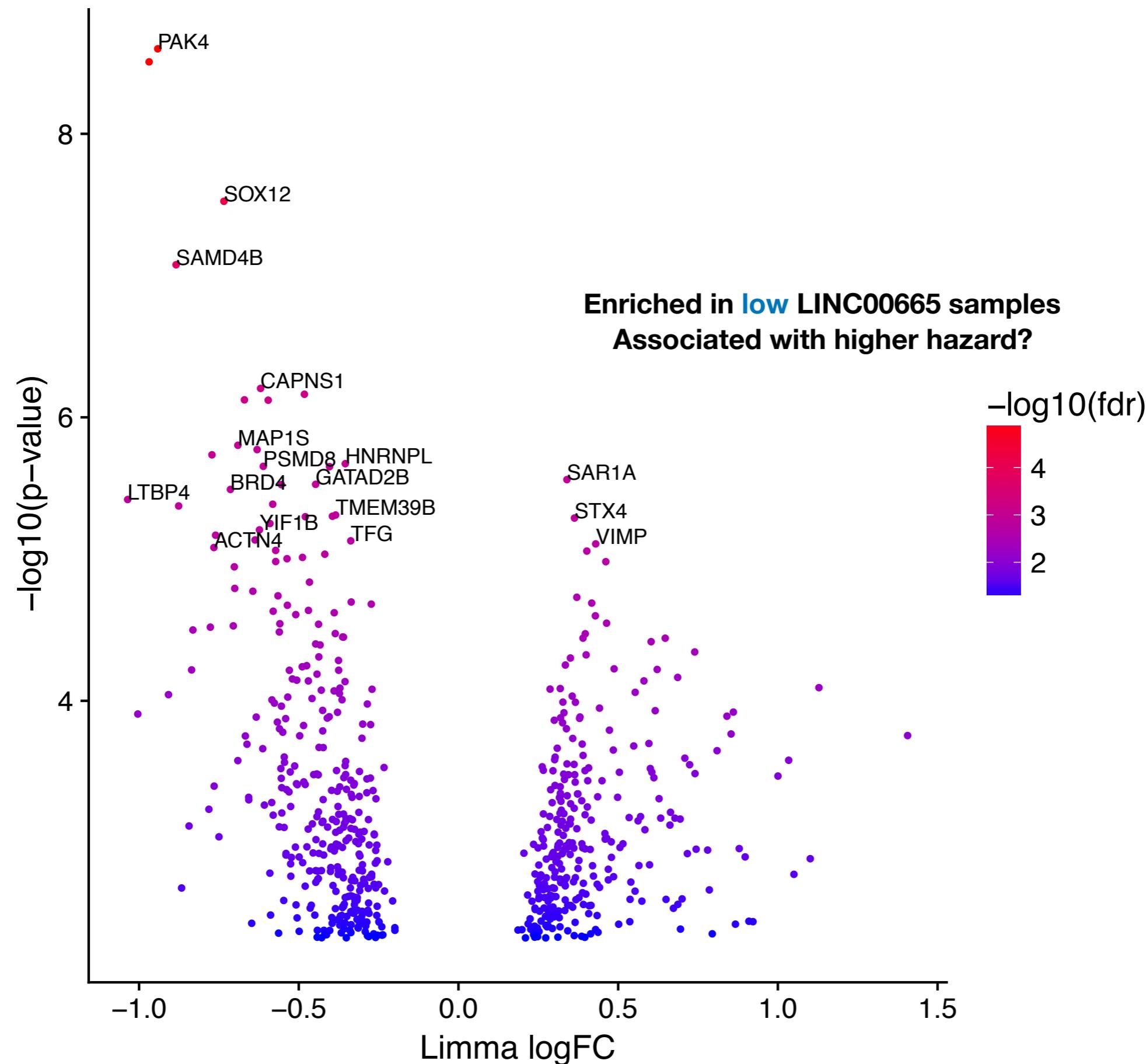
1,837 PCGs differentially expressed between high and low LINC00657 samples



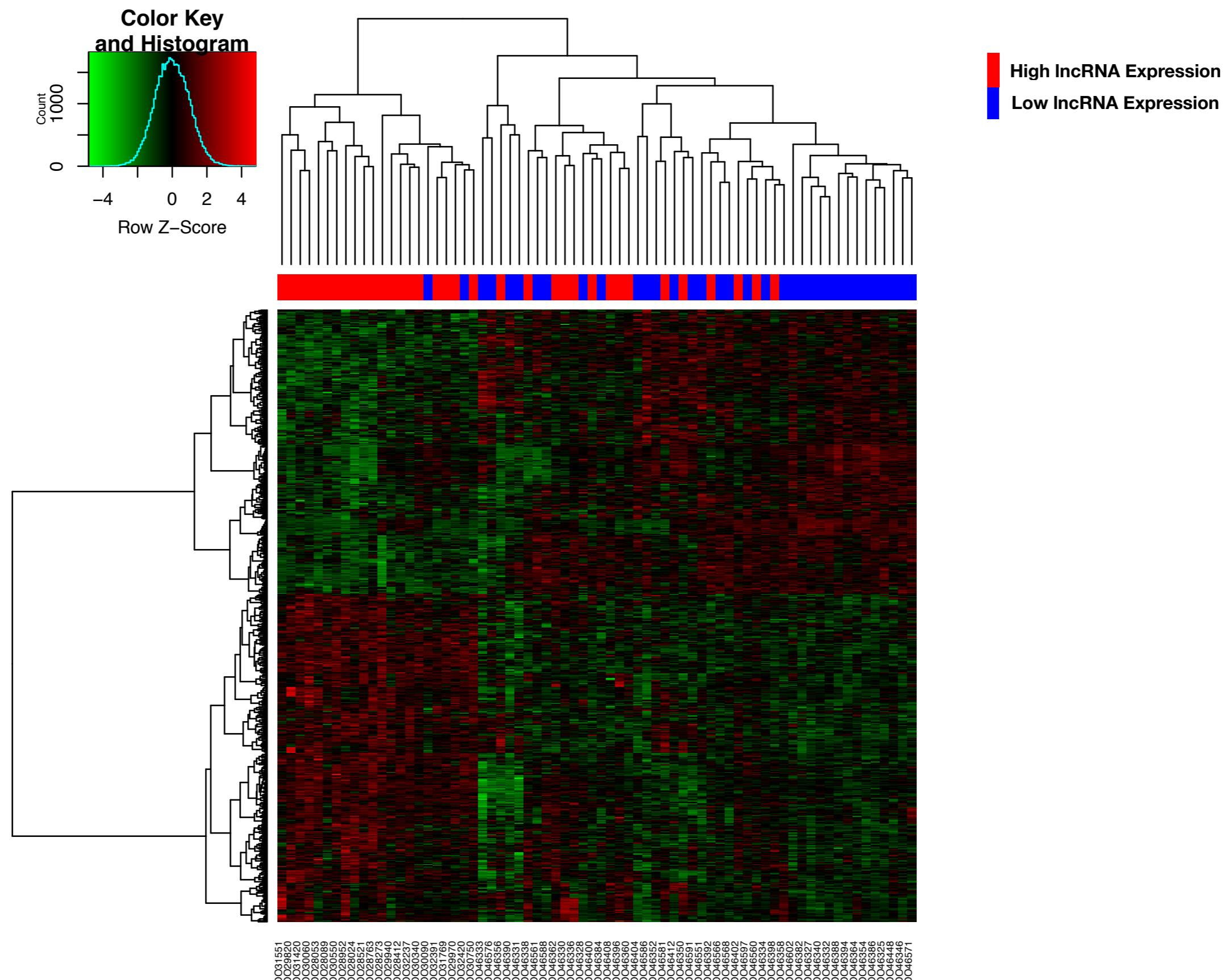
1,837 PCGs differentially expressed between high and low LINC00657 samples



641 PCGs differentially expressed between high and low LINC00665 samples



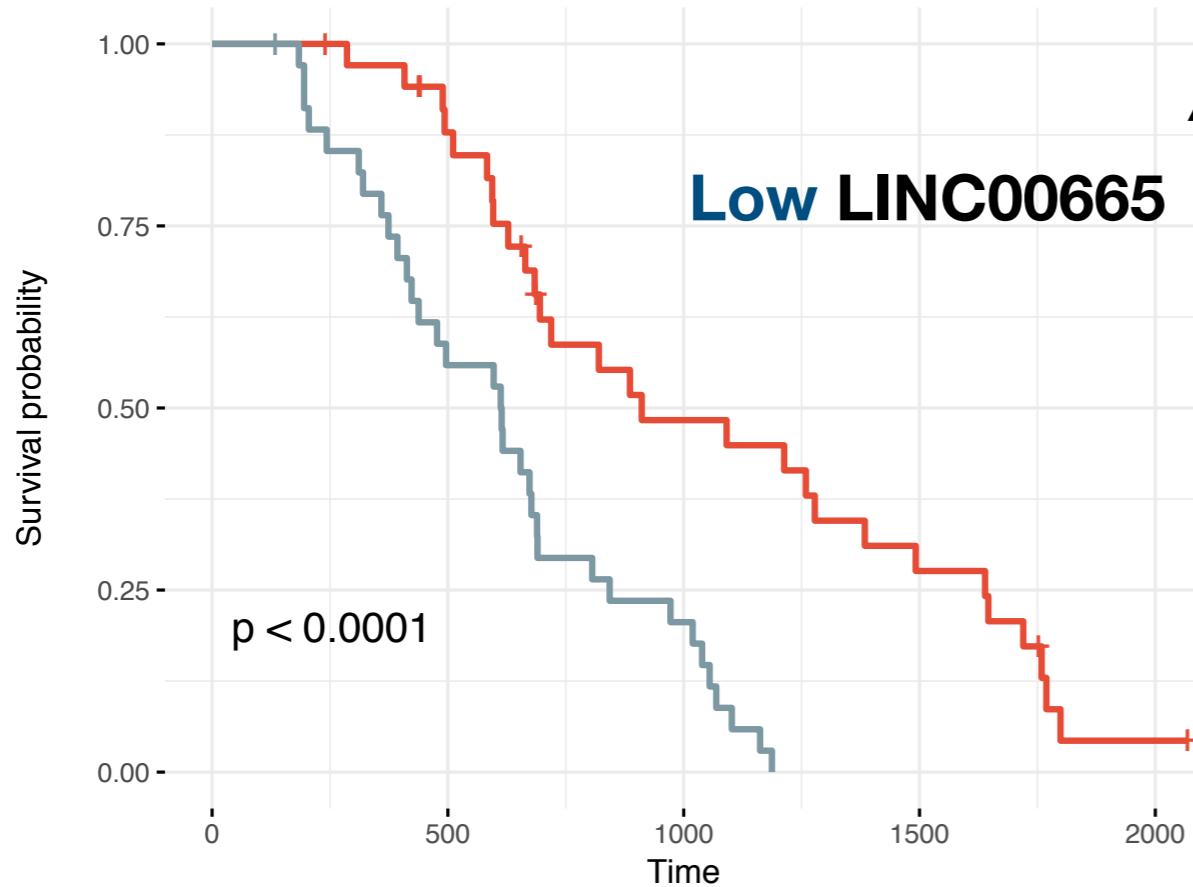
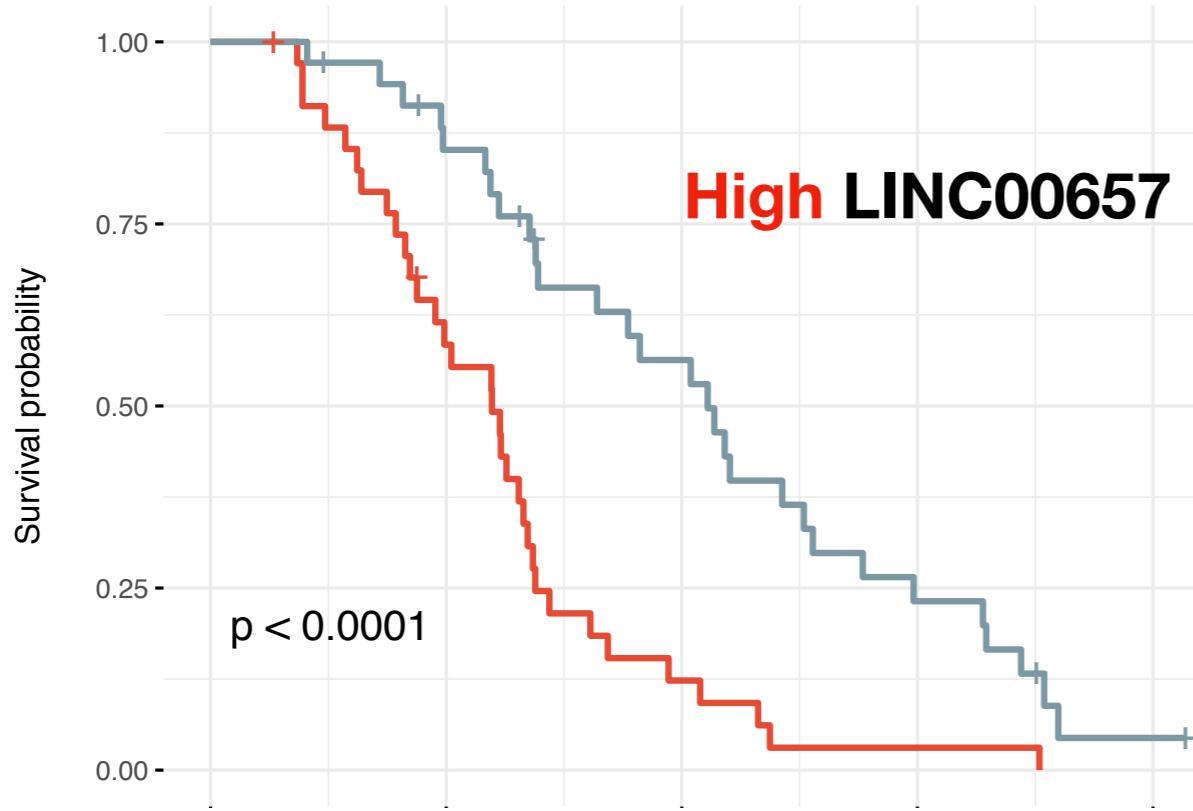
641 PCGs differentially expressed between high and low LINC00665 samples



Predicting regulatory mechanisms of LINC00657 and LINC00665

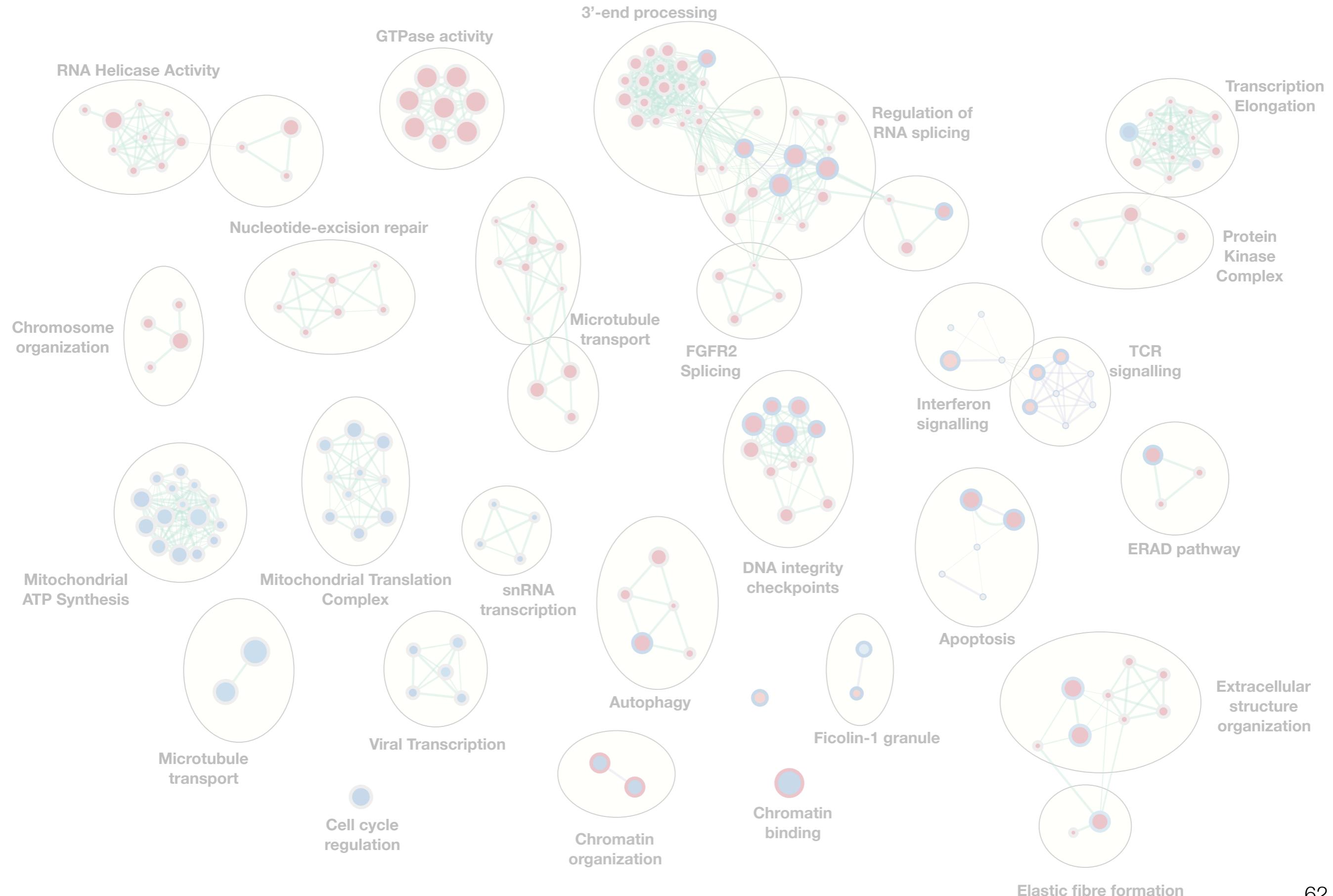
- Differential expression analysis of PCGs between high and low lncRNA expressing patients
 - LINC00657: 1,837 differentially expressed (DE) PCGs
 - LINC00665: 641 differentially expressed (DE) PCGs
 - All PCGs with median expression > 5FPKM
 - All DE genes have an FDR adjusted p-val < 0.05
- **Hypothesis:** Differentially expressed genes that are enriched in **high** LINC00657 and **low** LINC00665 patients are associated with pathways that lead to tumour progression
- **Identify pathways differentially expressed between the two groups**
 - **Which genes distinguish high and low groups ? What pathways are they enriched in?**

Pathway enrichment analysis of DE genes: hypothesis

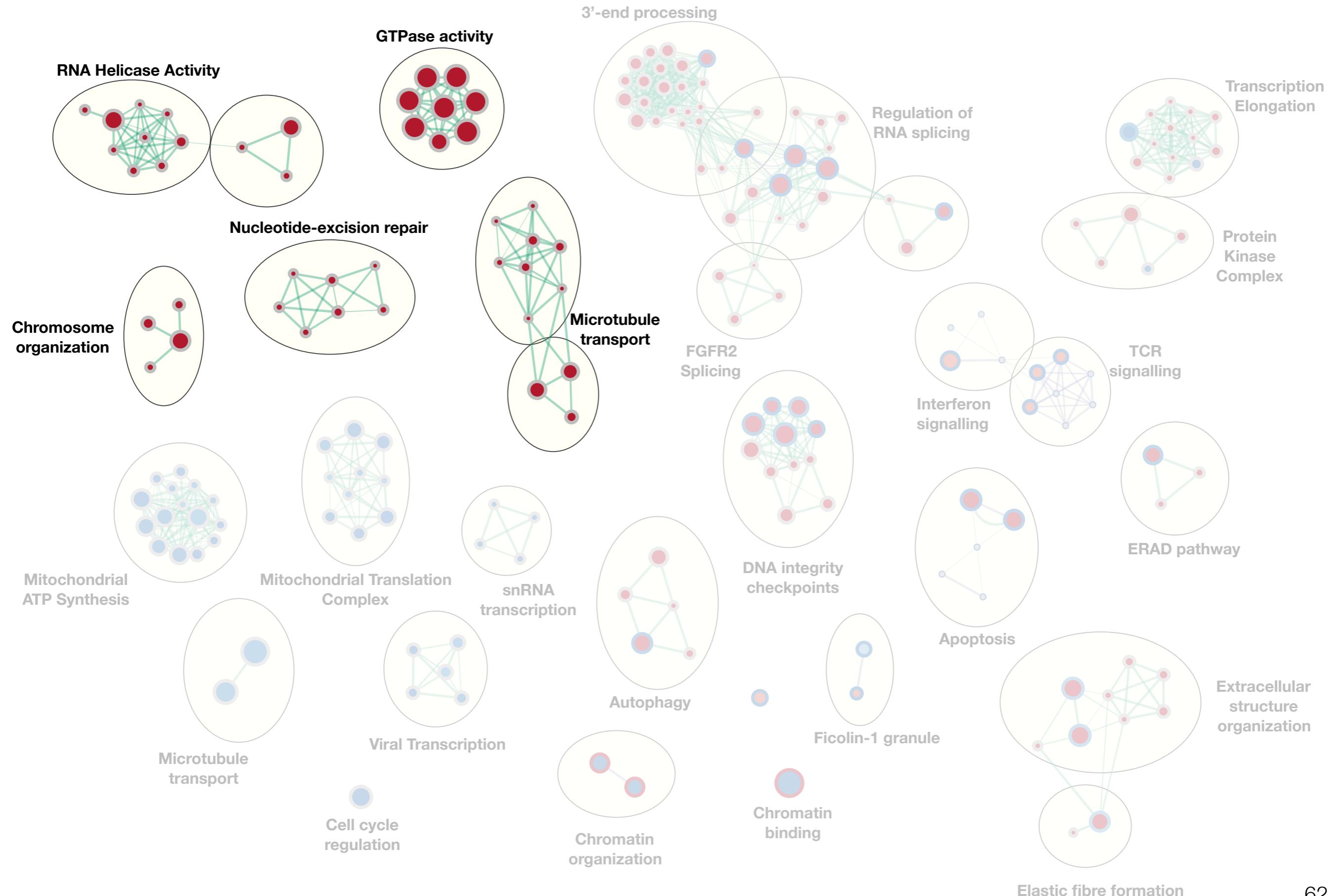


Pathways enriched by genes differentially expressed in these groups are associated with tumour development

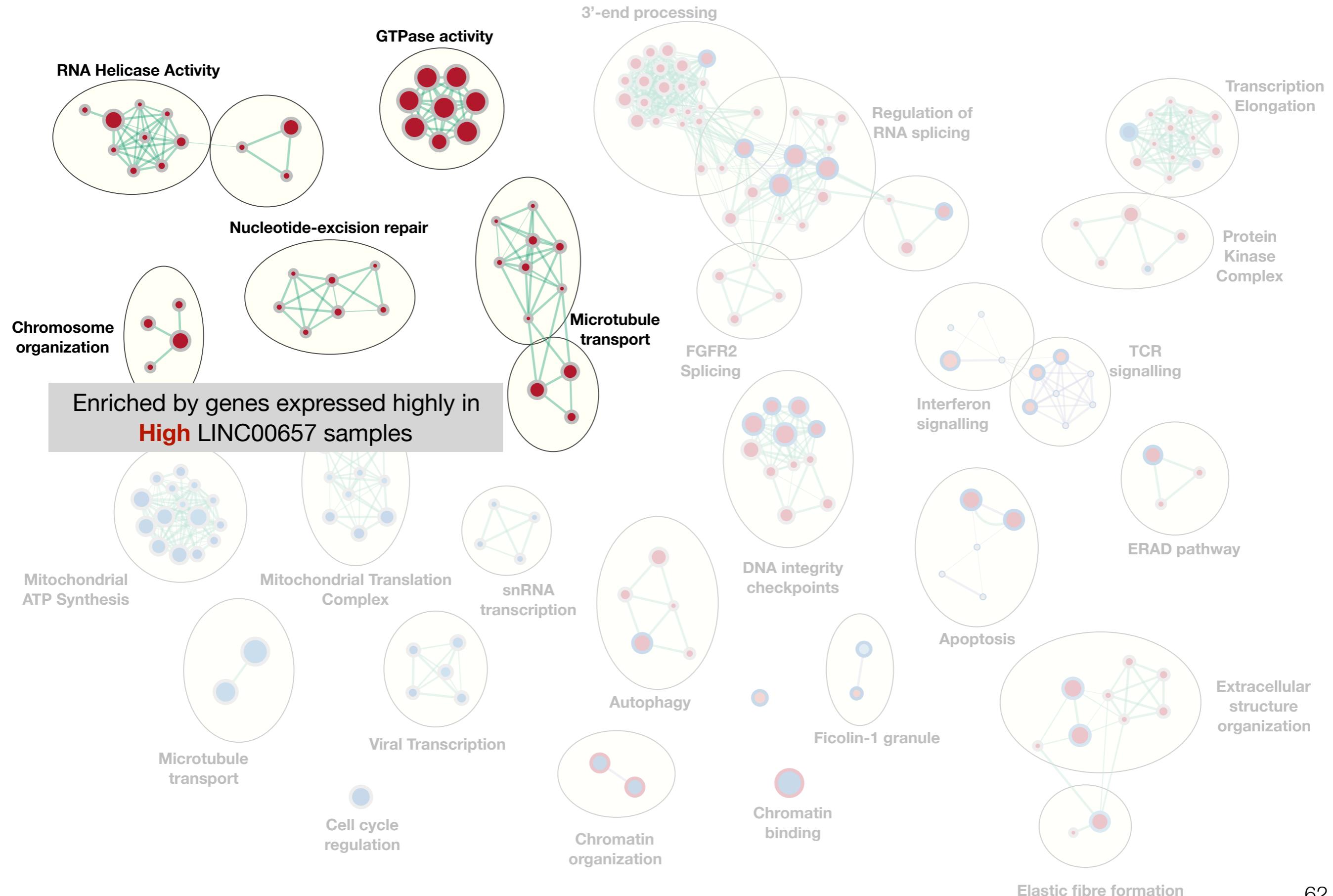
LINC00657 and LINC00665 associate with several common pathways



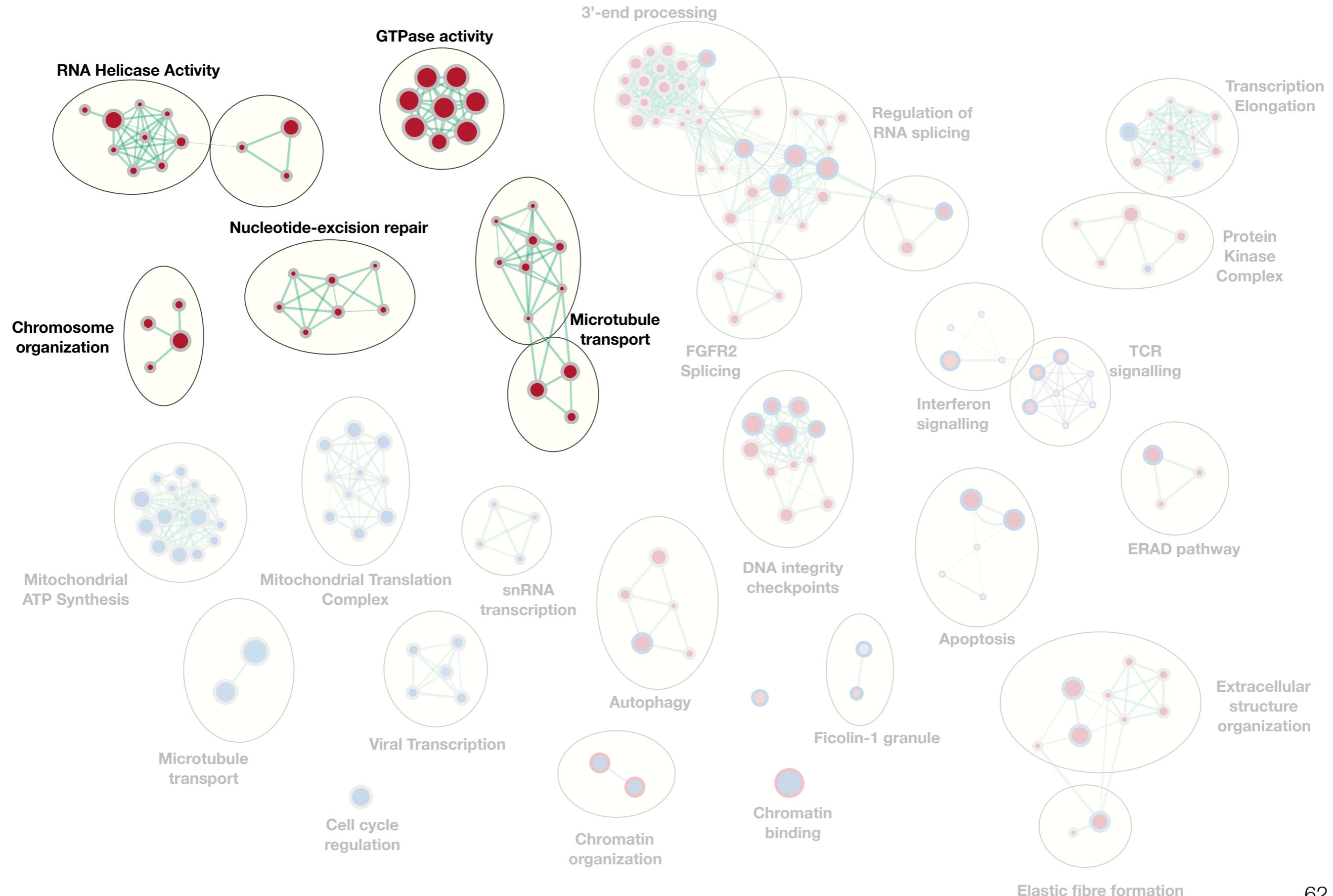
LINC00657 and LINC00665 associate with several common pathways



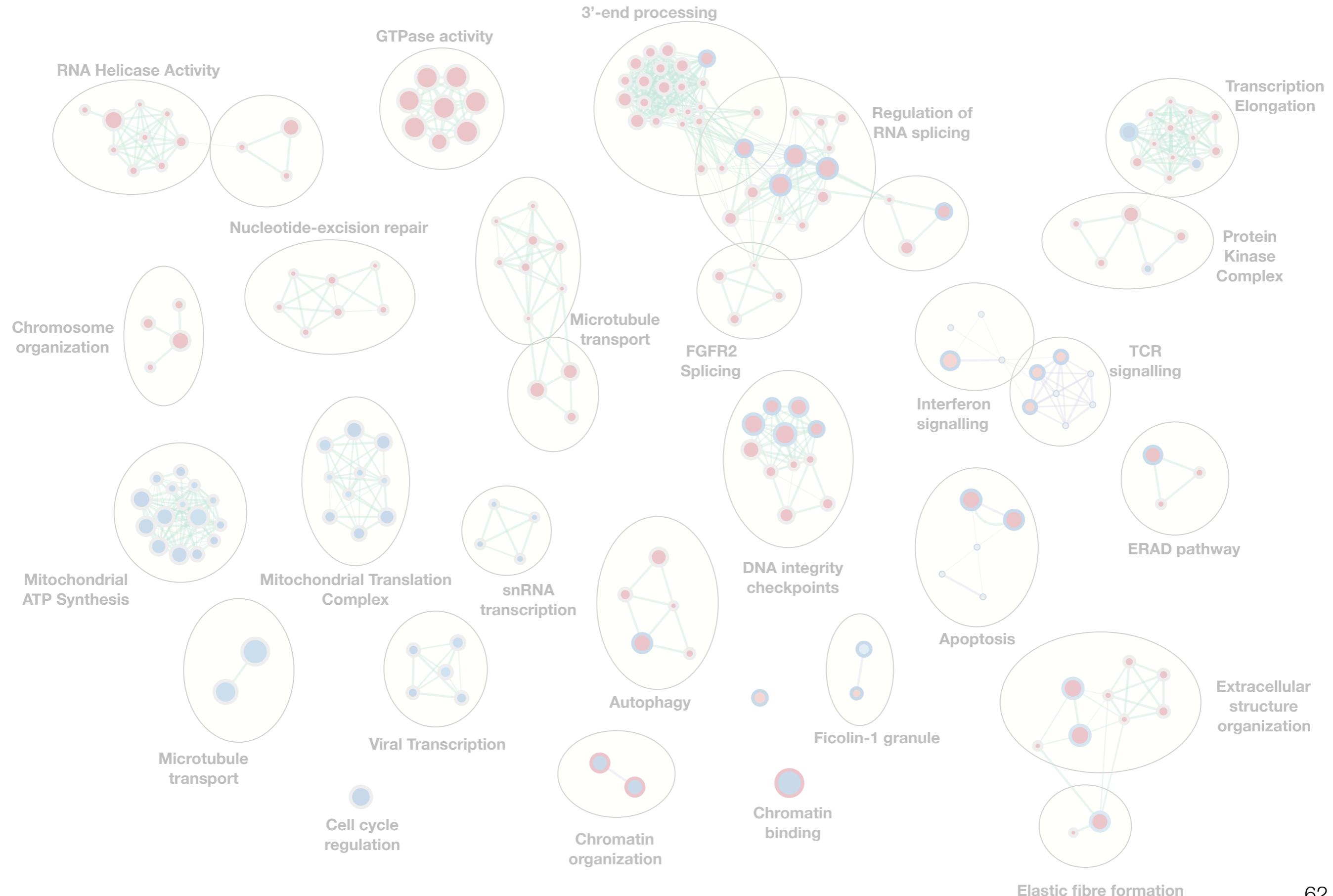
LINC00657 and LINC00665 associate with several common pathways



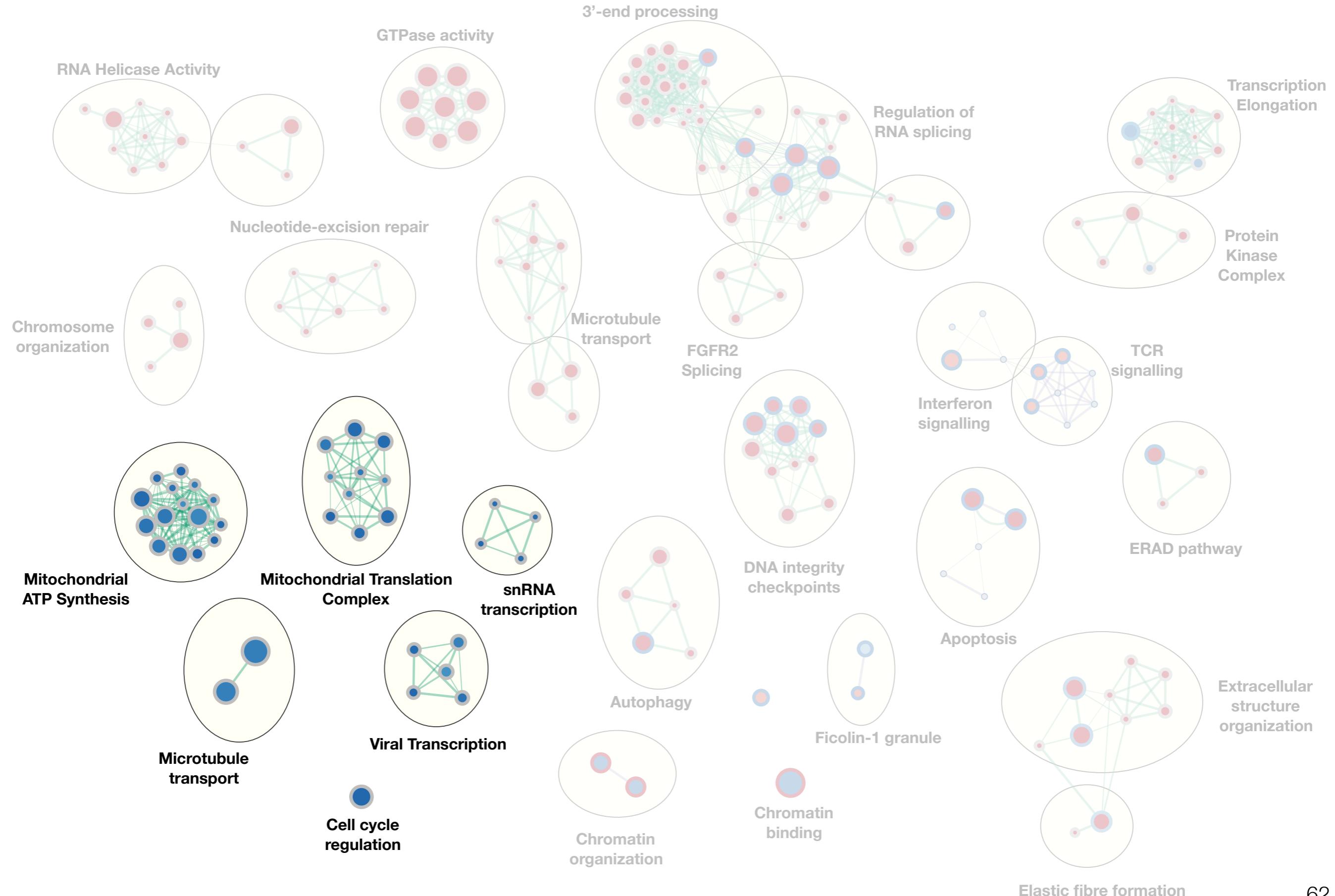
LINC00657 and LINC00665 associate with several common pathways



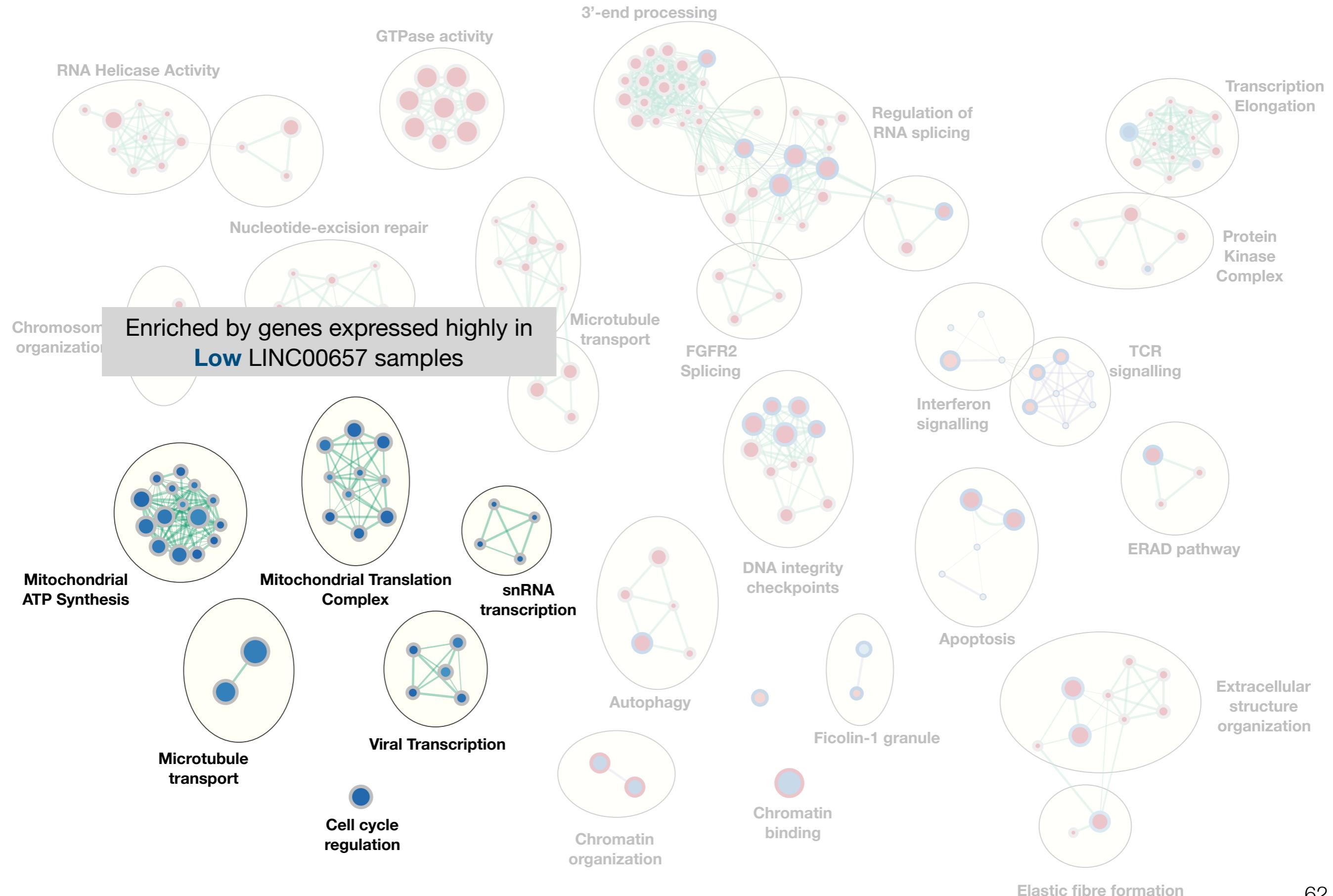
LINC00657 and LINC00665 associate with several common pathways



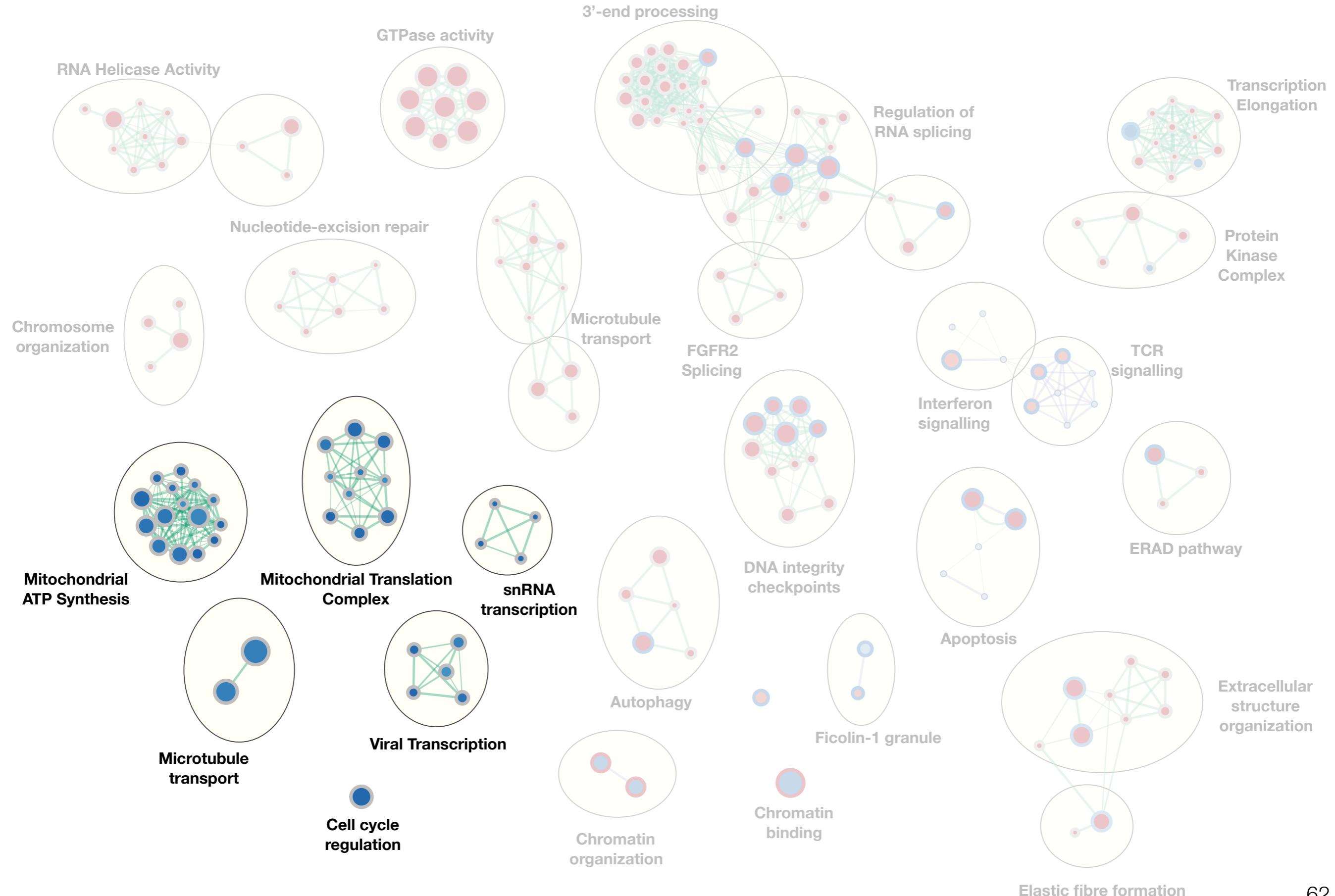
LINC00657 and LINC00665 associate with several common pathways



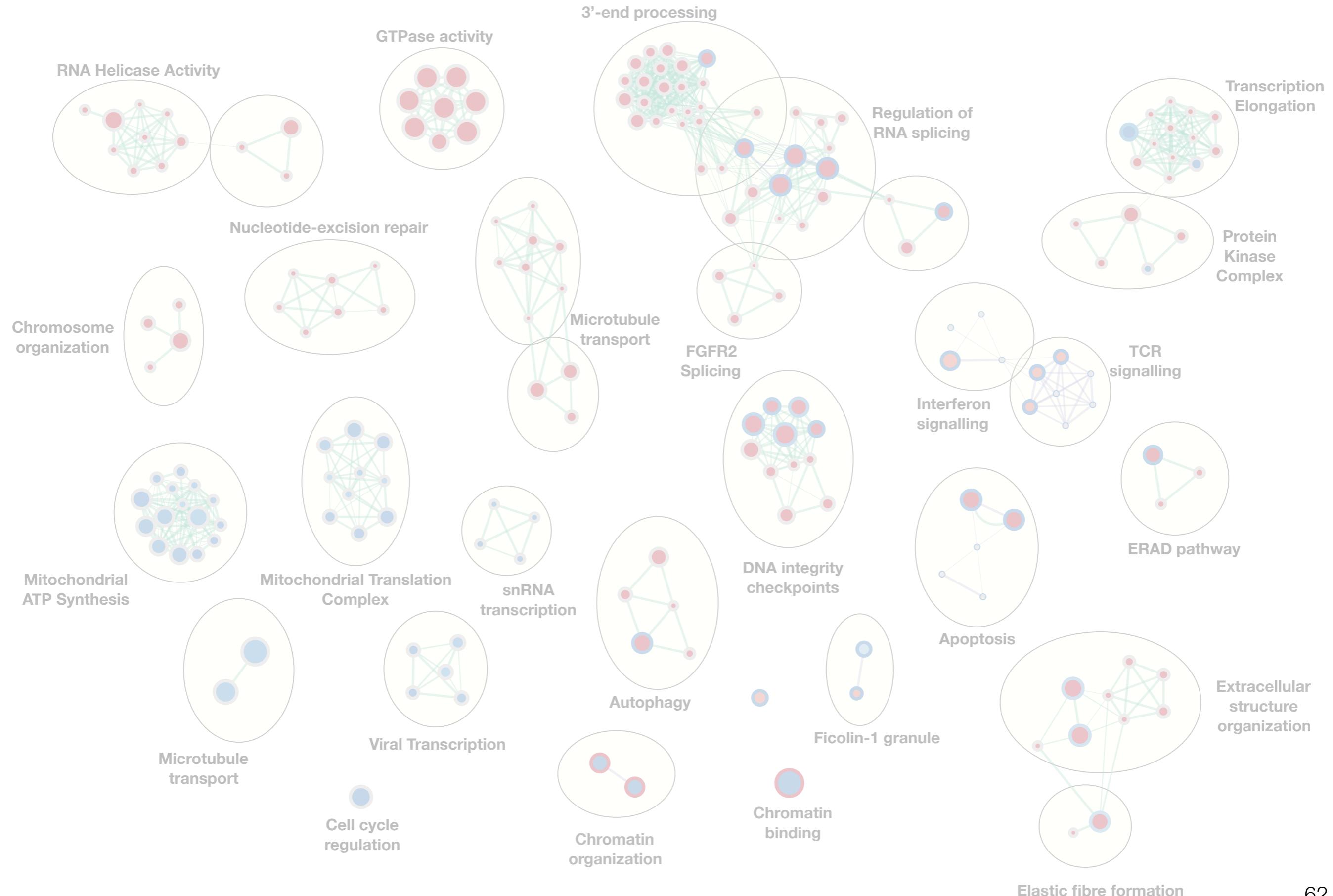
LINC00657 and LINC00665 associate with several common pathways



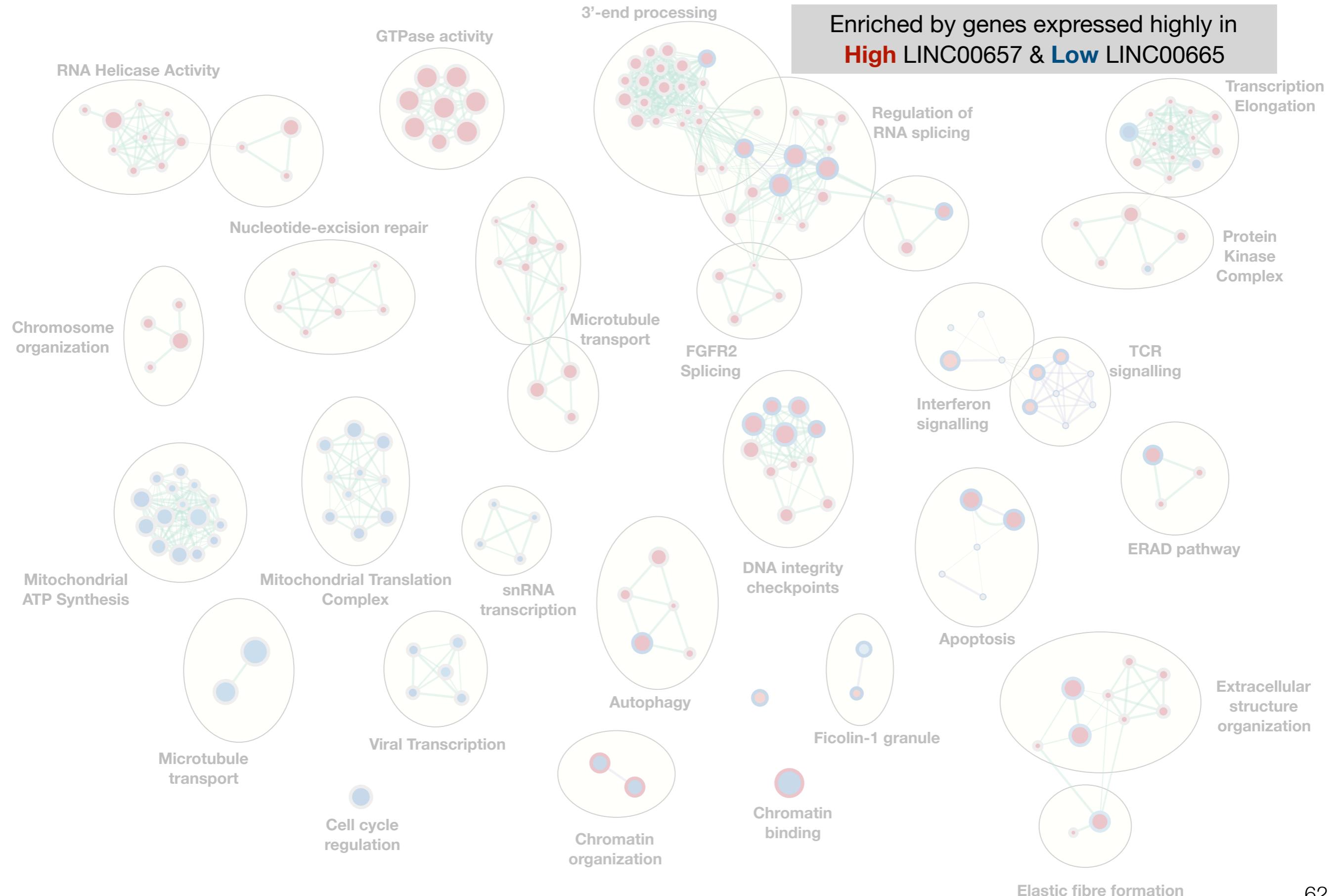
LINC00657 and LINC00665 associate with several common pathways



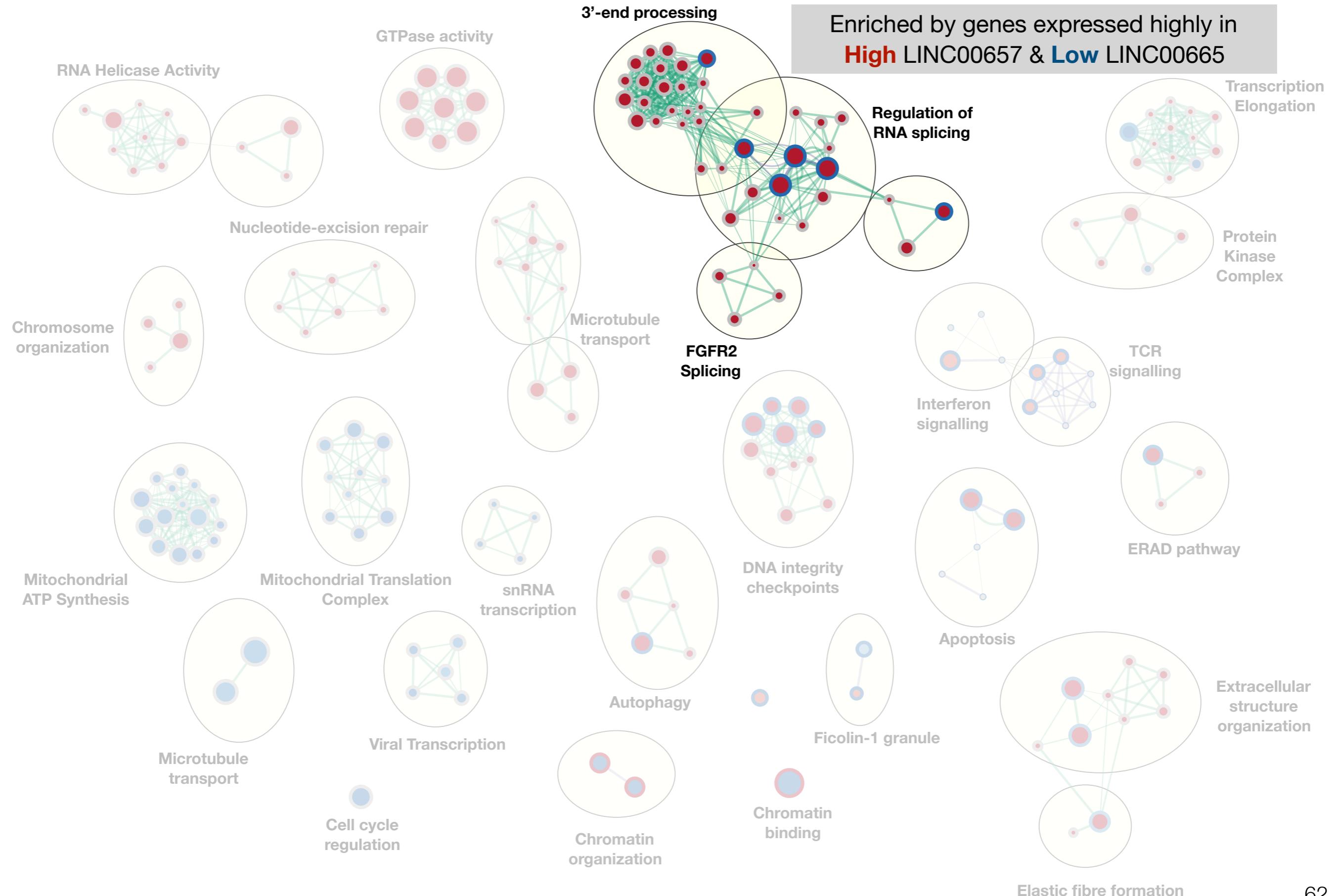
LINC00657 and LINC00665 associate with several common pathways



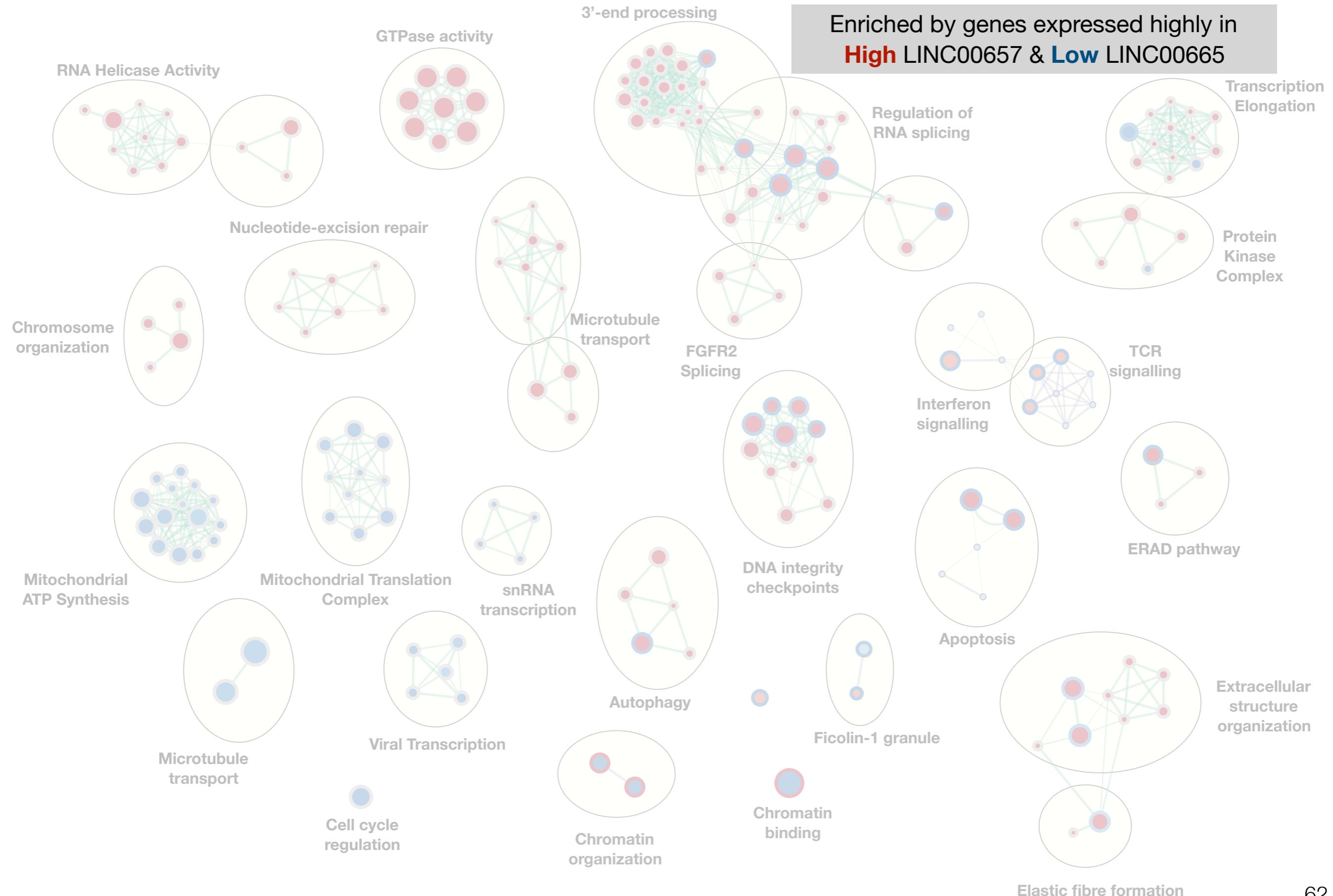
LINC00657 and LINC00665 associate with several common pathways



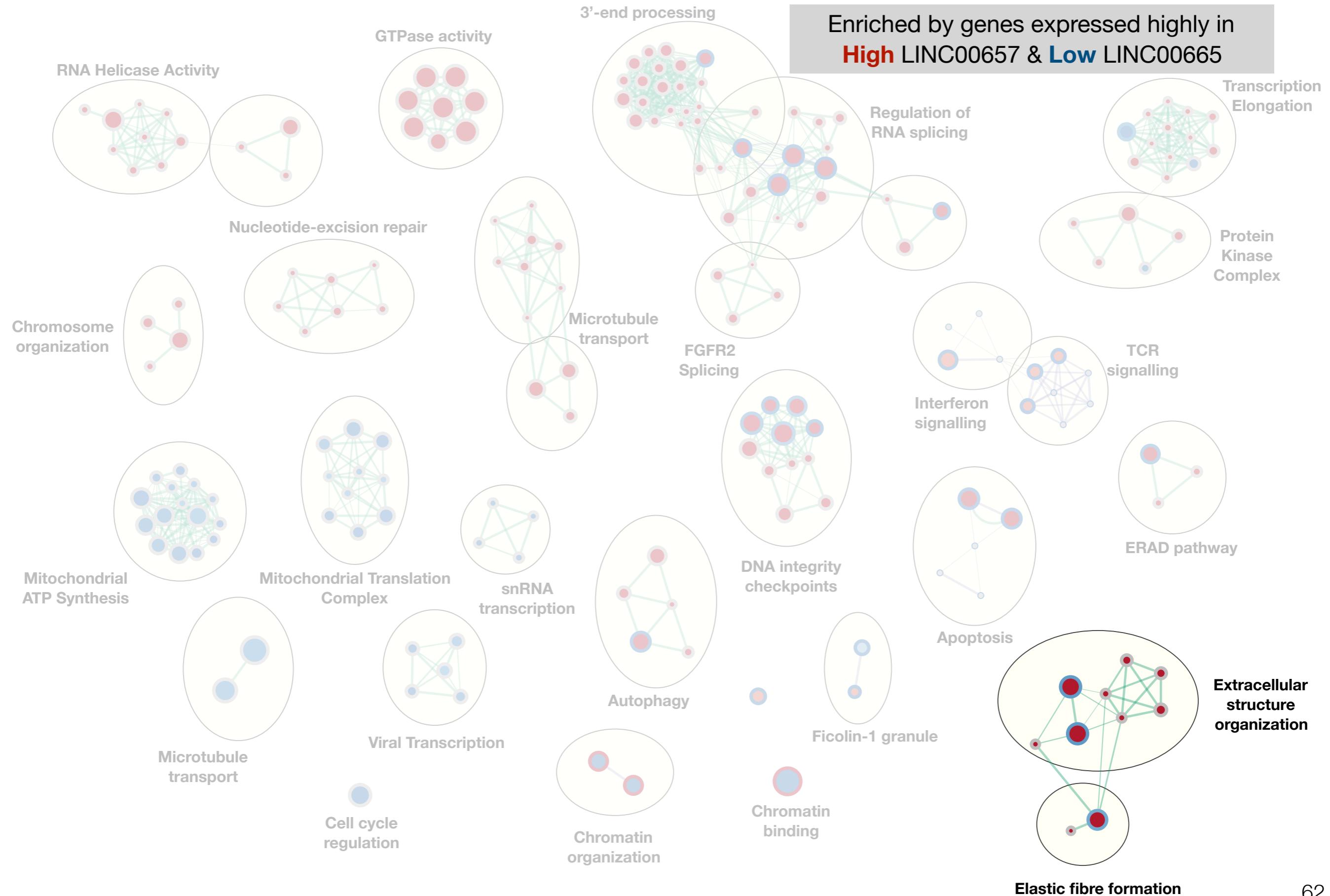
LINC00657 and LINC00665 associate with several common pathways



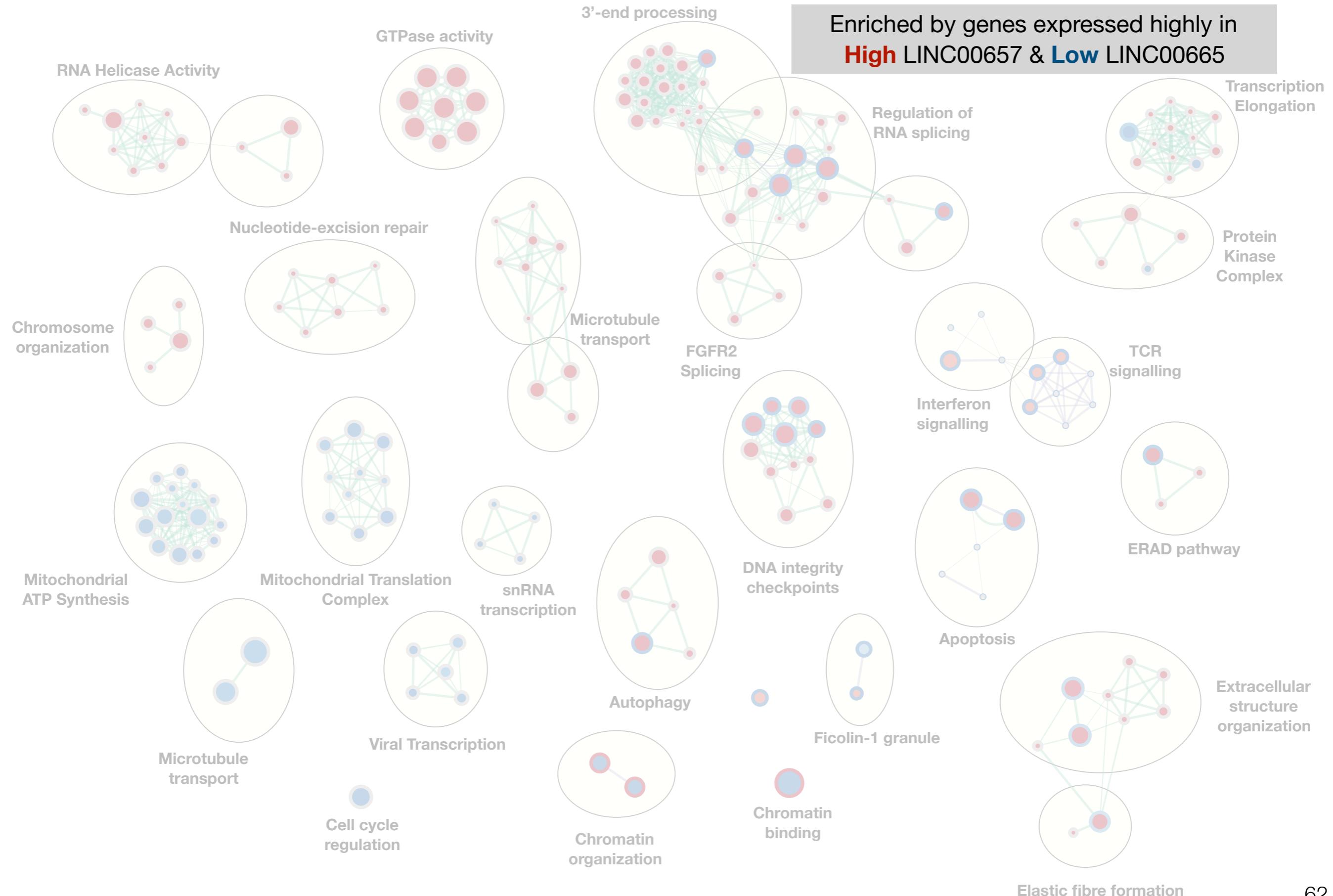
LINC00657 and LINC00665 associate with several common pathways



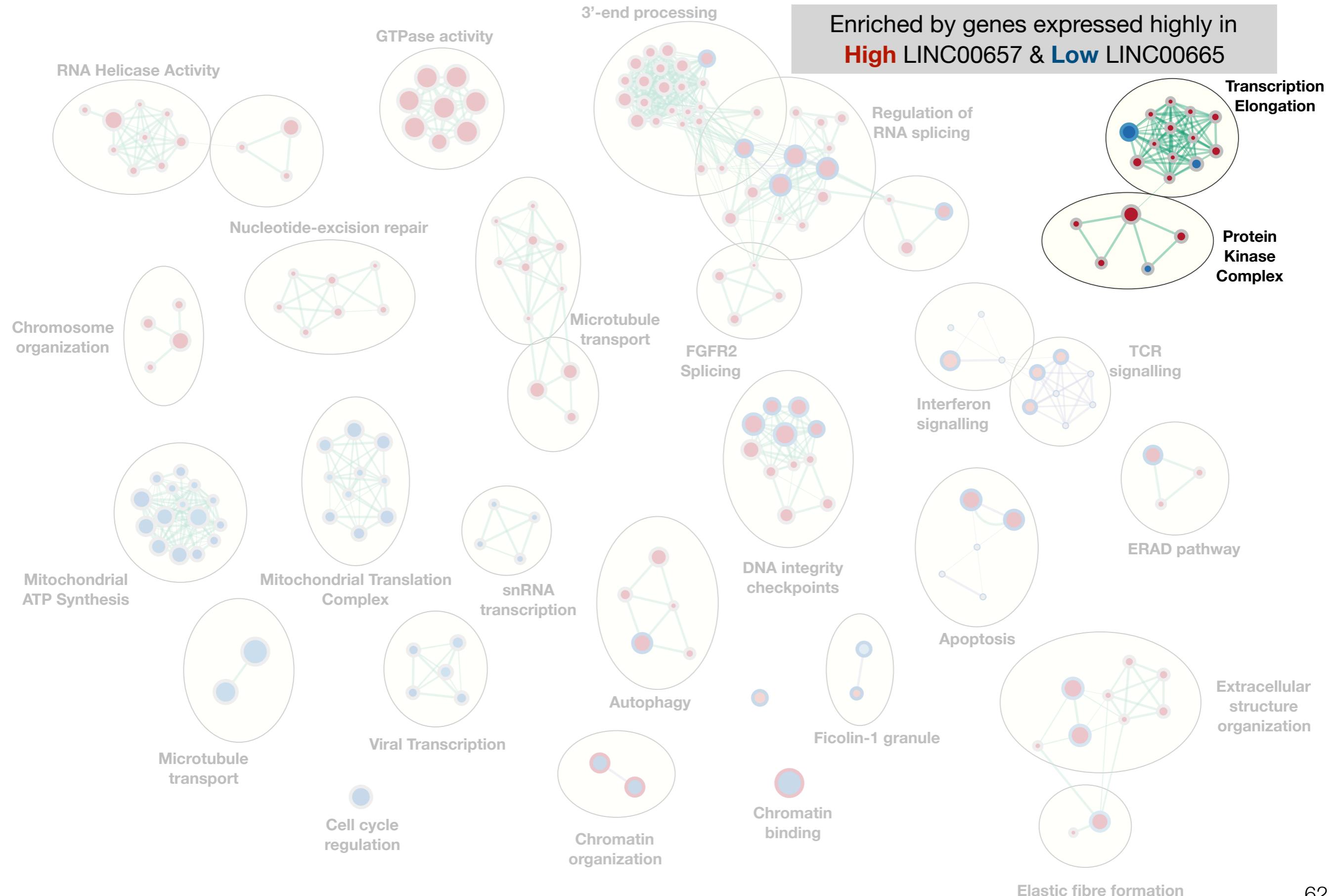
LINC00657 and LINC00665 associate with several common pathways



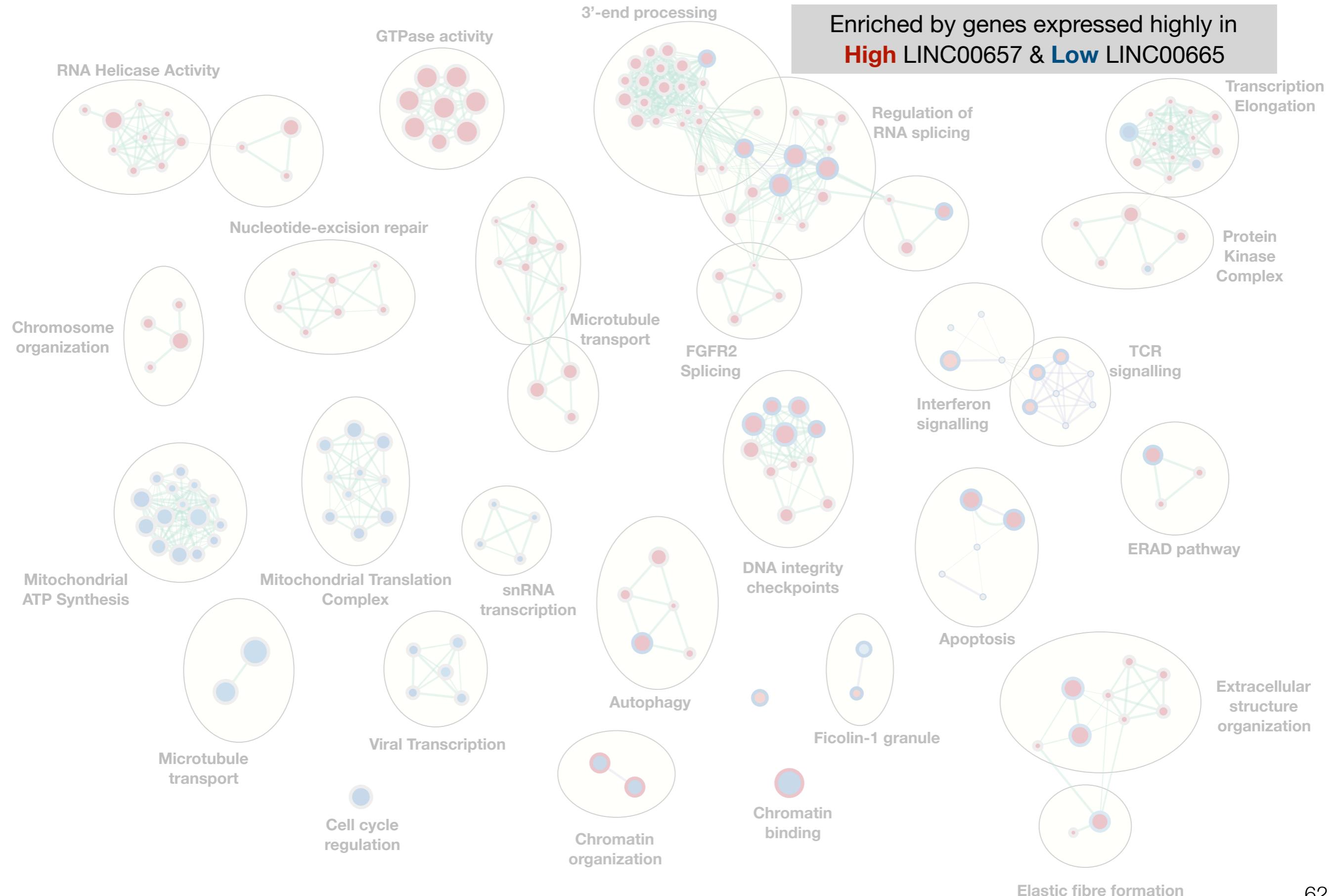
LINC00657 and LINC00665 associate with several common pathways



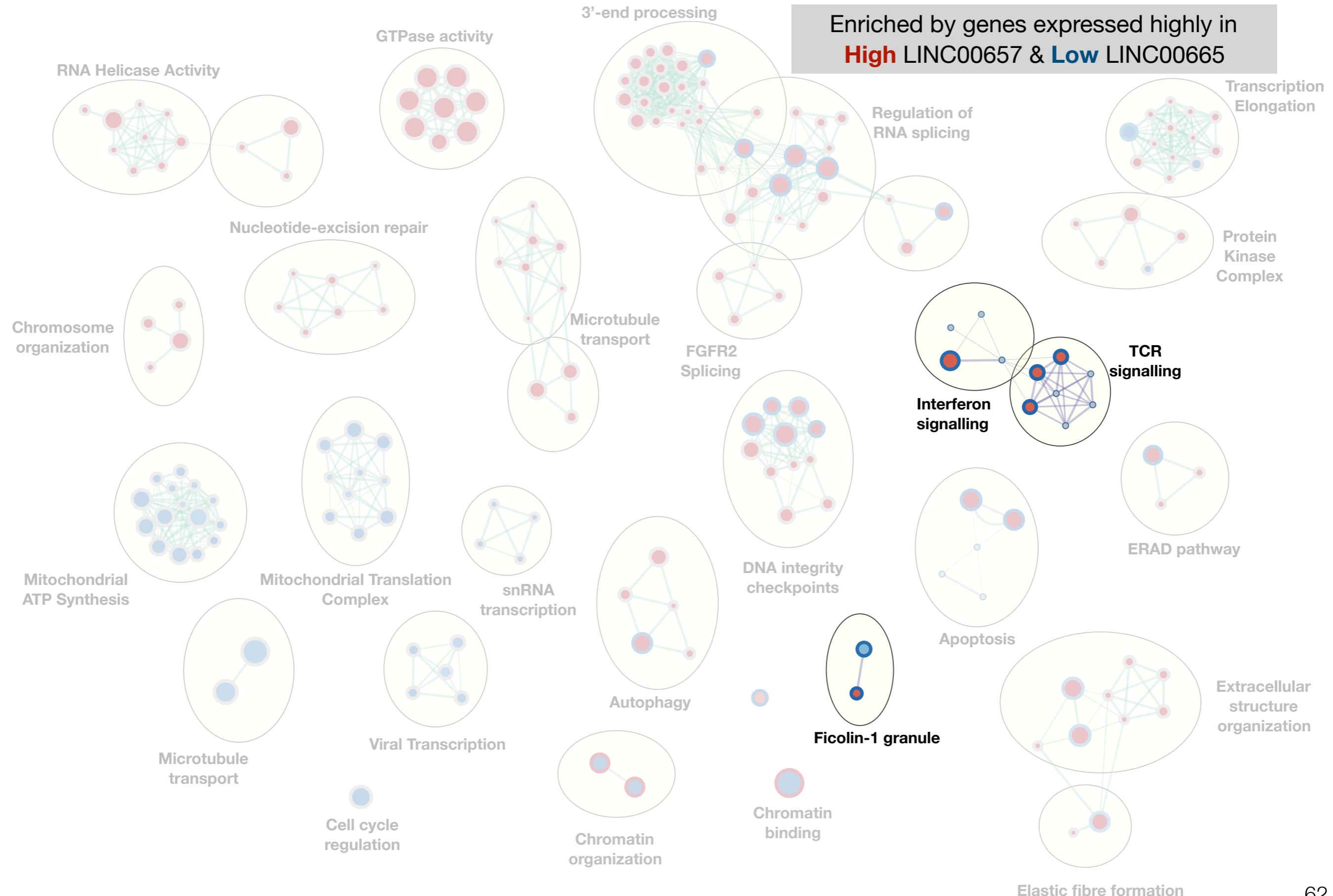
LINC00657 and LINC00665 associate with several common pathways



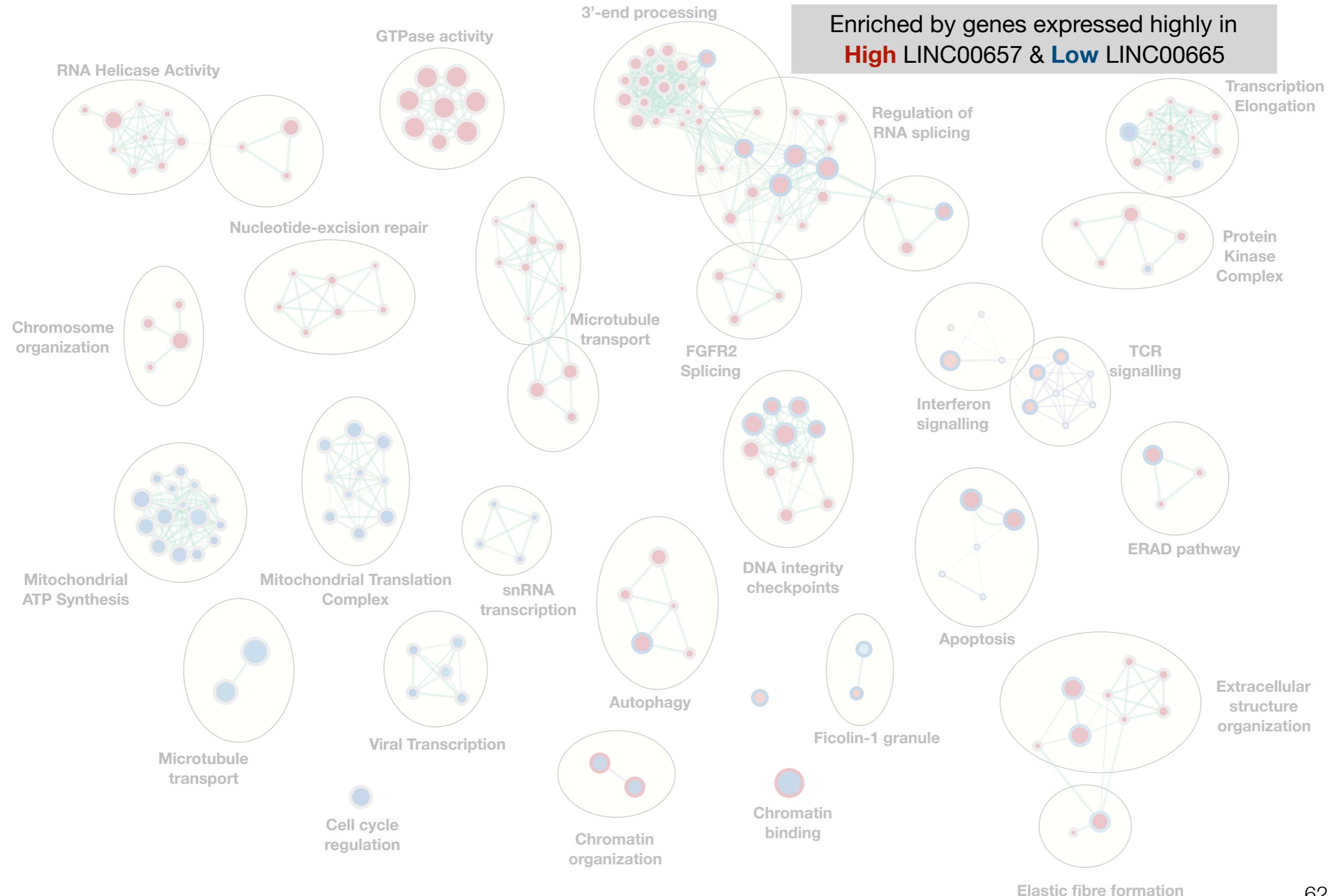
LINC00657 and LINC00665 associate with several common pathways



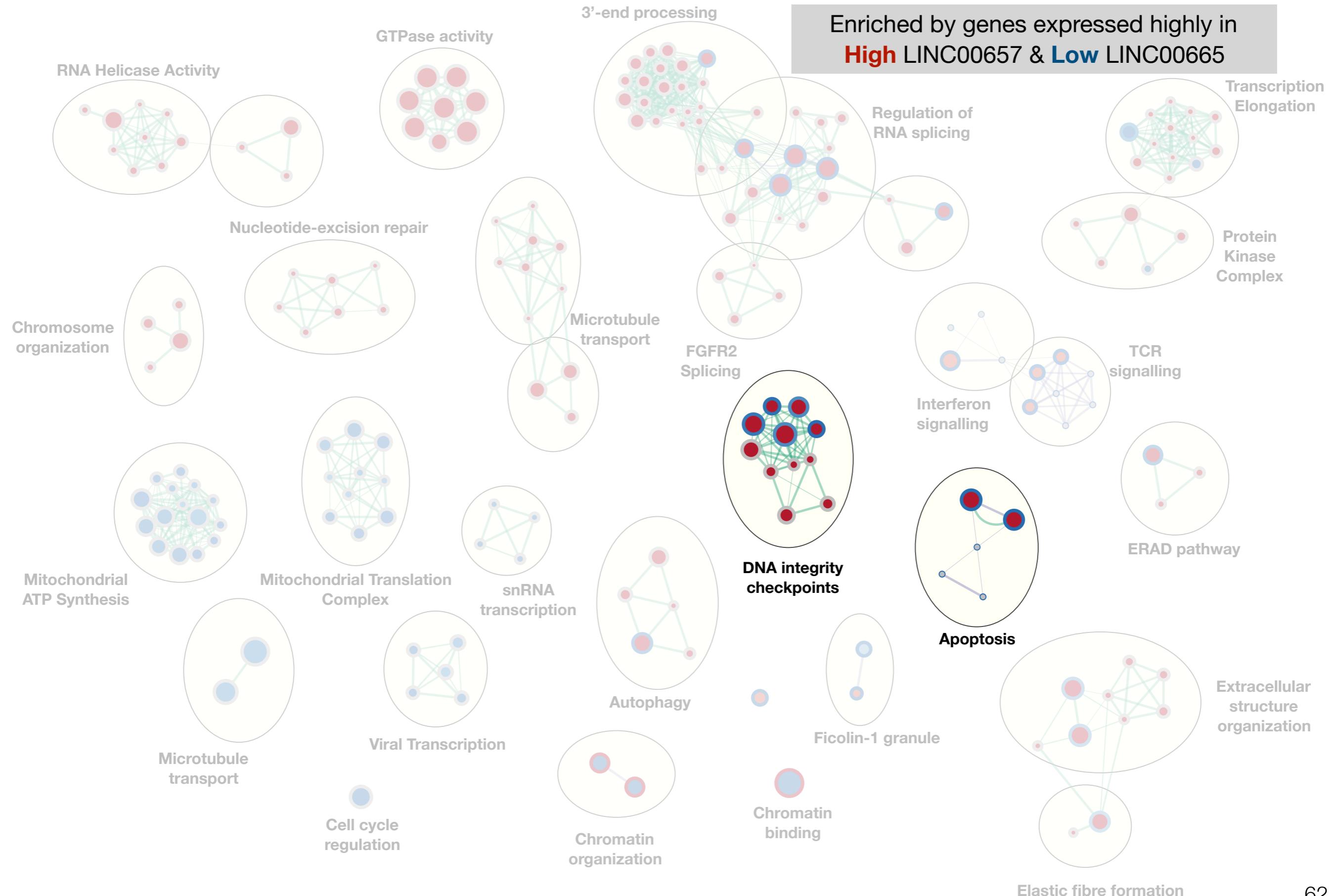
LINC00657 and LINC00665 associate with several common pathways



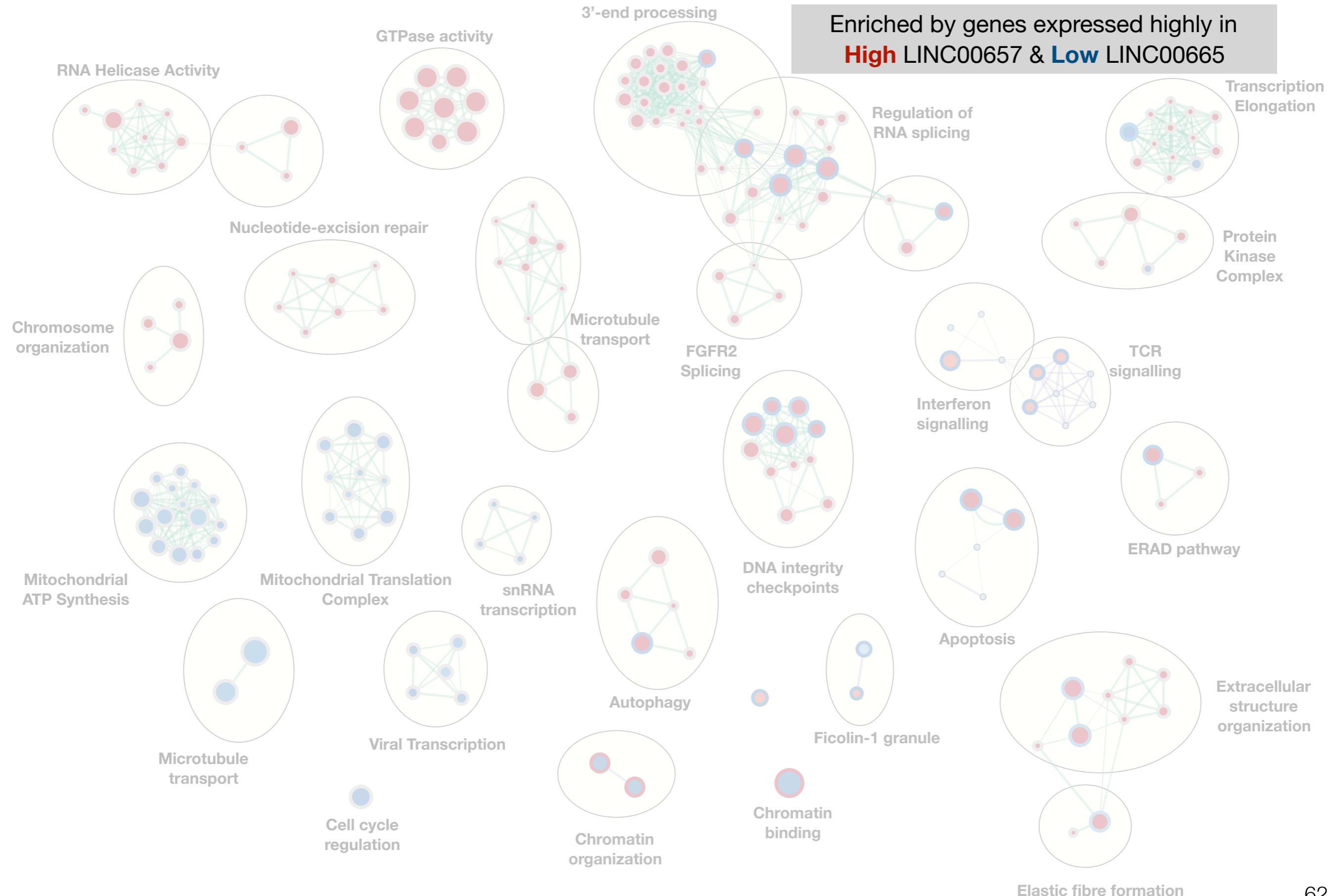
LINC00657 and LINC00665 associate with several common pathways



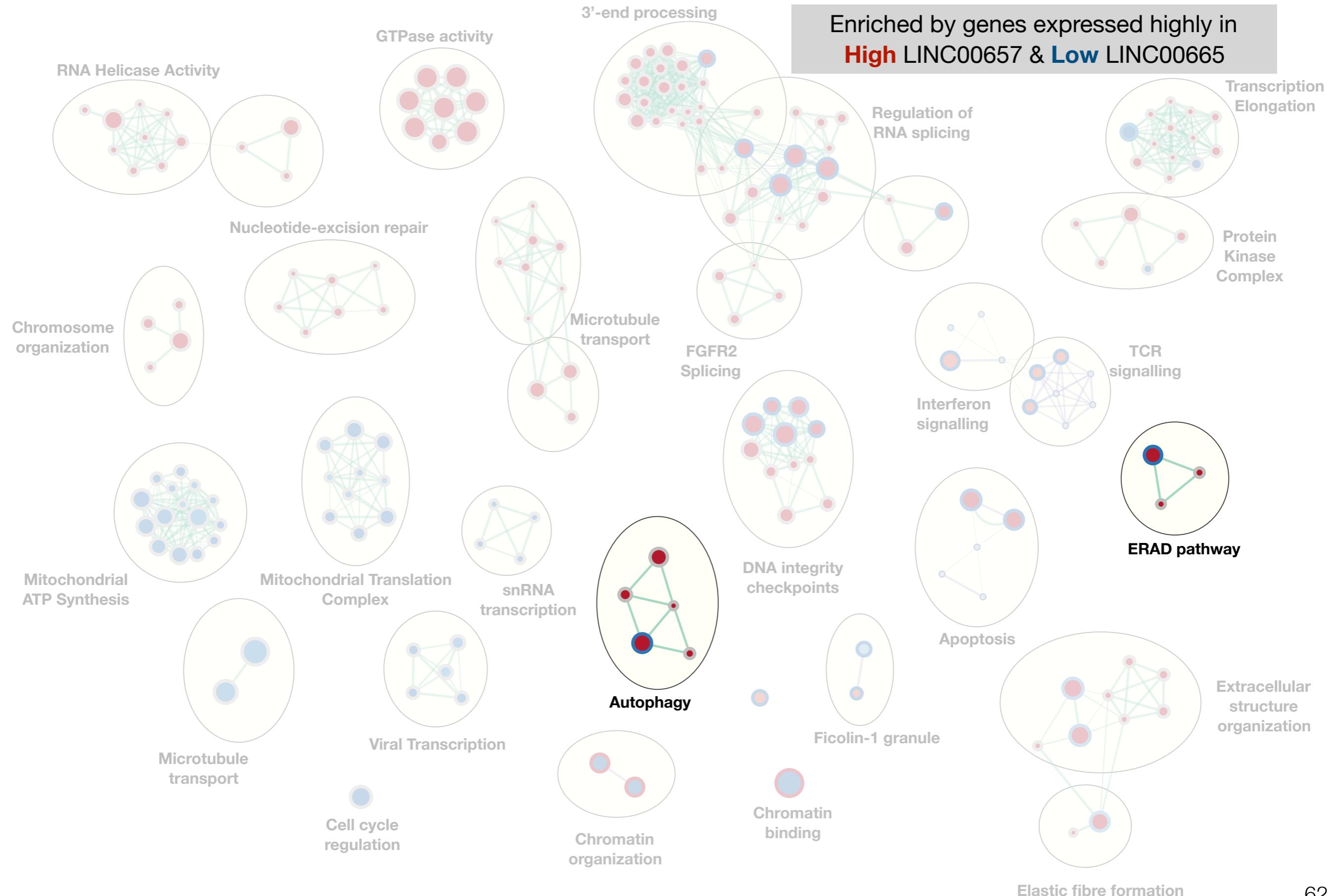
LINC00657 and LINC00665 associate with several common pathways



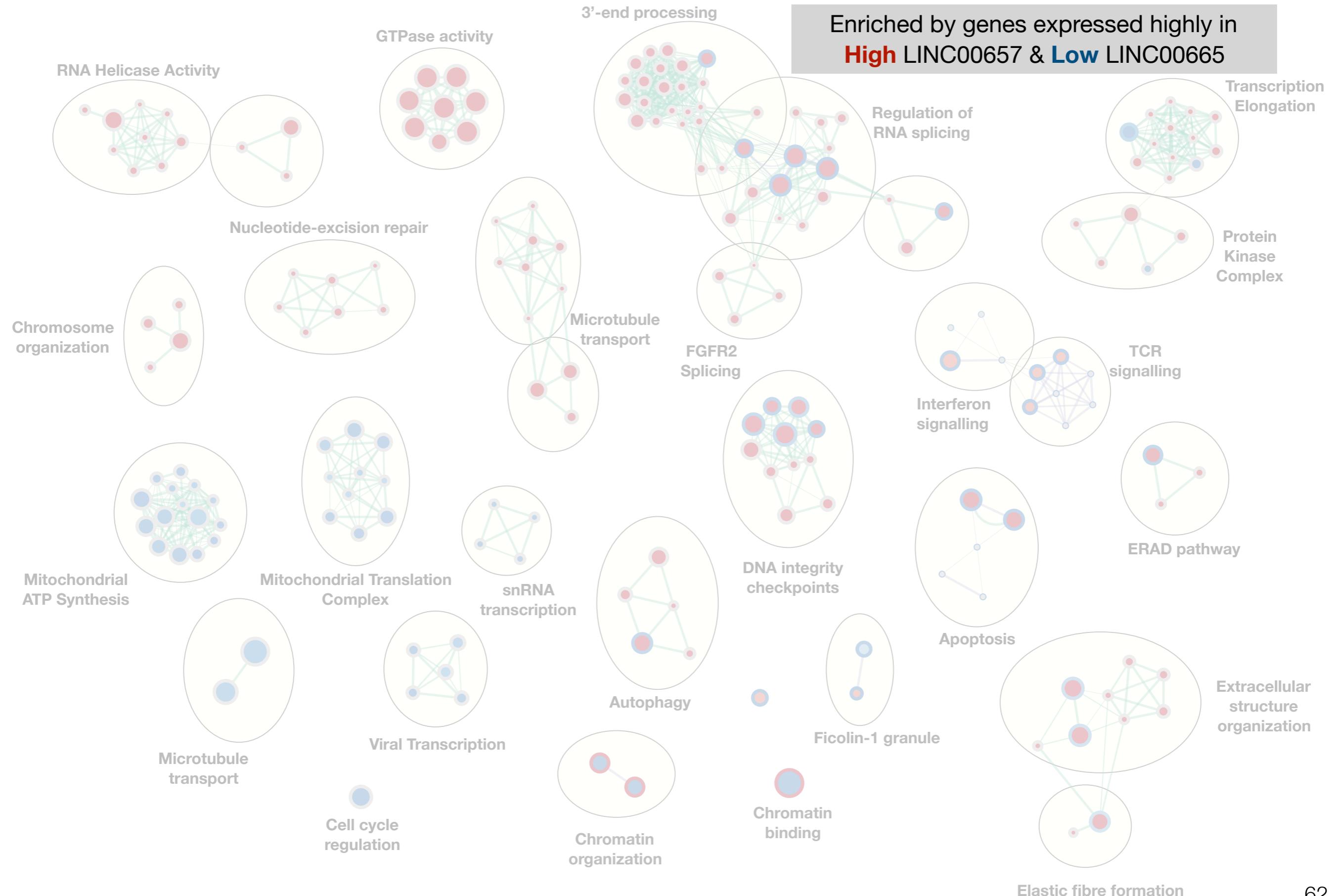
LINC00657 and LINC00665 associate with several common pathways



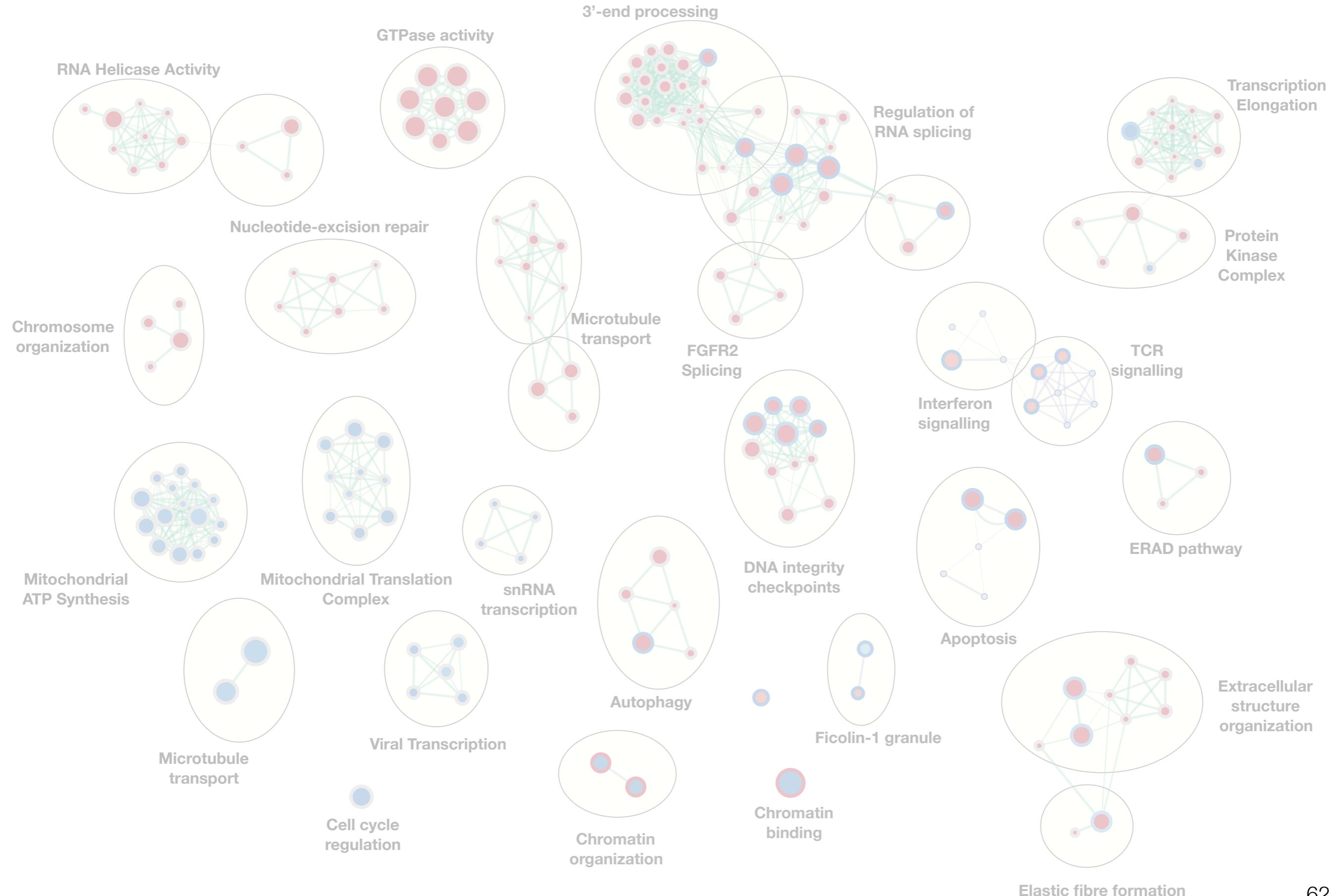
LINC00657 and LINC00665 associate with several common pathways



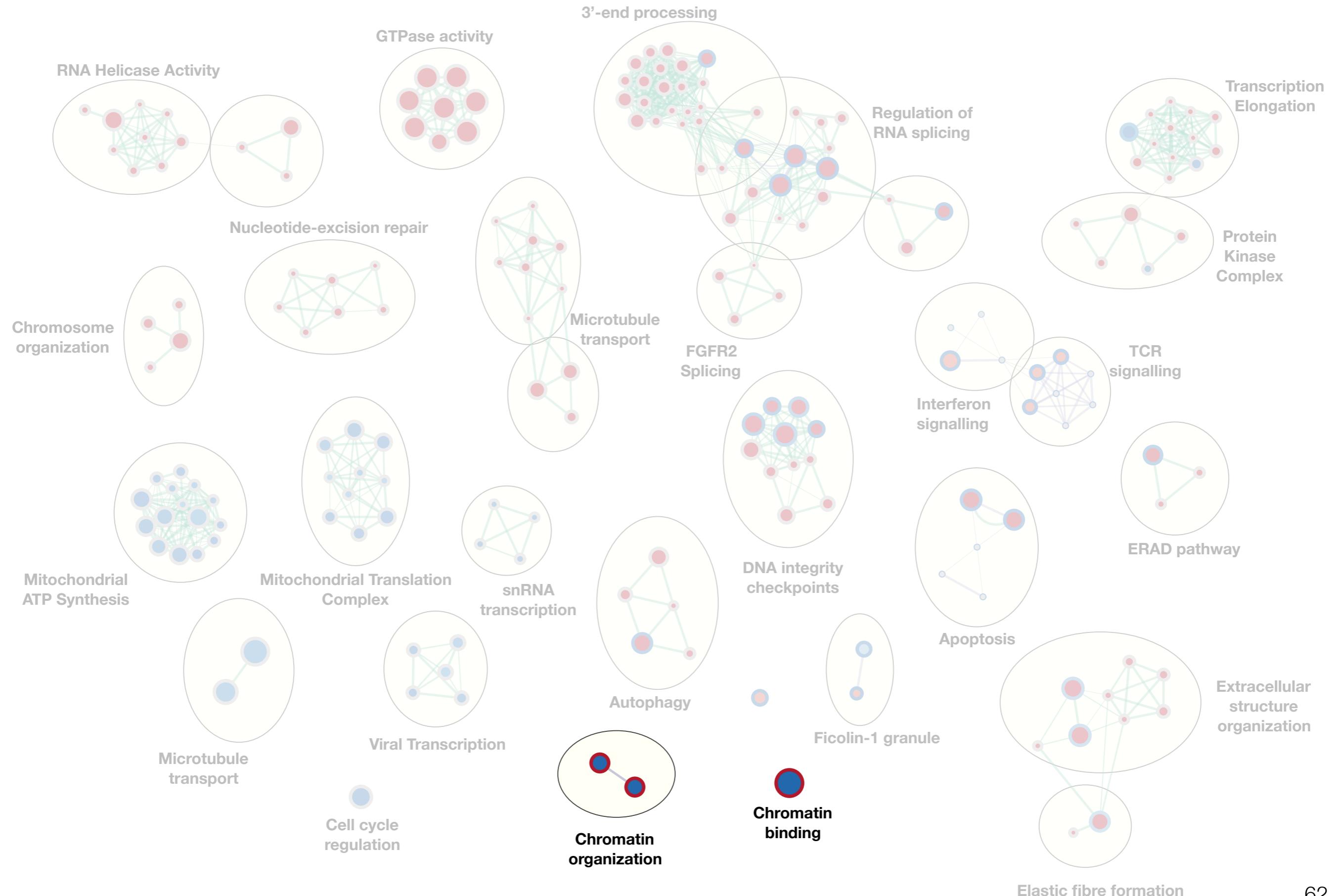
LINC00657 and LINC00665 associate with several common pathways



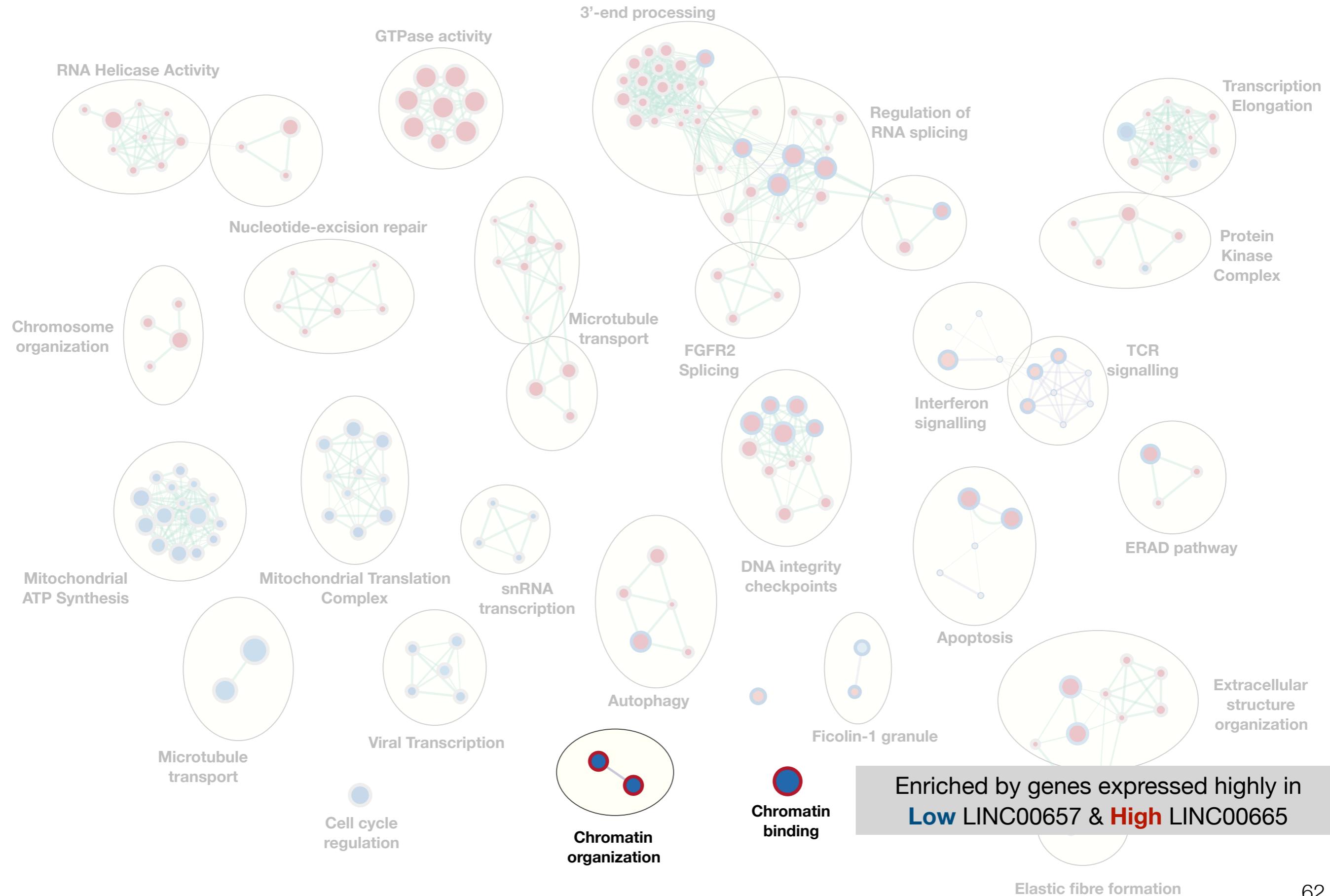
LINC00657 and LINC00665 associate with several common pathways



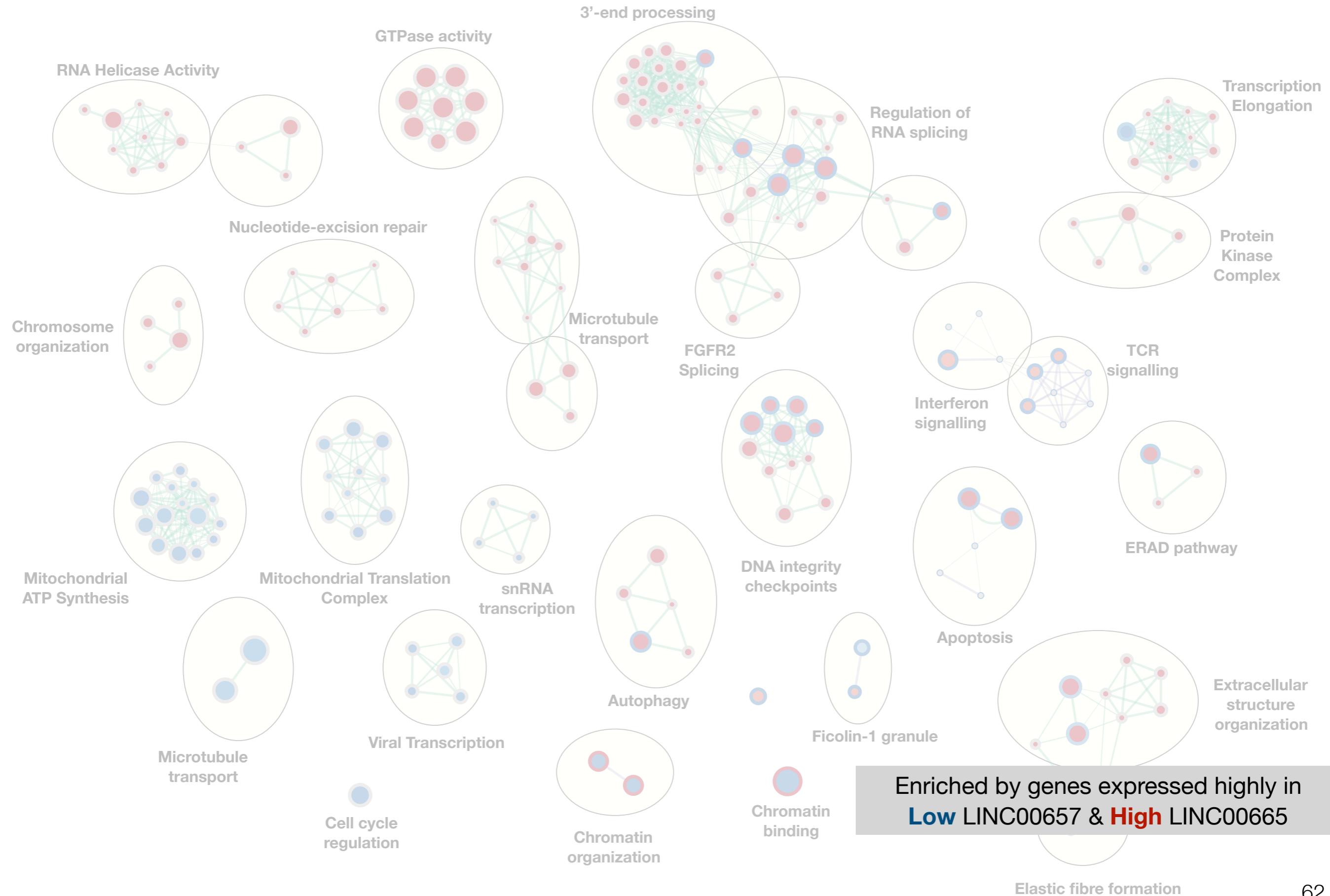
LINC00657 and LINC00665 associate with several common pathways



LINC00657 and LINC00665 associate with several common pathways

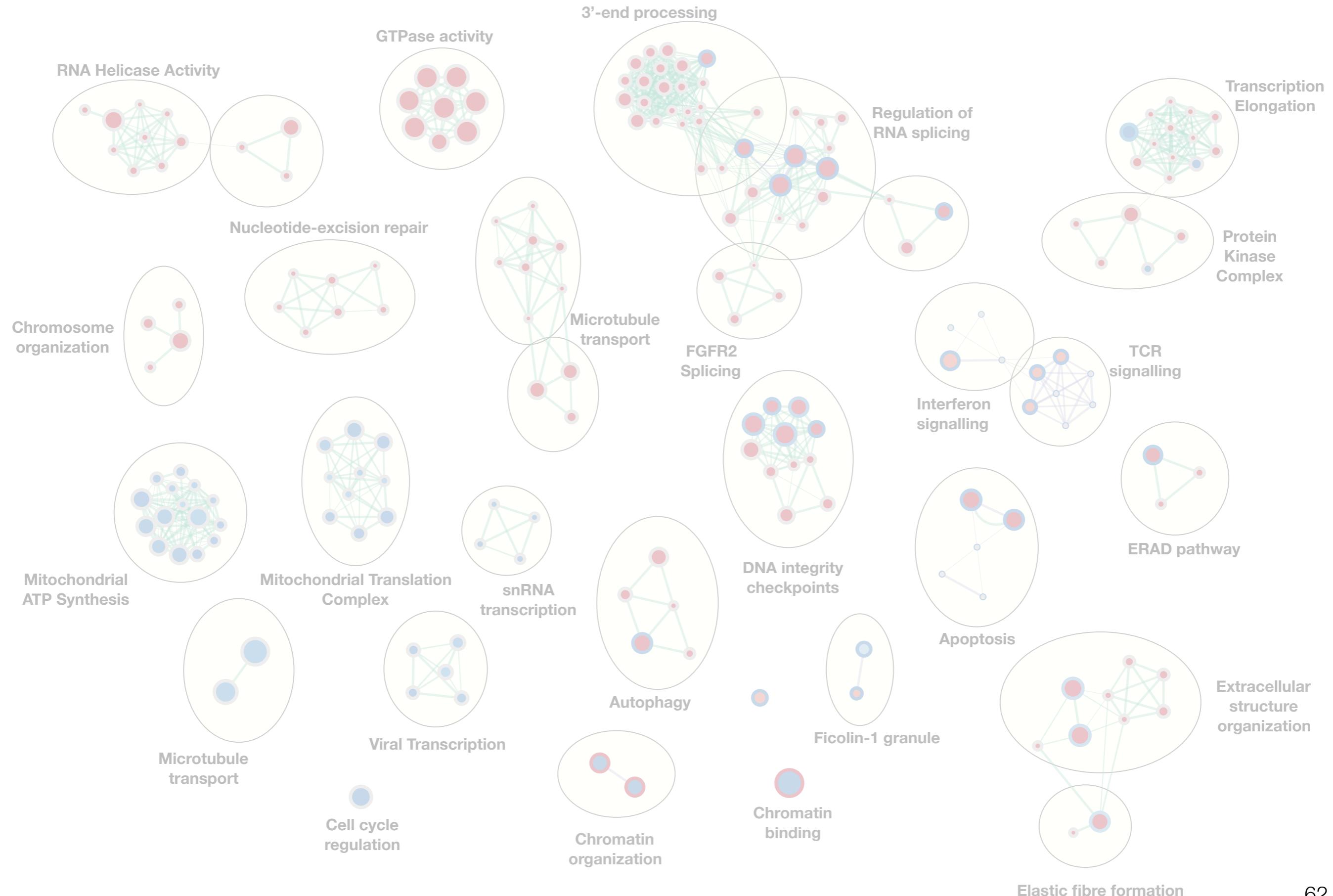


LINC00657 and LINC00665 associate with several common pathways

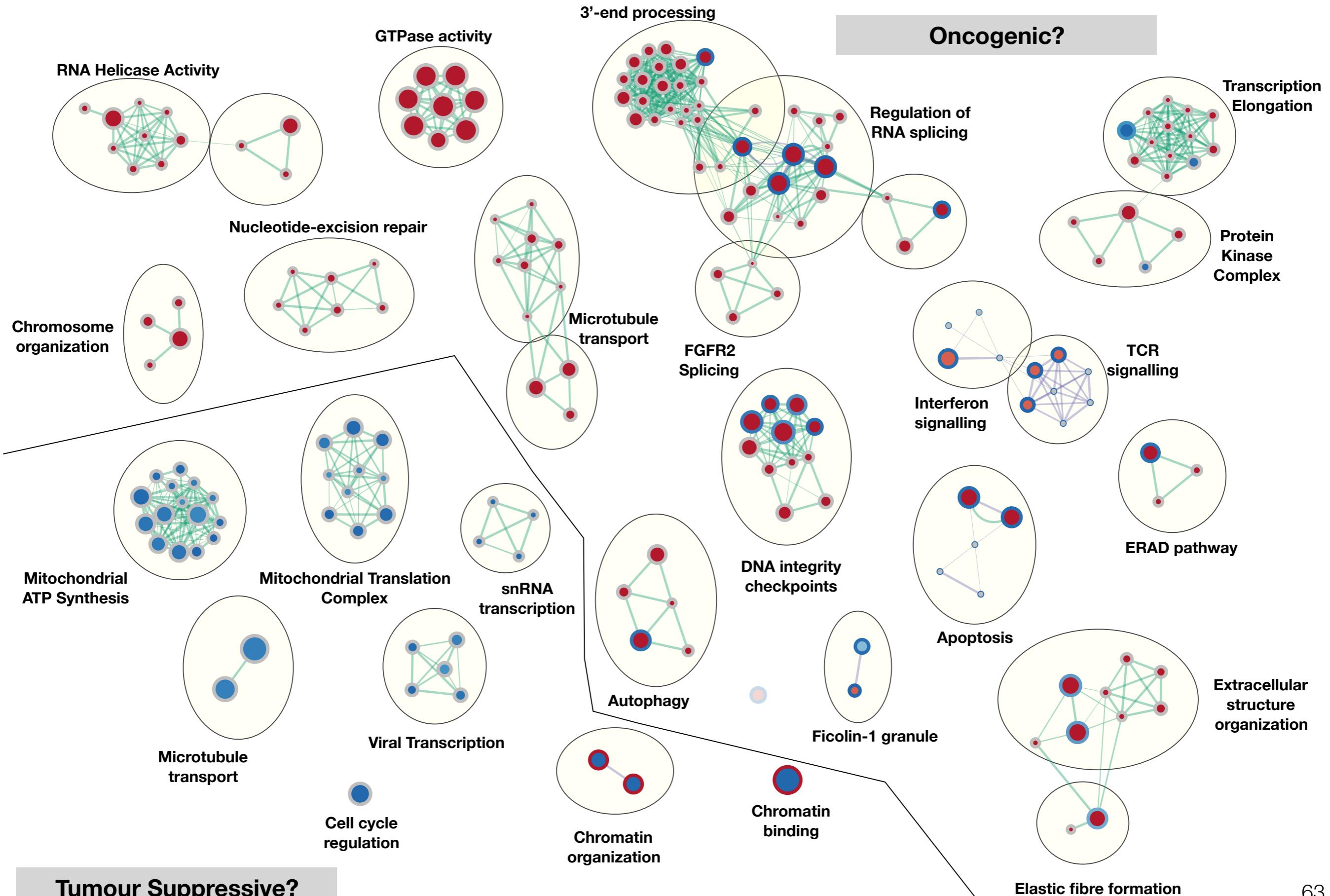


Enriched by genes expressed highly in
Low LINC00657 & **High** LINC00665

LINC00657 and LINC00665 associate with several common pathways



LINC00657 and LINC00665 associate with several common pathways



Tumour Suppressive?

63

Thesis: Genomic characterization of clinically relevant lncRNAs in multiple cancer types

1. Review of lncRNAs
2. Summary of the last SCM
3. Summary of progress since last meeting
4. Future directions

Thesis: Genomic characterization of clinically relevant lncRNAs in multiple cancer types

1. Review of lncRNAs

2. Summary of the last SCM

3. Summary of progress since last meeting

4. Future directions

Next Steps

1. Validation of survival associated lncRNAs in additional datasets

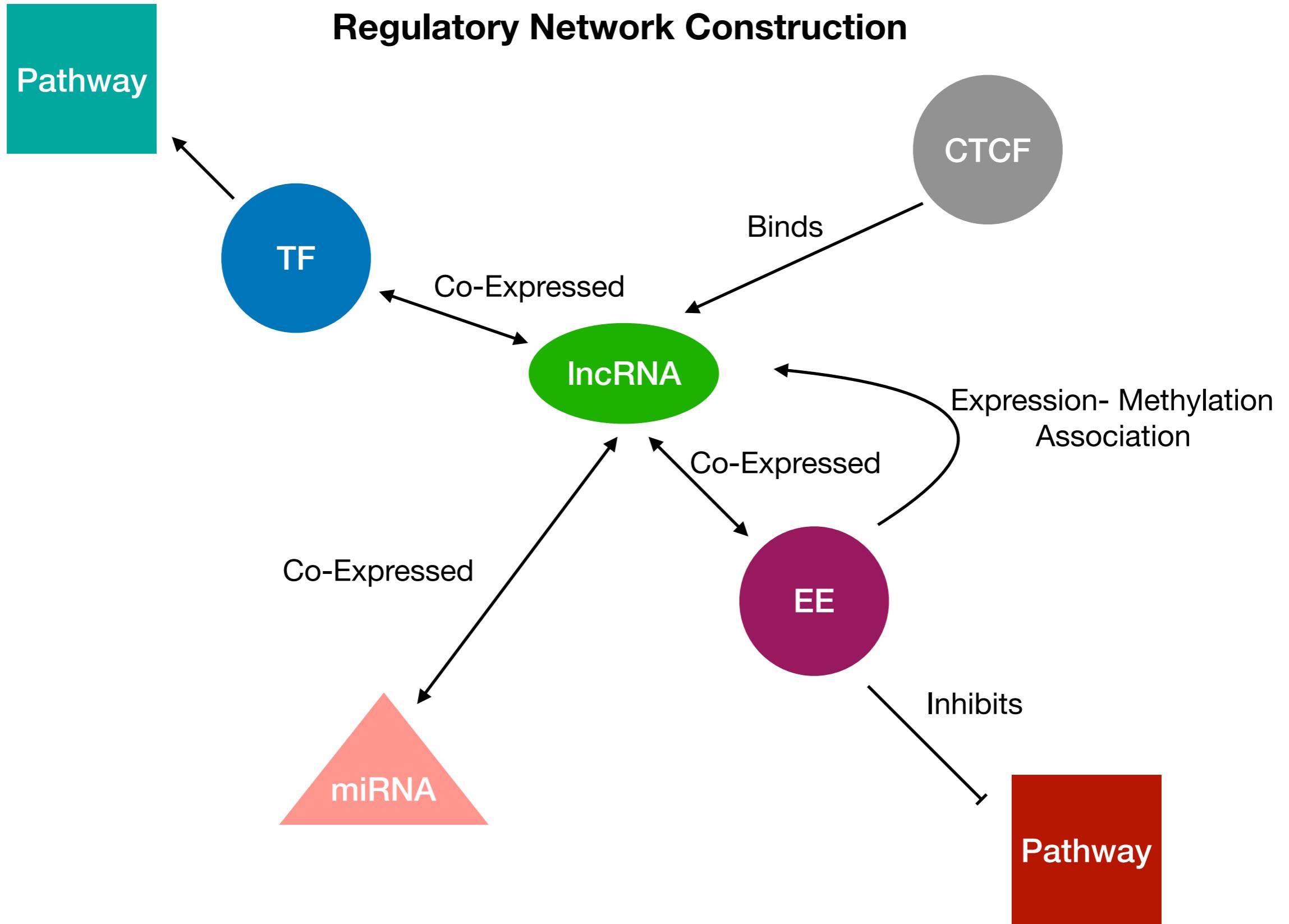
- TCGA datasets
- Identify clinical and experimental collaborations
- Compare performance of lncRNAs against other predictors (clinical features, combinations of lncRNA candidates)
- Additional predictors of survival will include:
 - CNAs
 - lncRNA promoter methylation
 - SNVs in lncRNA regulatory regions (and gene bodies)
- Validate known prognostic lncRNAs such as *FAL1* in our data

Next Steps

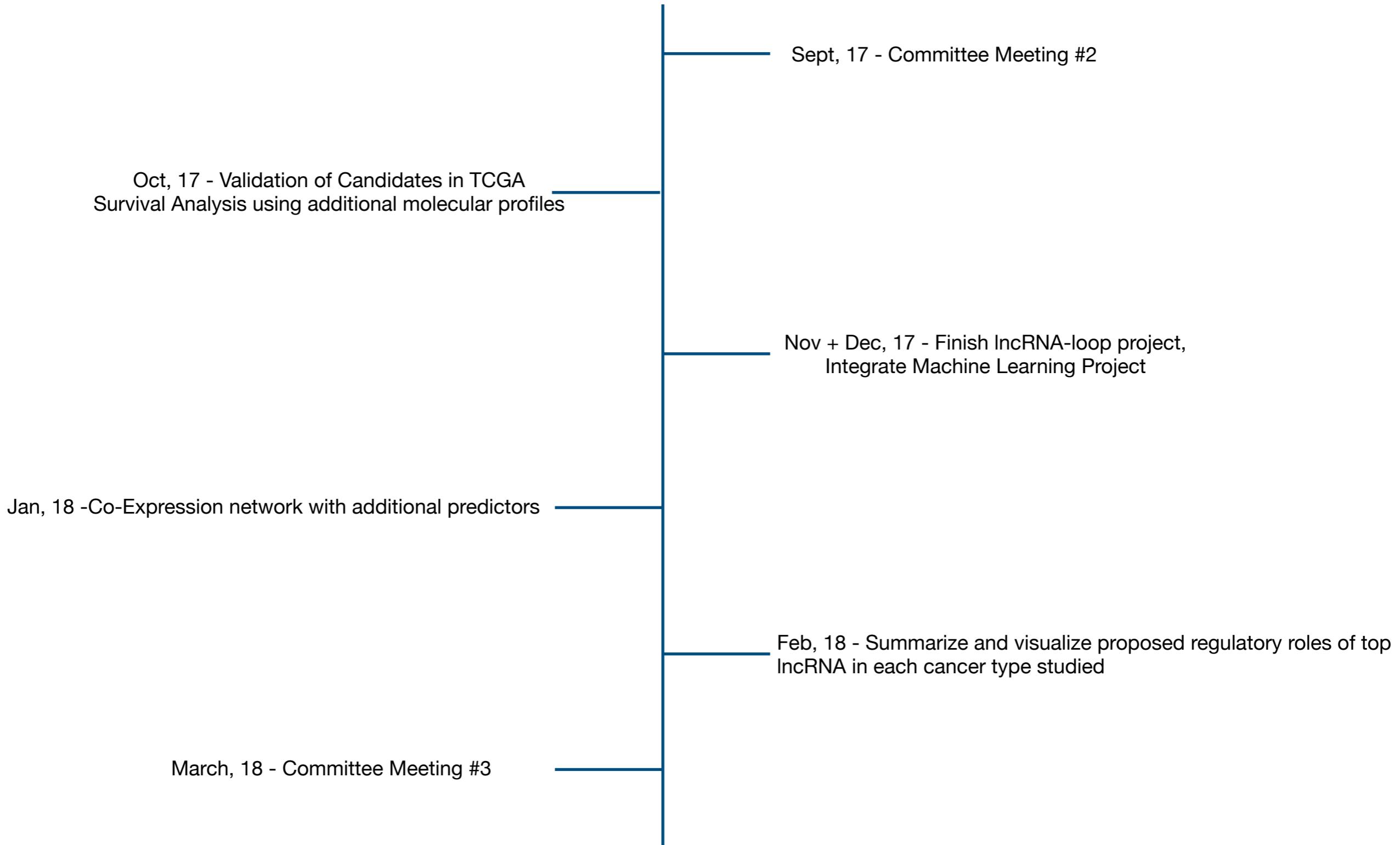
Next Steps

2. In-depth analysis of candidate lncRNAs to predict cellular mechanisms that could be experimentally validated
 - Predict interactions through co-expression analysis and assess permutations
 - Associate lncRNA molecular profiles to other mRNA molecular profiles
 - lncRNA E → DNA methylation ?
 - lncRNA E → More TP53 mutations ?
 - lncRNA E → Distal PCG E ?
 - lncRNA promoter methylation → PCG E ?
 - lncRNA copy number → lncRNA E
 - Integrate ENCODE ChIP-Seq binding data for Transcription Factors
 - Integrate available protein level data to evaluate protein-protein co-expression as well as lncRNA-co-expression to build a lncRNA-protein regulatory network

Future Directions



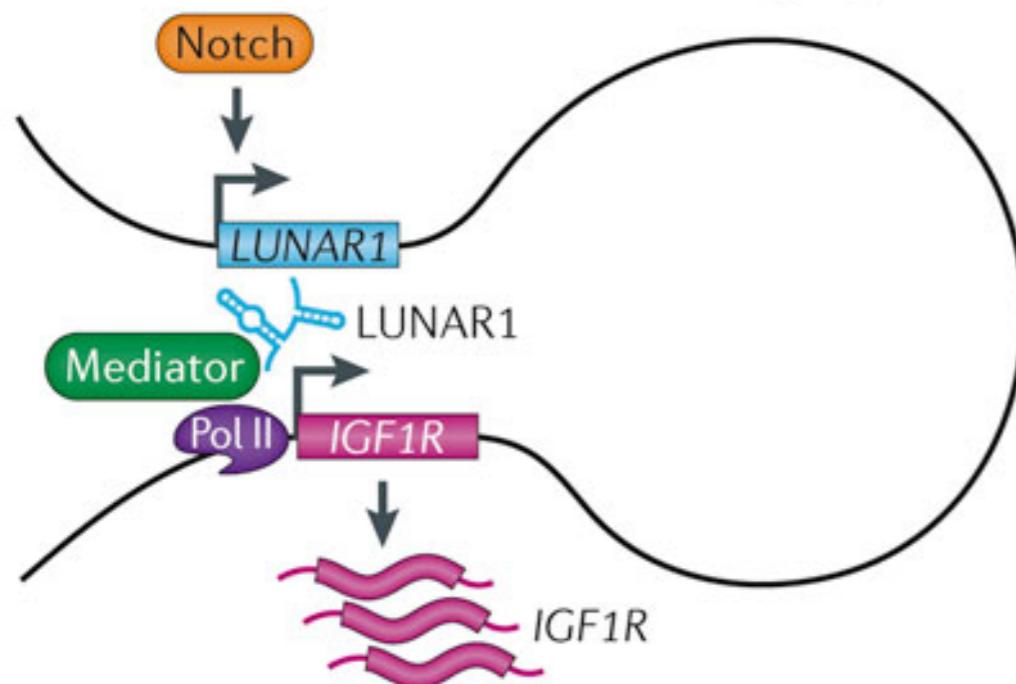
Timeline



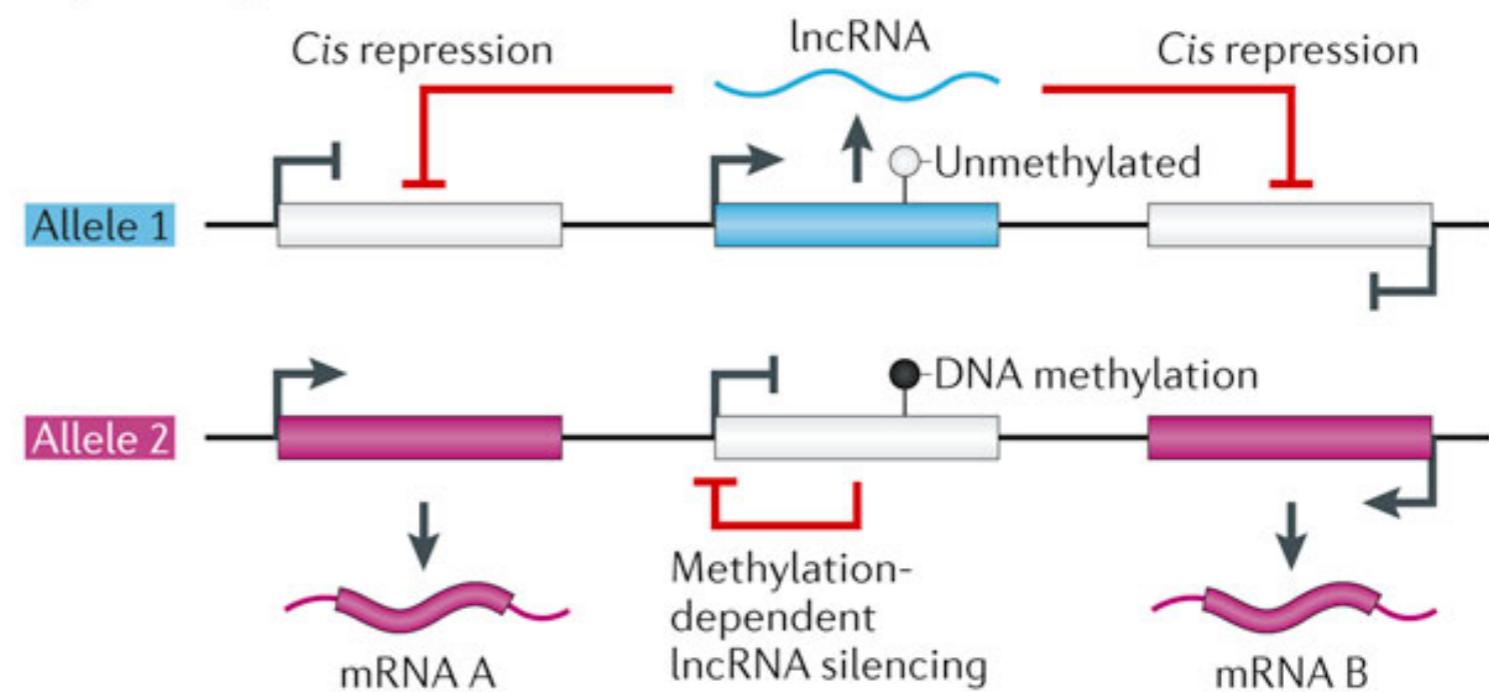
Thank you!

Cis-Regulatory Mechanisms of lncRNAs

a Enhancer RNAs and chromosome looping



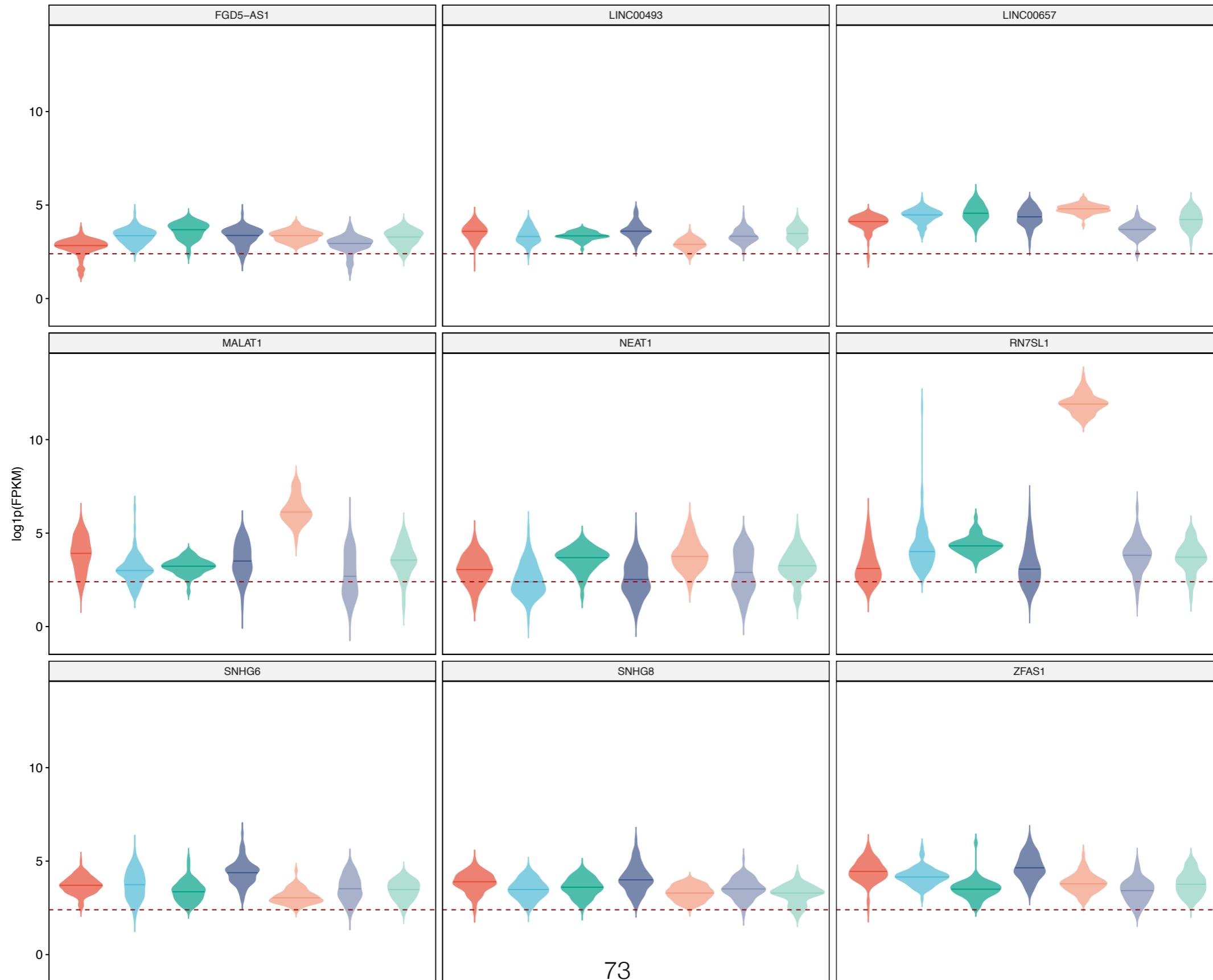
c Imprinted gene clusters



FDR problem

9 most highly expressed lncRNAs in all Cancer types

Cancer
Kidney Adenocarcinoma, clear cell type
Kidney Adenocarcinoma, chromophobe type
Pancreas Pancreatic ductal carcinoma
Kidney Adenocarcinoma, papillary type
Breast Infiltrating duct carcinoma
Ovary Serous cystadenocarcinoma
Liver Hepatocellular carcinoma

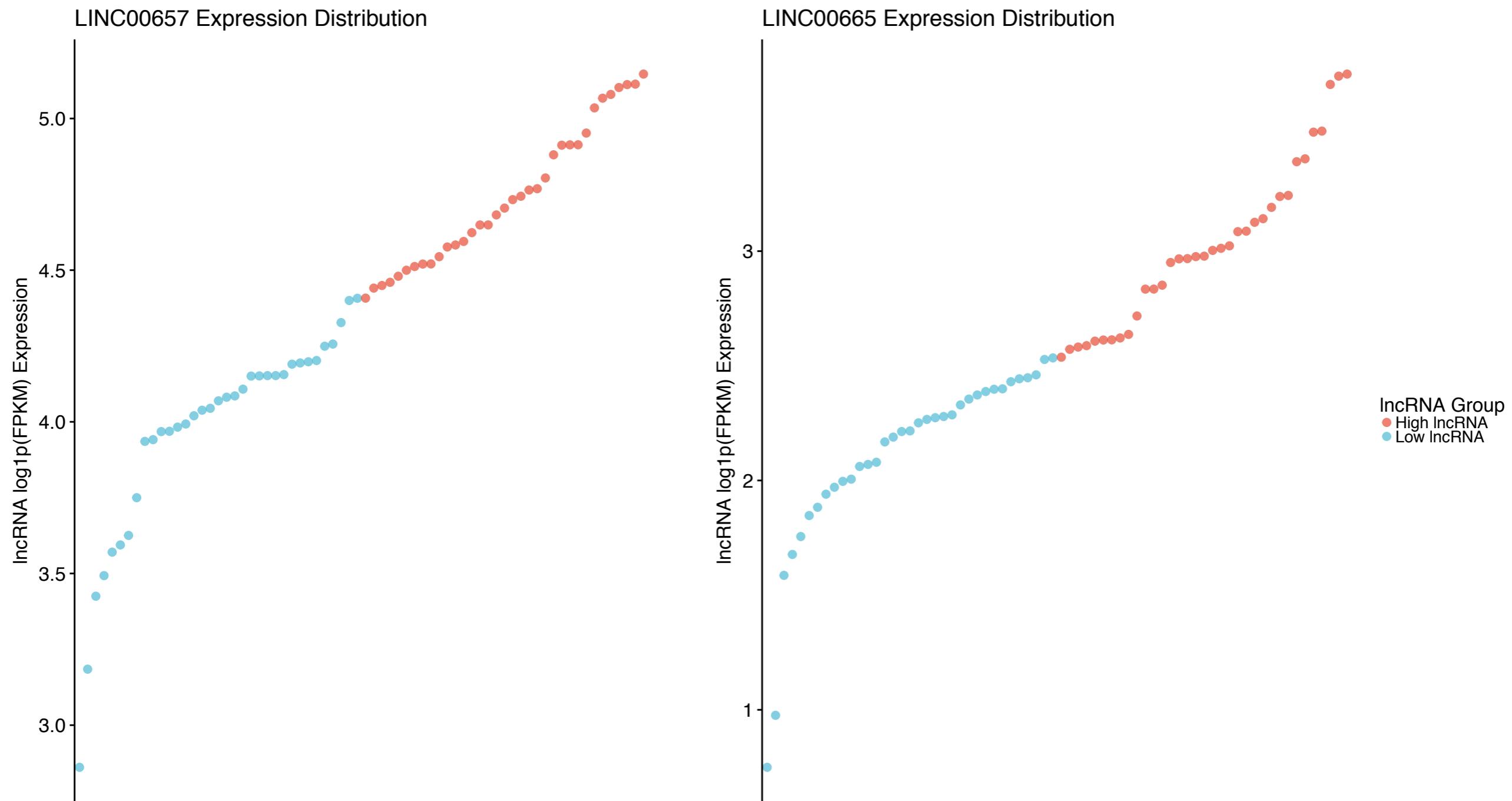


42 lncRNA-Cancer Survival Associations with p-values < 0.05

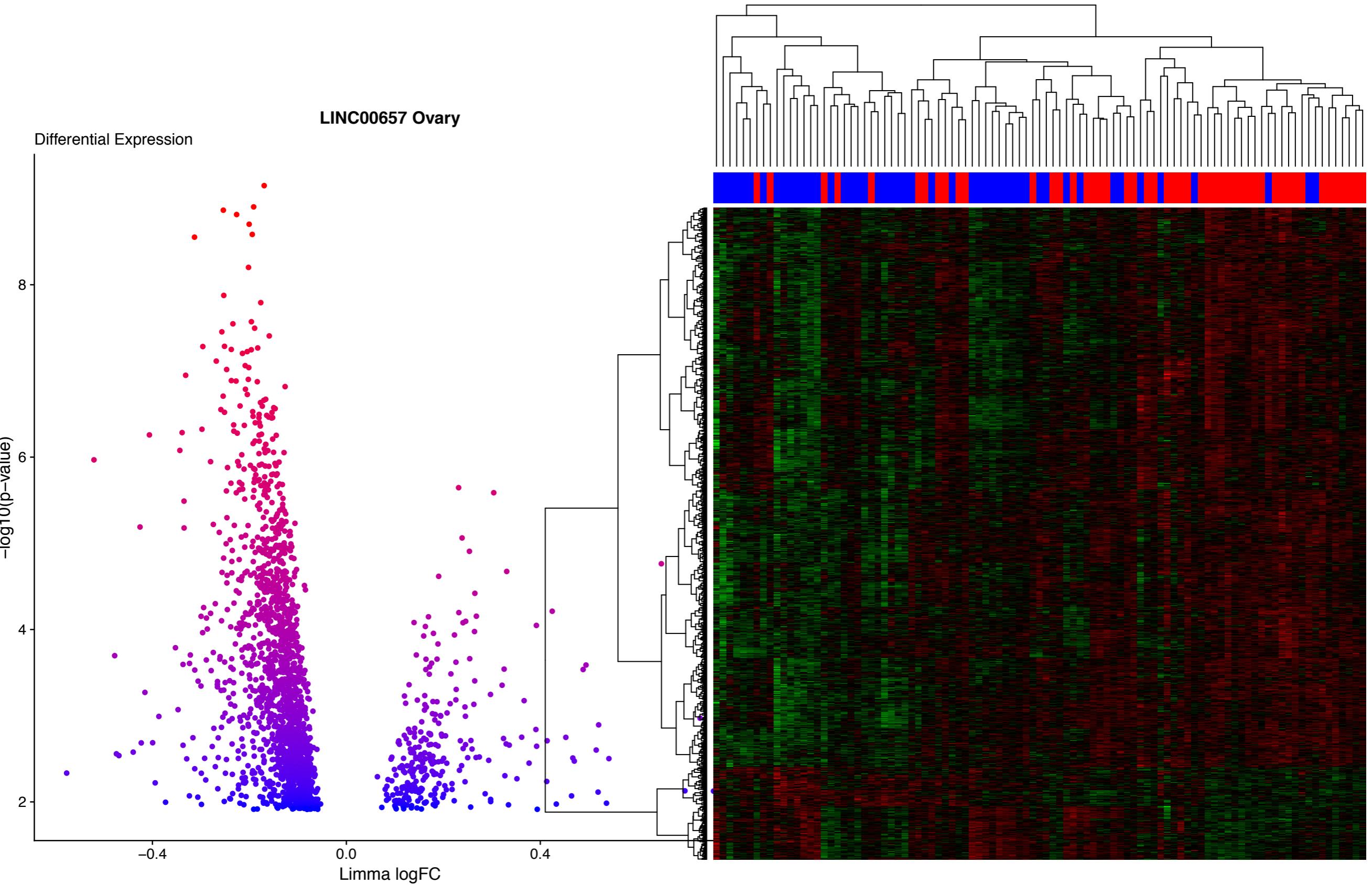
42 lncRNA-Cancer Survival Associations with p-values < 0.05

	gene	coef	HR	pval	canc	fdr
1	LINC00657	1.09089497046119	2.97693715151224	6.714862e-05	Ovary Serous cystadenocarcinoma	0.01103476
2	LINC00665	-1.23789638086889	0.28999361285883	4.933441e-05	Ovary Serous cystadenocarcinoma	0.01103476
3	ZNF503-AS2	-1.21096375798916	0.29791002786222	5.909242e-05	Ovary Serous cystadenocarcinoma	0.01103476
4	AC009336.24	-0.915329978890643	0.400384485861799	6.039664e-04	Ovary Serous cystadenocarcinoma	0.05014560
5	RP11-622K12.1	-0.969070407483862	0.379435594700213	6.102913e-04	Ovary Serous cystadenocarcinoma	0.05014560
6	AP000355.2	-1.50935541867795	0.221052418307288	4.777307e-04	Liver Hepatocellular carcinoma	0.05014560
7	RP11-622A1.2	-1.42689280391215	0.240053658412211	1.416813e-03	Liver Hepatocellular carcinoma	0.09978412
8	SNHG15	0.778032349402381	2.17718411047022	3.557492e-03	Ovary Serous cystadenocarcinoma	0.19487153
9	RP11-220I1.1	-1.31154573241281	0.269403308950893	3.371553e-03	Liver Hepatocellular carcinoma	0.19487153
10	SNHG12	0.764311259067298	2.14751478385111	4.537974e-03	Ovary Serous cystadenocarcinoma	0.22372214
11	SNHG1	0.743189814071518	2.10263183341979	5.168054e-03	Ovary Serous cystadenocarcinoma	0.23162277
12	GS1-251I9.4	0.725427456398557	2.06561387114612	6.281176e-03	Ovary Serous cystadenocarcinoma	0.23892283
13	RP11-295G20.2	1.23237659225222	3.42937007343585	6.300196e-03	Liver Hepatocellular carcinoma	0.23892283
14	OSER1-AS1	0.699393073114752	2.01253087763027	9.189207e-03	Ovary Serous cystadenocarcinoma	0.31979542
15	RP11-157P1.4	0.66631704328976	1.94705318673016	9.730084e-03	Ovary Serous cystadenocarcinoma	0.31979542
16	LINC00493	1.07830444424117	2.93969091314685	1.054080e-02	Liver Hepatocellular carcinoma	0.32478839
17	U47924.27	0.680113191614456	1.97410117128317	1.176238e-02	Ovary Serous cystadenocarcinoma	0.34110901
18	AC006126.4	-0.918822841885361	0.398988437230102	1.473127e-02	Kidney Adenocarcinoma, clear cell type	0.36995216
19	NEAT1	0.642552733080203	1.90132827320059	1.365801e-02	Ovary Serous cystadenocarcinoma	0.36995216
20	MMP24-AS1	0.998979579568135	2.71550945287663	1.500820e-02	Liver Hepatocellular carcinoma	0.36995216
21	FGD5-AS1	-0.866125022806513	0.420578126385491	1.824250e-02	Kidney Adenocarcinoma, clear cell type	0.40879774
22	ADORA2A-AS1	-1.0238468792094	0.359210439705118	1.798022e-02	Liver Hepatocellular carcinoma	0.40879774
23	RP11-49I11.1	0.979396006633756	2.66284741394427	1.917262e-02	Liver Hepatocellular carcinoma	0.41096102
24	RP11-304L19.5	-0.727854754389591	0.48294391292074	2.111194e-02	Pancreas Pancreatic ductal carcinoma	0.43367449
25	MMP24-AS1	0.825276522328797	2.28251184356216	2.453876e-02	Kidney Adenocarcinoma, clear cell type	0.45015634
26	NCBP2-AS2	0.601706131619117	1.82523022838774	2.465359e-02	Ovary Serous cystadenocarcinoma	0.45015634
27	NEAT1	-0.931193692129404	0.394083015701056	2.327743e-02	Liver Hepatocellular carcinoma	0.45015634
28	CTB-131K11.1	0.579258152521435	1.78471395398422	2.992787e-02	Ovary Serous cystadenocarcinoma	0.48644962
29	DANCR	-0.569578402465635	0.56576391309669	2.961290e-02	Ovary Serous cystadenocarcinoma	0.48644962
30	FAM83H-AS1	-0.572351978944779	0.564196897737599	2.971973e-02	Ovary Serous cystadenocarcinoma	0.48644962
31	DANCR	0.910329458260571	2.48514114880777	3.058811e-02	Liver Hepatocellular carcinoma	0.48644962
32	RP11-834C11.4	-0.563198488147974	0.569384977143017	3.520396e-02	Ovary Serous cystadenocarcinoma	0.52126585
33	RPPH1	-0.639134723510223	0.527748875230572	3.412066e-02	Pancreas Pancreatic ductal carcinoma	0.52126585
34	RP11-139H15.1	0.895710166661311	2.44907442303575	3.594937e-02	Liver Hepatocellular carcinoma	0.52126585
35	CTC-503J8.6	-0.551357810106765	0.576166953702487	3.954573e-02	Ovary Serous cystadenocarcinoma	0.53743244
36	MFI2-AS1	0.539284764289622	1.7147799516152	3.840014e-02	Ovary Serous cystadenocarcinoma	0.53743244
37	RP11-469A15.2	-0.881443557345828	0.414184580728924	4.033469e-02	Liver Hepatocellular carcinoma	0.53743244
38	HOXD-AS1	-0.521580642937459	0.593581565558328	4.176834e-02	Ovary Serous cystadenocarcinoma	0.54188927
39	SNHG7	0.734356702478046	2.08414083853393	4.533632e-02	Kidney Adenocarcinoma, clear cell type	0.55960215
40	RP11-834C11.4	-0.600714941409506	0.548419408156069	4.563384e-02	Pancreas Pancreatic ductal carcinoma	0.55960215
41	DBH-AS1	-0.83141909496413	0.435430929832266	4.653892e-02	Liver Hepatocellular carcinoma	0.55960215
42	SNHG8	-0.599508861430317	0.549081244858215	4.976660e-02	Pancreas Pancreatic ductal carcinoma	0.58416513

Top Two Candidates in Ovarian Cancer



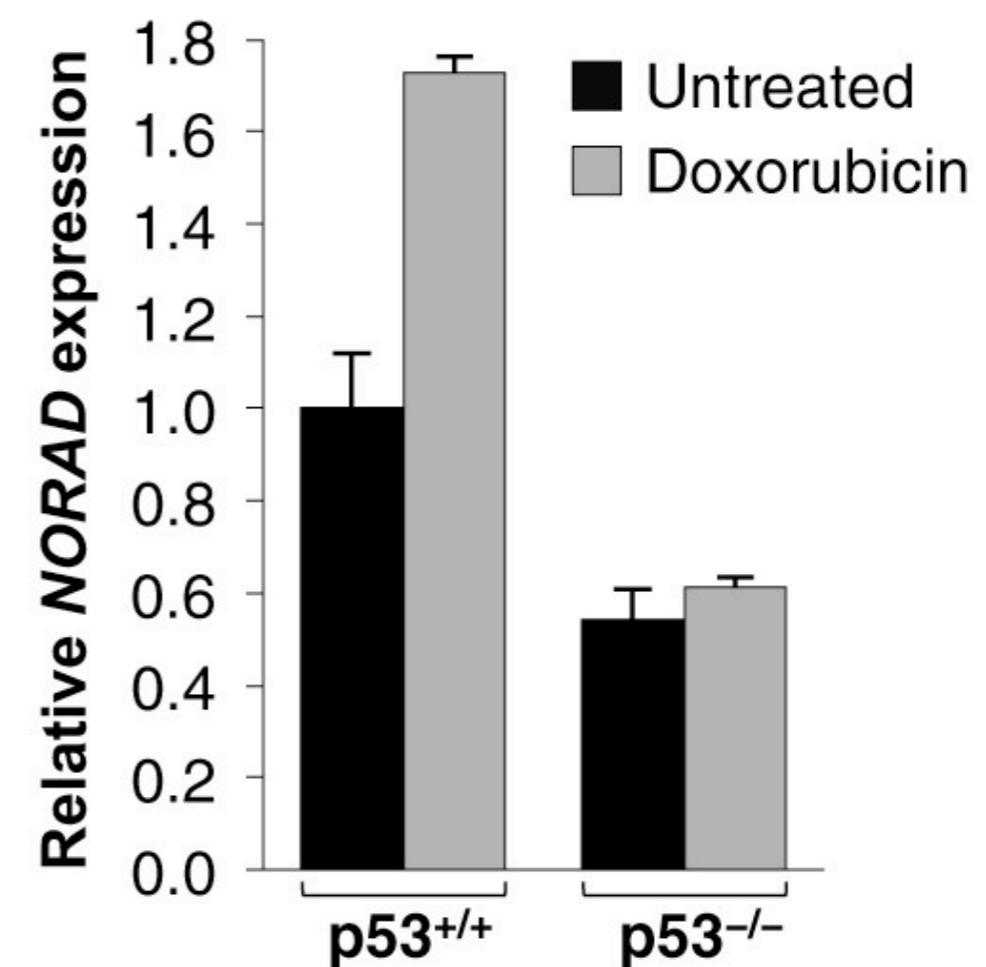
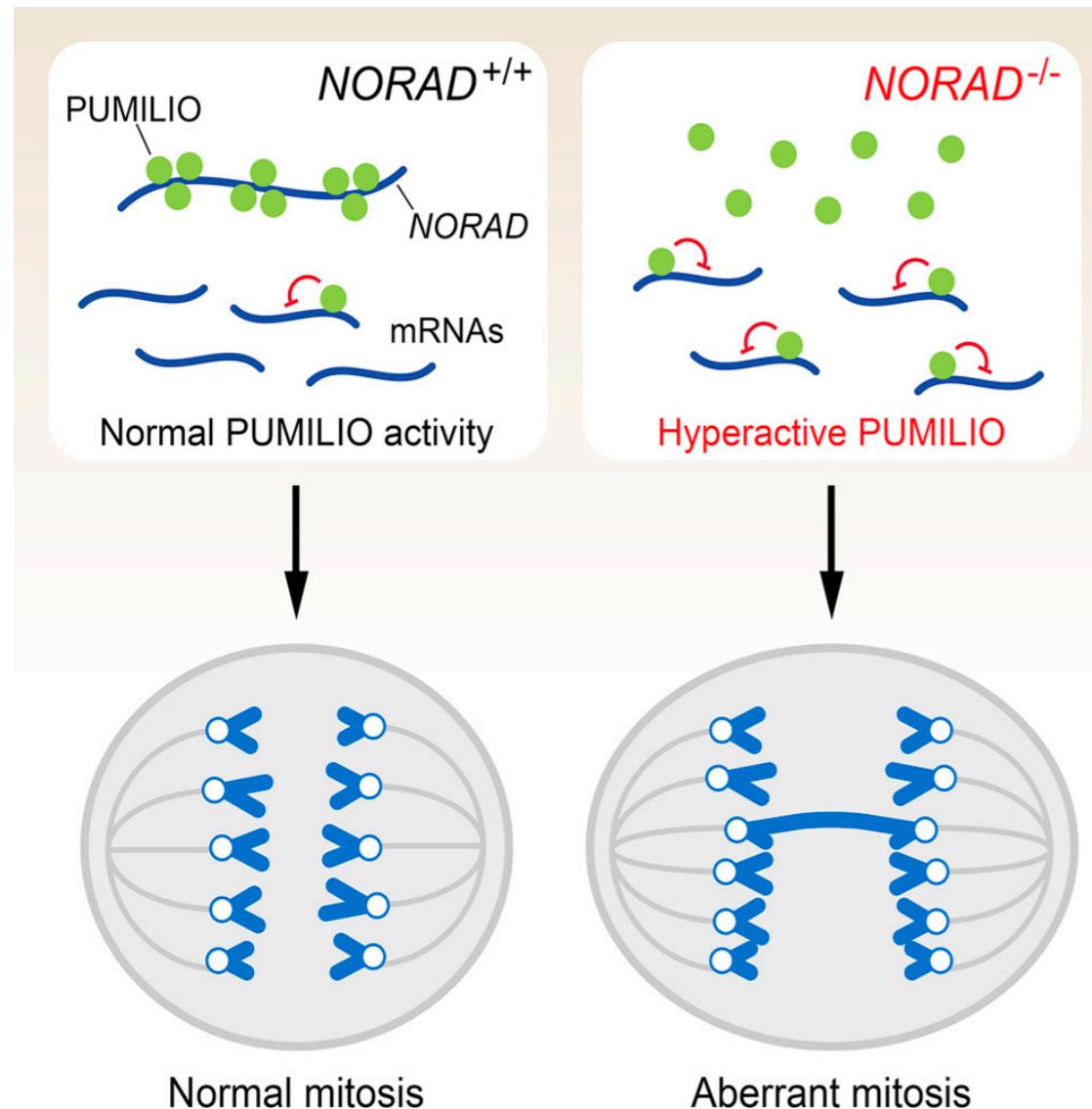
Differential Expression of PCGS between high and low LINC00657 GTEX ovary samples



Calculating Power for Survival Analysis

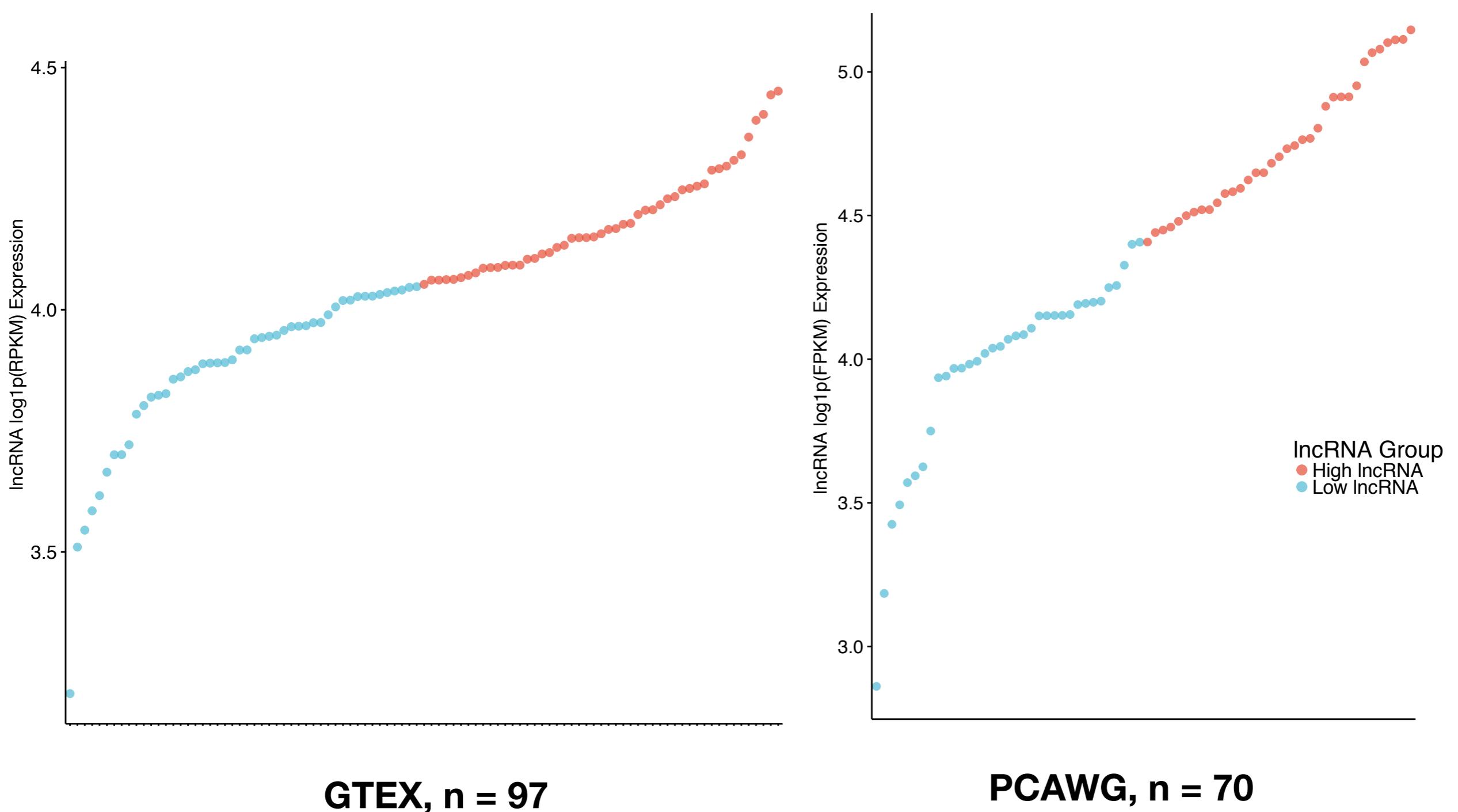
- With 70 patients split into two equal groups
 - “Experimental” and “Control”, n= 35 each
 - Postulated HR of 2
 - alpha = 0.05
 - Power = 73% for LINC00657 as predictor
 - Power = 78% for LINC00665 as predictor

Annotated Function of LINC00657 (NORAD)



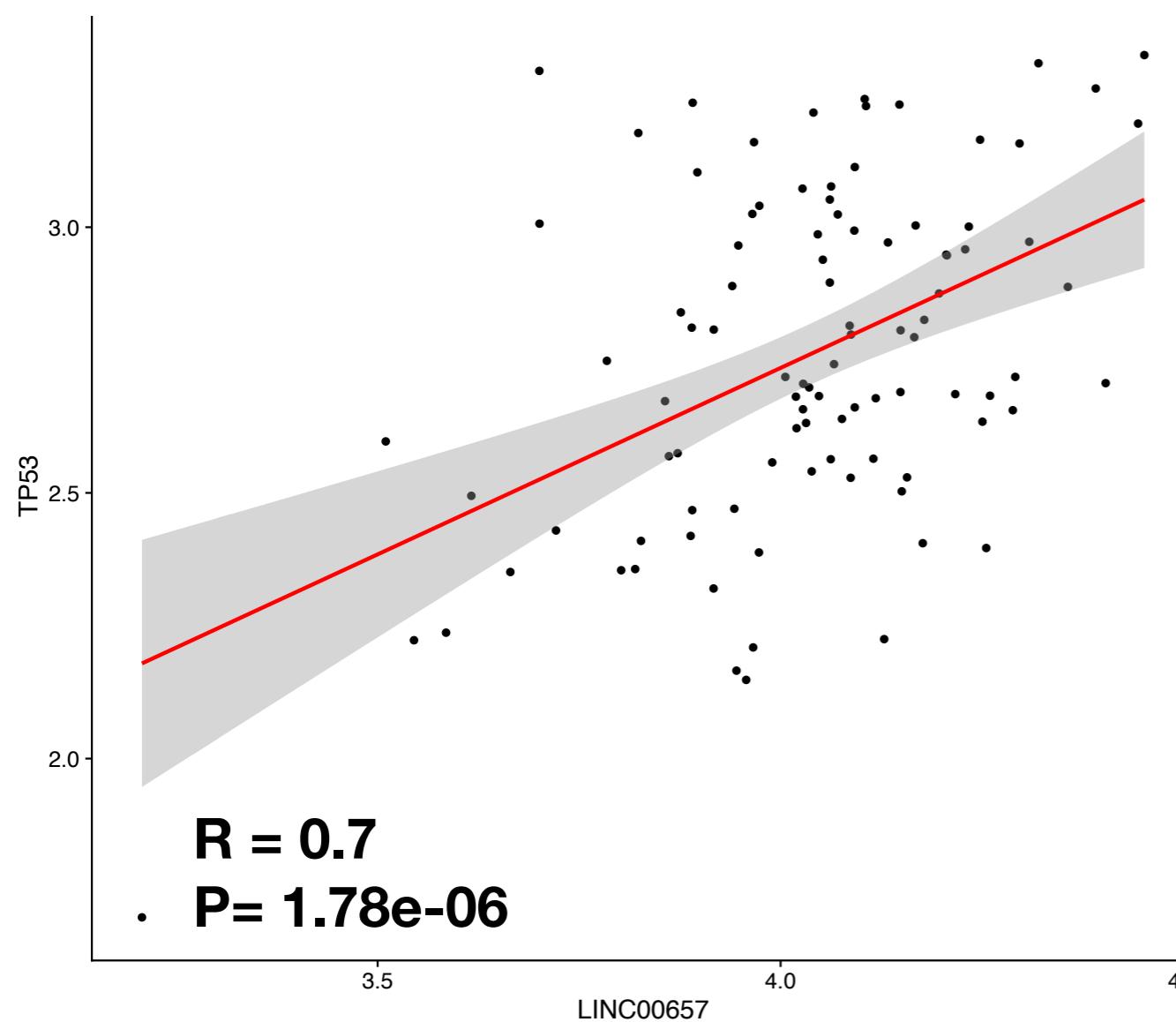
- DNA damage induced NORAD expression in a TP53-dependent manner in HCT116 human colon cancer cells
- Conditional knockout or knockdown of NORAD expression caused multiple mitotic errors, including anaphase bridges, mitotic slippage, and significant aneuploidy

In-depth analysis of LINC00657 in normal and tumour ovarian samples

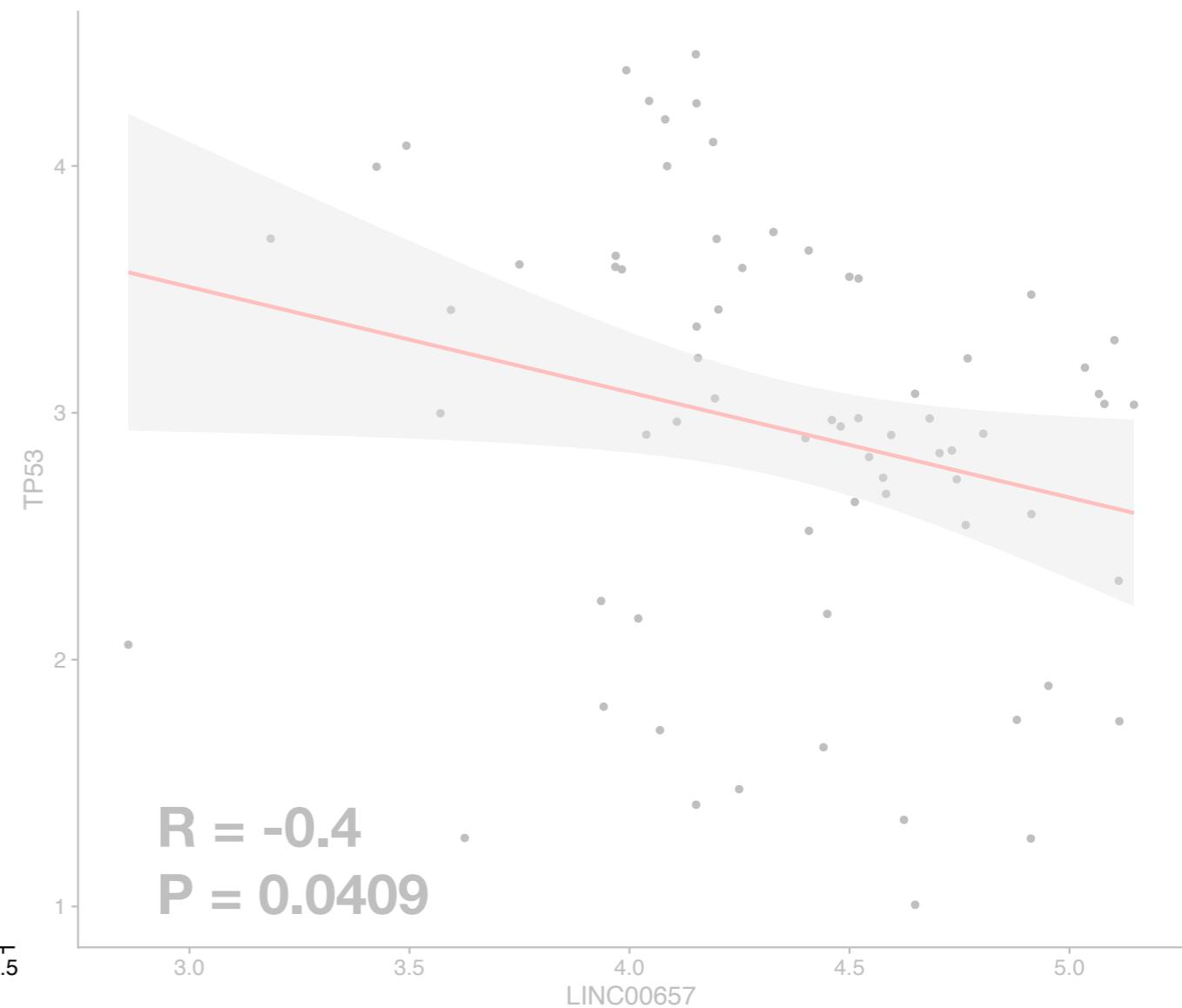


In-depth analysis of LINC00657 in normal and tumour ovarian samples

**LINC00657-TP53 Co-Expression
in 97 Normal Ovary Samples**

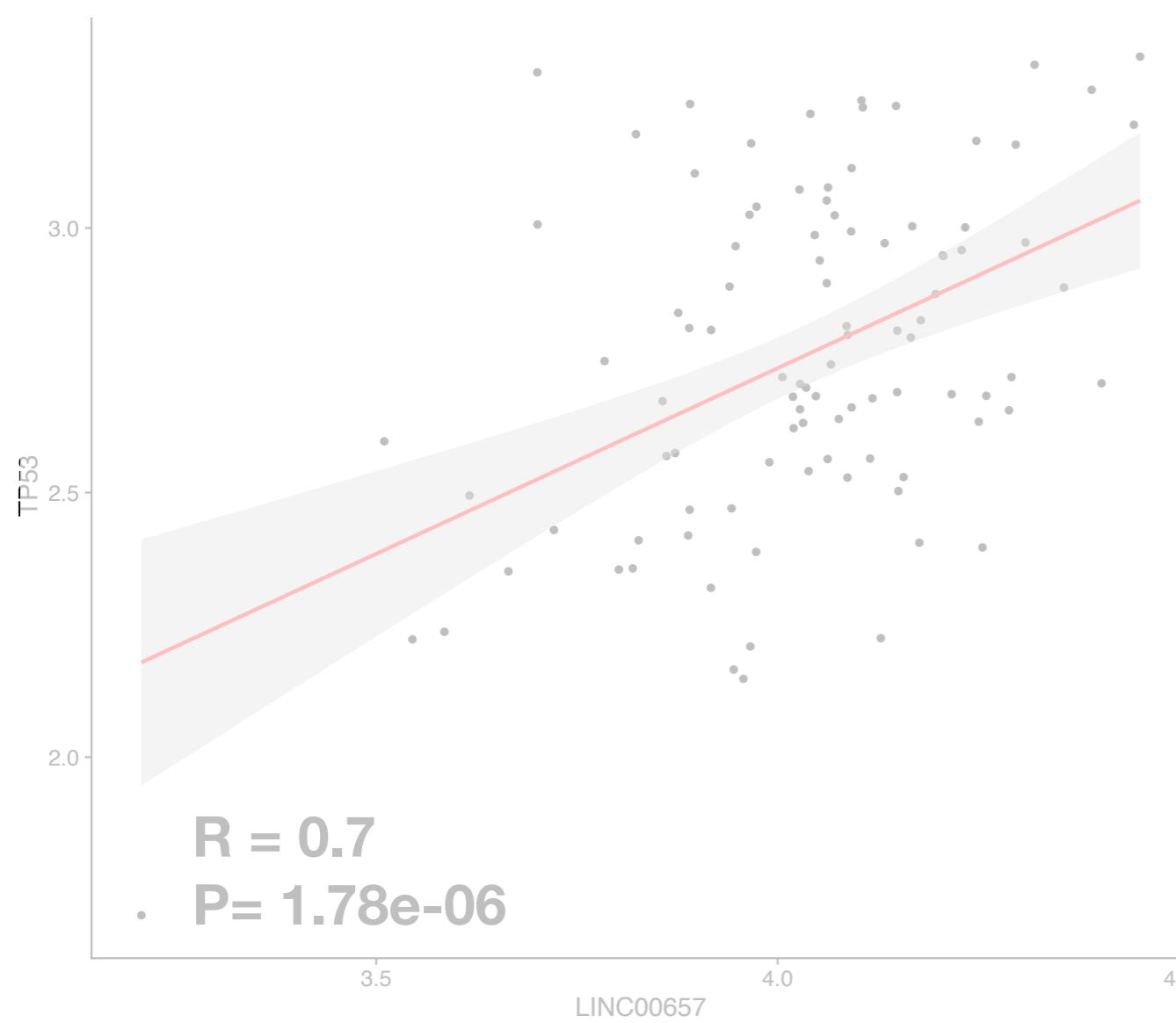


**LINC00657-TP53 Co-Expression
in 70 Tumour Ovary Samples**

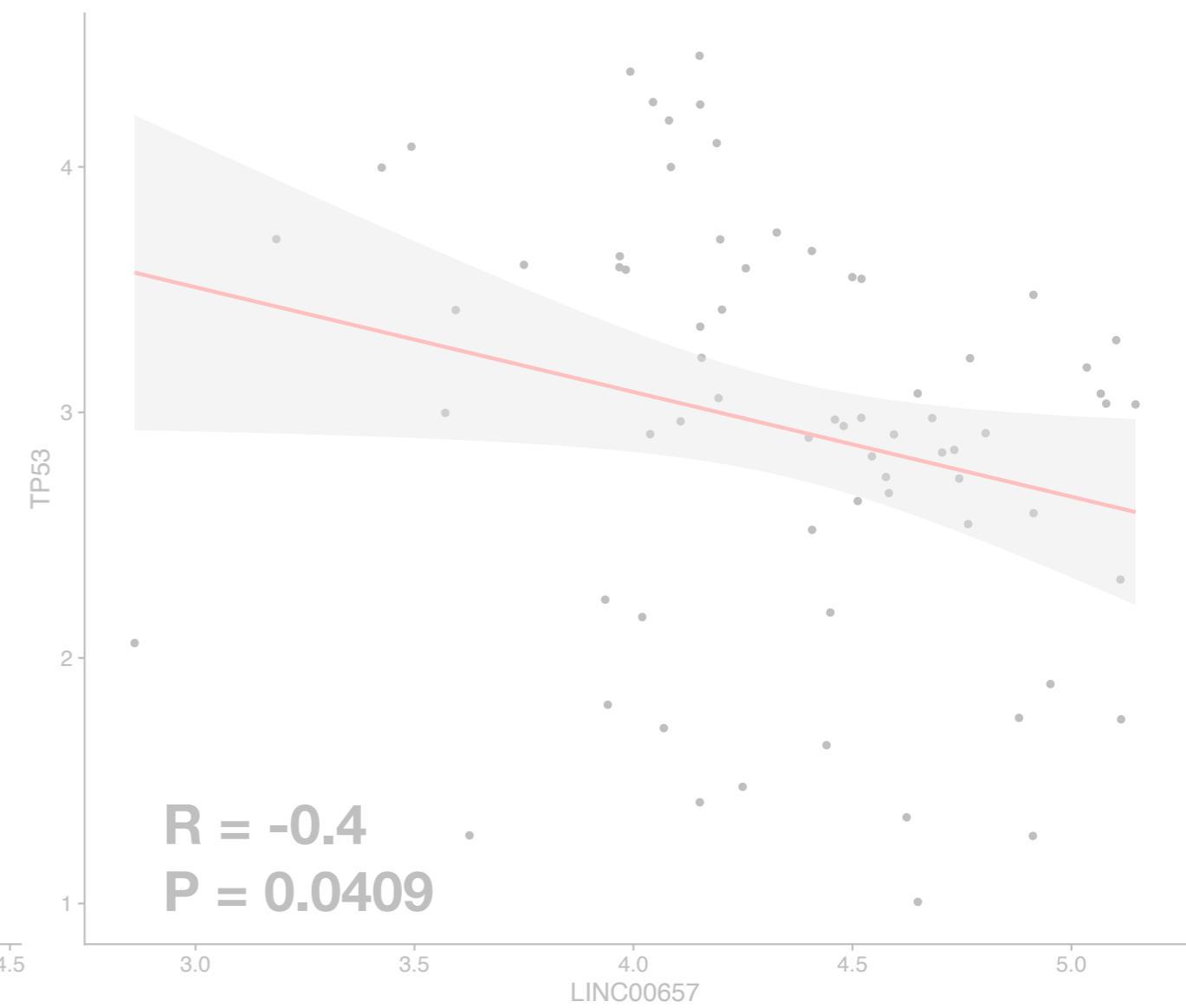


In-depth analysis of LINC00657 in normal and tumour ovarian samples

LINC00657-TP53 Co-Expression
in 97 Normal Ovary Samples

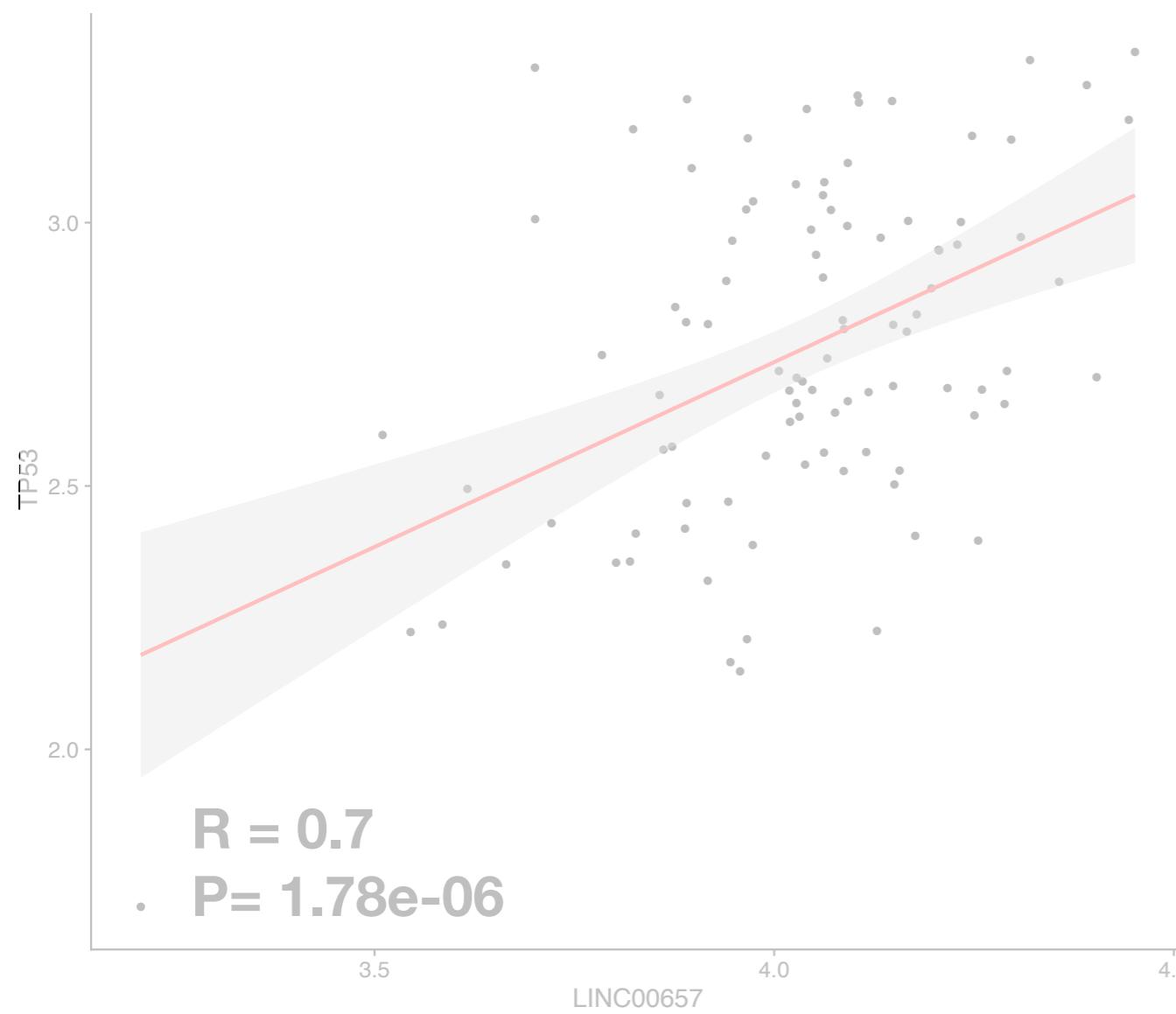


LINC00657-TP53 Co-Expression
in 70 Tumour Ovary Samples

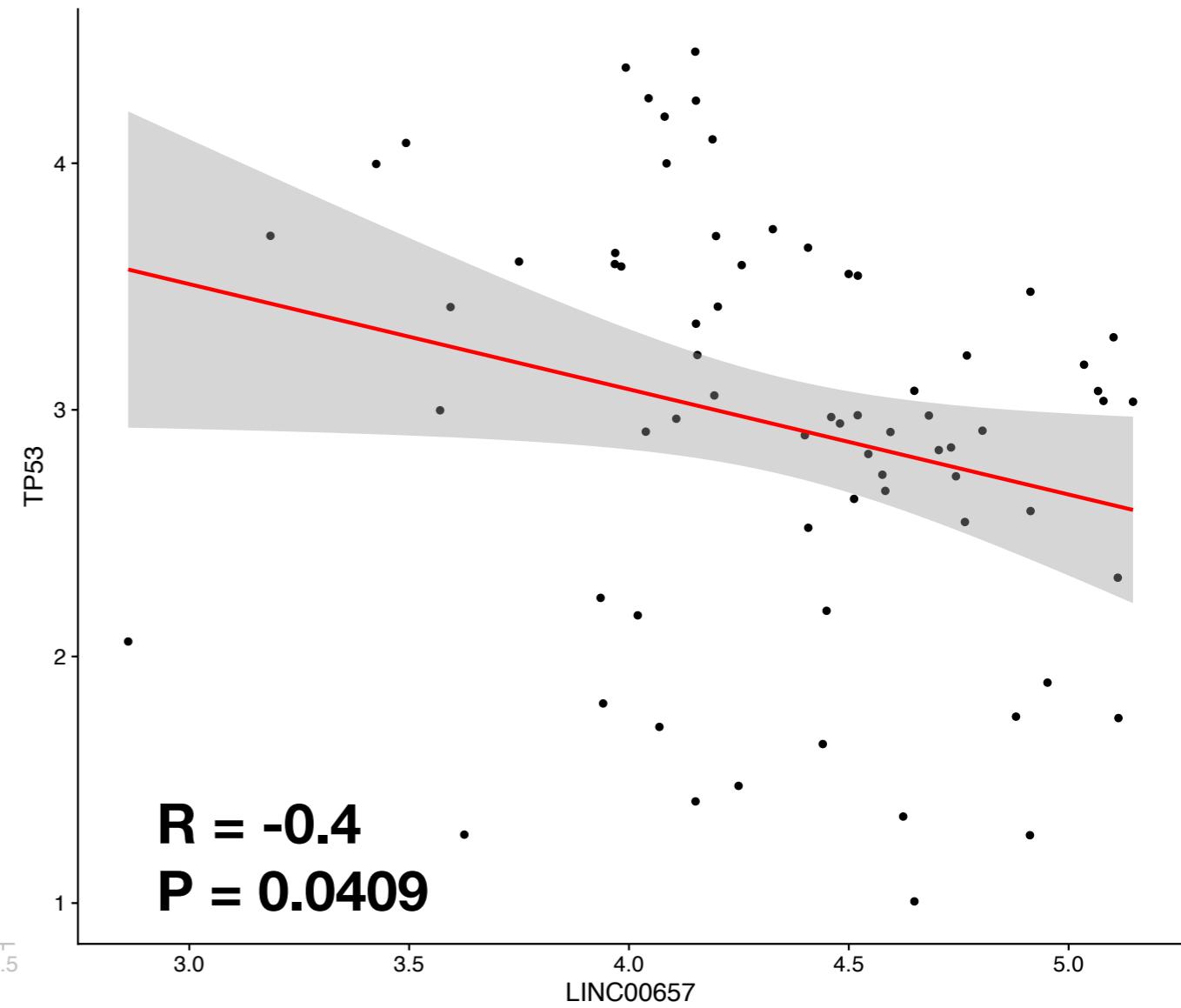


In-depth analysis of LINC00657 in normal and tumour ovarian samples

LINC00657-TP53 Co-Expression
in 97 Normal Ovary Samples

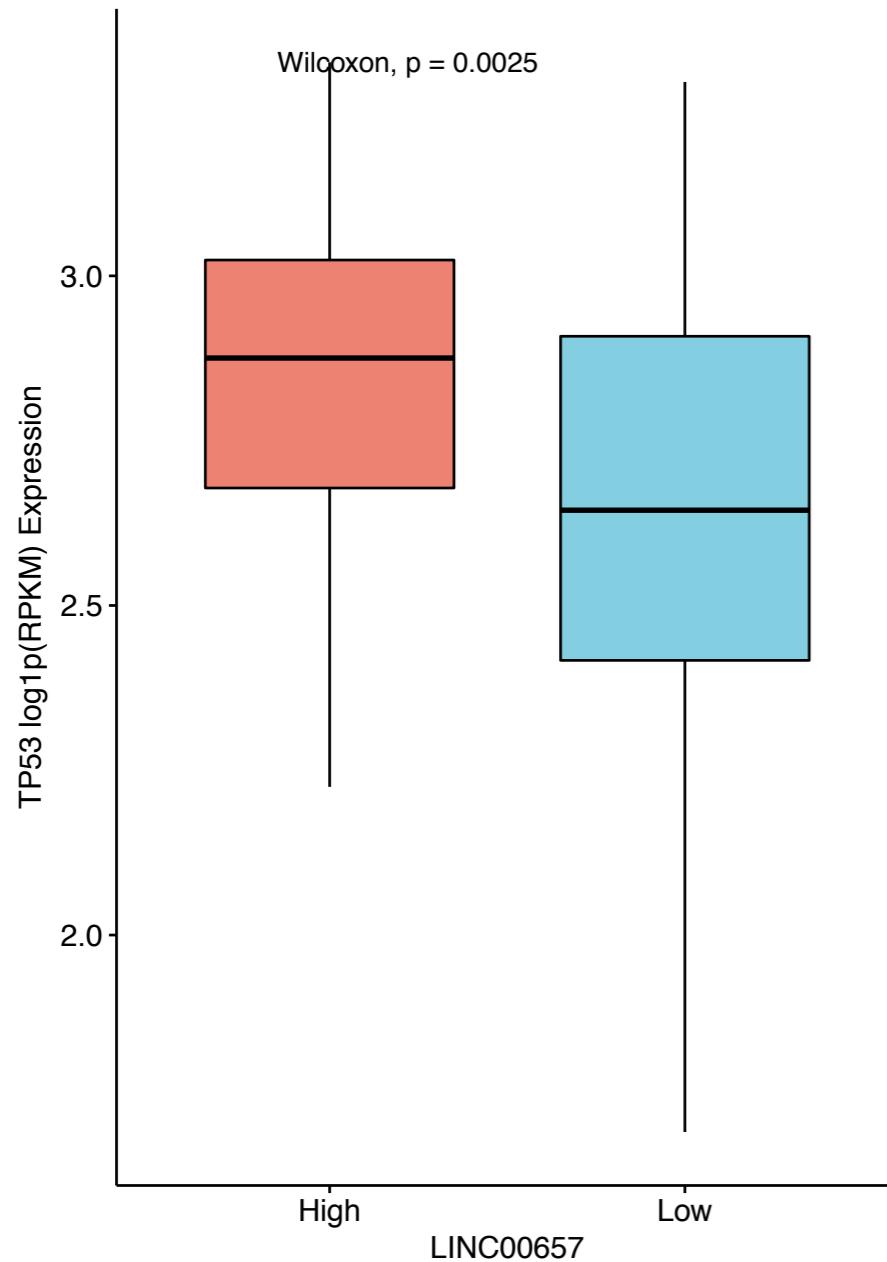


LINC00657-TP53 Co-Expression
in 70 Tumour Ovary Samples

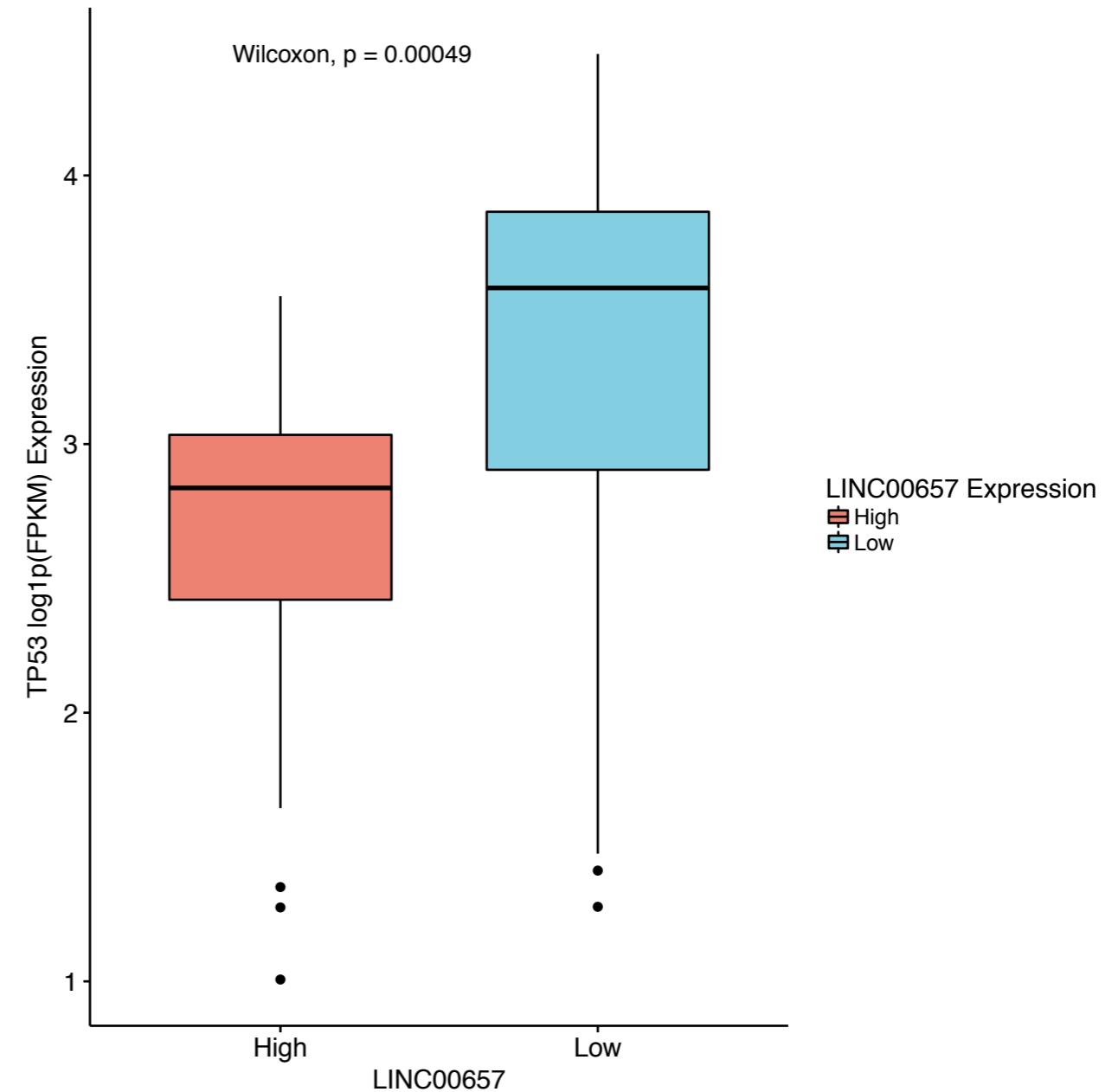


In-depth analysis of LINC00657 in normal and tumour ovarian samples

TP53 Expression Relative to LINC00657
in 97 Normal Ovary Samples

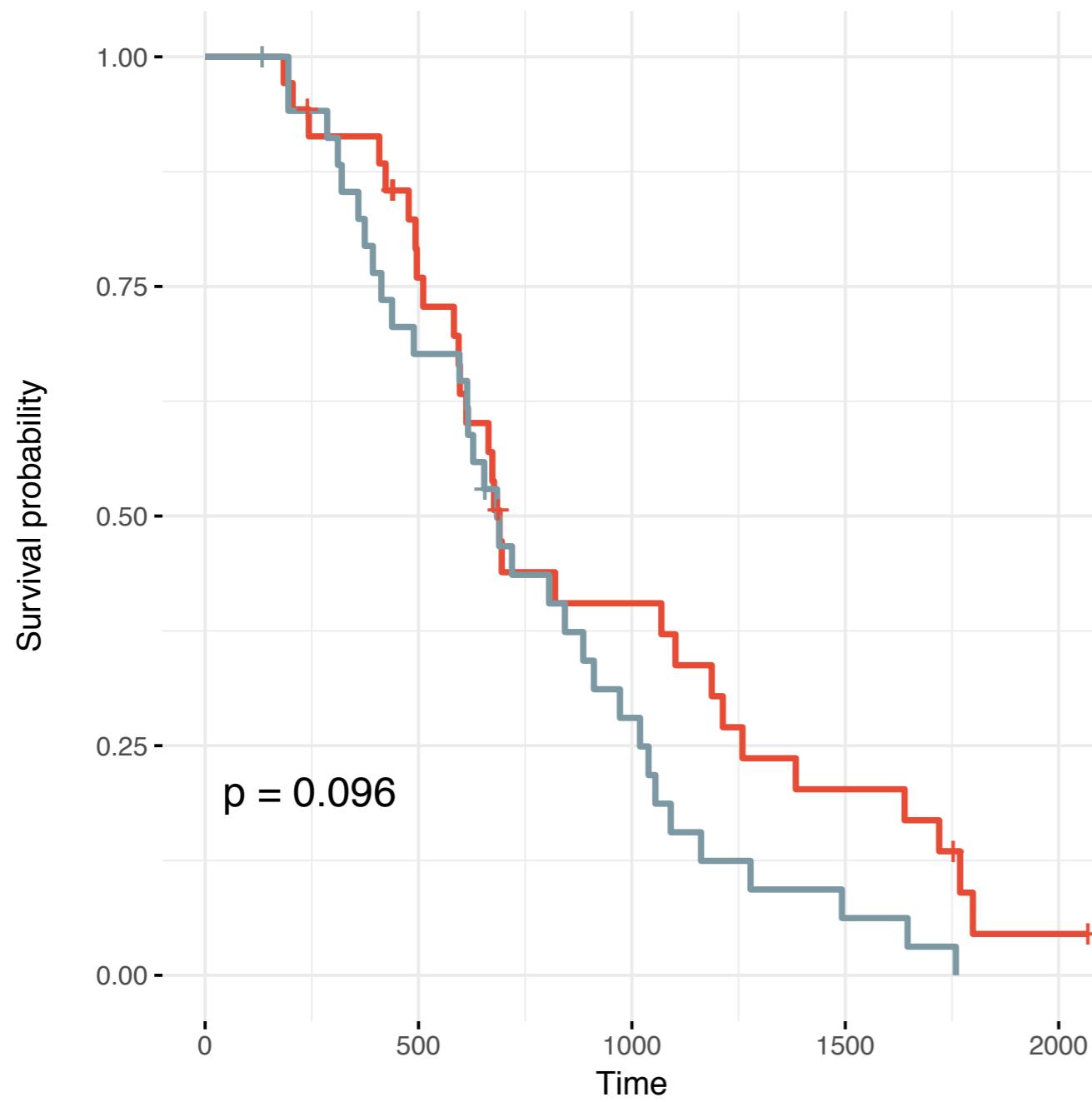


TP53 Expression Relative to LINC00657
in 70 Tumour Ovary Samples



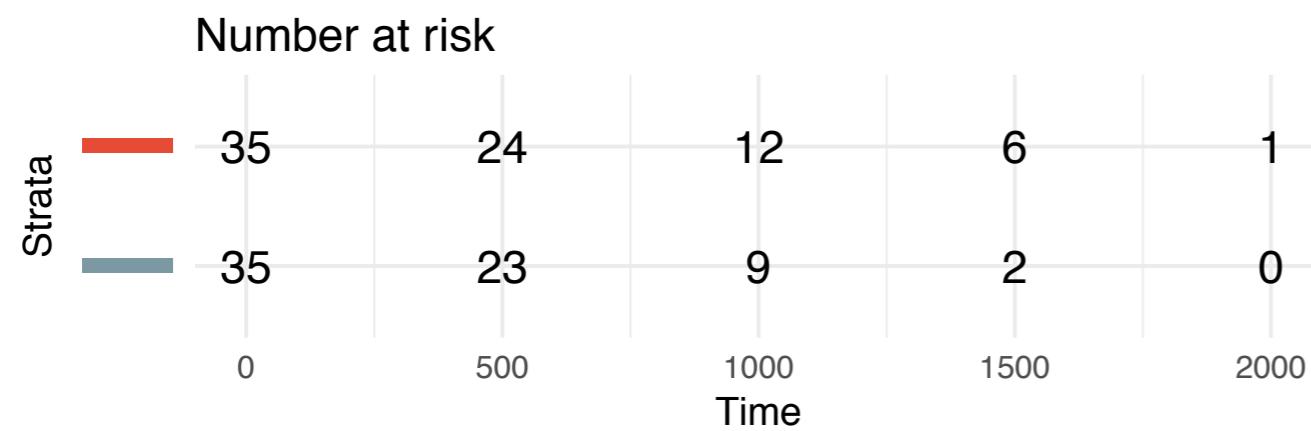
1. LINC00657 and TP53 are negatively correlated in cancer
2. **TP53 more abundant in those with low LINC00657 expression (low risk group)**
 - Thus —> TP53 high expressing patients should also have better prognosis than low expressing patients

Survival analysis using TP53 expression as predictor



Strata
TP53=high
TP53=low

Violates proportionality assumption



In-depth analysis of LINC00657 in normal and tumour ovarian samples

- NORAD may be acting independently of TP53
 - Further analysis through co-expression
 - Analysis of variants and methylation patterns in NORAD promoter
- May be inducible by TP53 in some cancers or normal tissues but have an ovarian cancer specific mechanism of dysregulation caused by:
 - Copy number aberrations
 - Mutations in promoter or enhancer regions of NORAD