Supplementary Material for PRER: A Patient Representation with Pairwise Relative Expressions on Biological Networks

Halil İbrahim Kuru Mustafa Buyukozkan Oznur Tastan

1 Supplementary Tables

1.1 Clinical and Molecular Patient Data

Number of Patients		
	censored	deceased
KIRC	297	156
OV	194	213
BRCA	628	99
LUAD	147	72
LUSC	106	78
COAD	260	60
HNSC	98	114
BLCA	70	52
UCEC	347	50
GBM	73	139

Table 1: Number of censored and deceased patients for each cancer type.

1.2 Prediction Performance

Random Survival Forest C-indices						
	C-index					p-value
	1^{st}	2^{nd}	median	3^{rd}	4^{th}	
KIRC	0.57	0.66	0.69	0.72	0.80	$\mathbf{2.16x10^{-11}}$
KIRC-prer	0.58	0.67	**0.70**	0.73	0.83	
OV	0.40	0.52	0.55	0.57	0.66	$\mathbf{3.94x10^{-8}}$
OV-prer	0.46	0.54	**0.57**	0.60	0.68	
BRCA	0.38	0.55	0.60	0.64	0.82	$\mathbf{6.15x10^{-16}}$
BRCA-prer	0.56	0.62	**0.66**	0.71	0.83	
LUSC	0.36	0.51	0.57	0.64	0.73	$\mathbf{8.64x10^{-3}}$
LUSC-prer	0.41	0.56	**0.59**	0.65	0.76	
COAD	0.22	0.43	0.49	0.53	0.70	**0.02**
COAD-prer	0.25	0.44	**0.50**	0.55	0.72	
HNSC	0.37	0.48	0.52	0.55	0.70	$\mathbf{3.76x10^{-4}}$
HNSC-prer	0.39	0.49	**0.54**	0.59	0.73	
UCEC	0.27	0.47	0.53	0.59	0.80	$\mathbf{9.66x10^{-18}}$
UCEC-prer	0.39	0.62	**0.66**	0.72	0.86	
GBM	0.35	0.51	0.56	0.61	0.70	**0.02**
GBM-prer	0.41	0.53	**0.57**	0.62	0.71	
BLCA	0.31	0.47	0.55	0.61	0.83	0.09
BLCA-prer	0.29	0.50	**0.55**	0.61	0.71	
LUAD	0.40	0.55	0.59	0.66	0.80	1
LUAD-prer	0.31	0.47	0.51	0.57	0.74	

Table 2: The comparison of RFSs models' performances that are trained with individual features and pairwise ranking representations (PRER) for different cancer types. The methods are compared at the mean with four different quartile values of the 100 repeated runs, where each run is conducted with a random split of train and test data.

2 Supplementary Figures

2.1 Important PRER Features

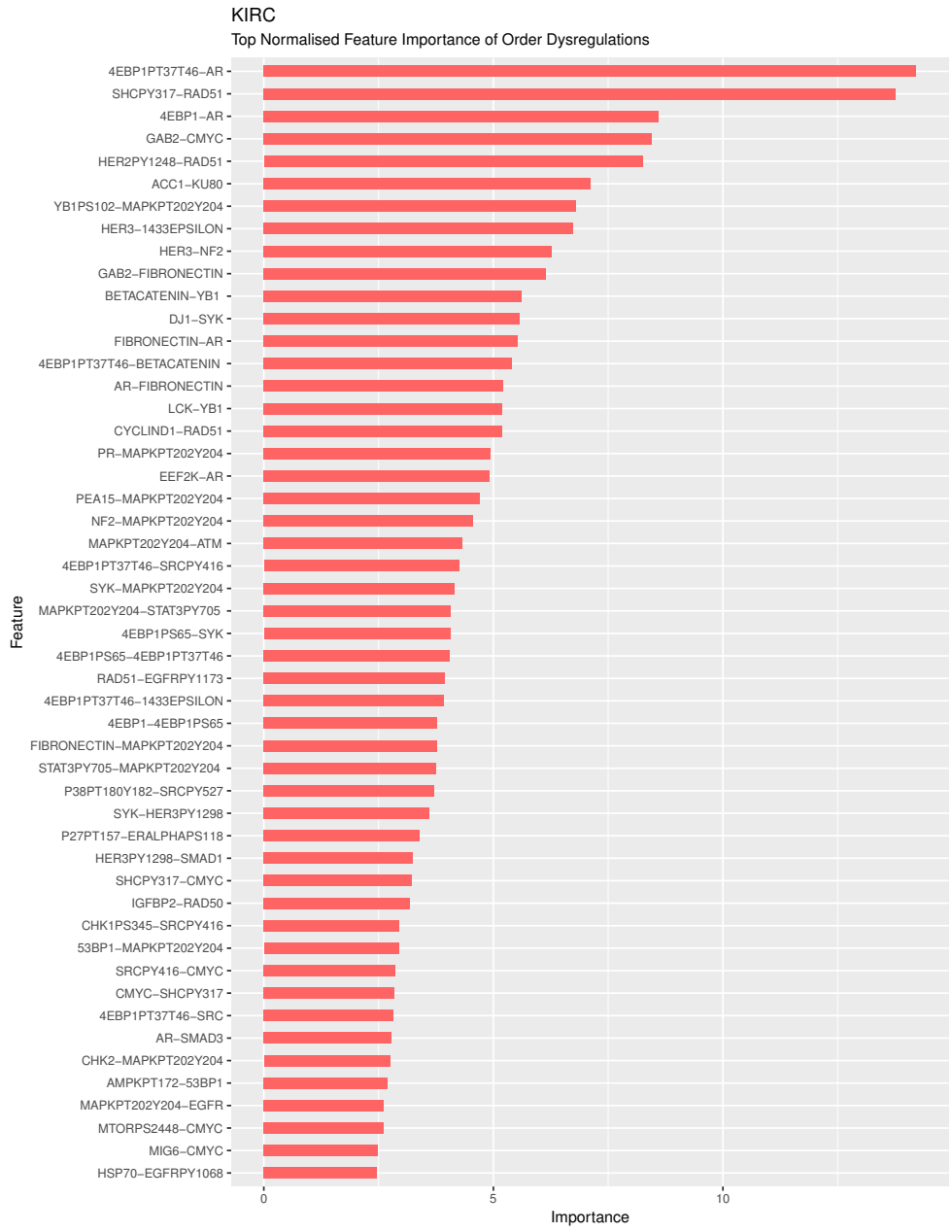


Figure 1: Variable importance of significant pairwise ranking embeddings for kidney renal clear cell carcinoma

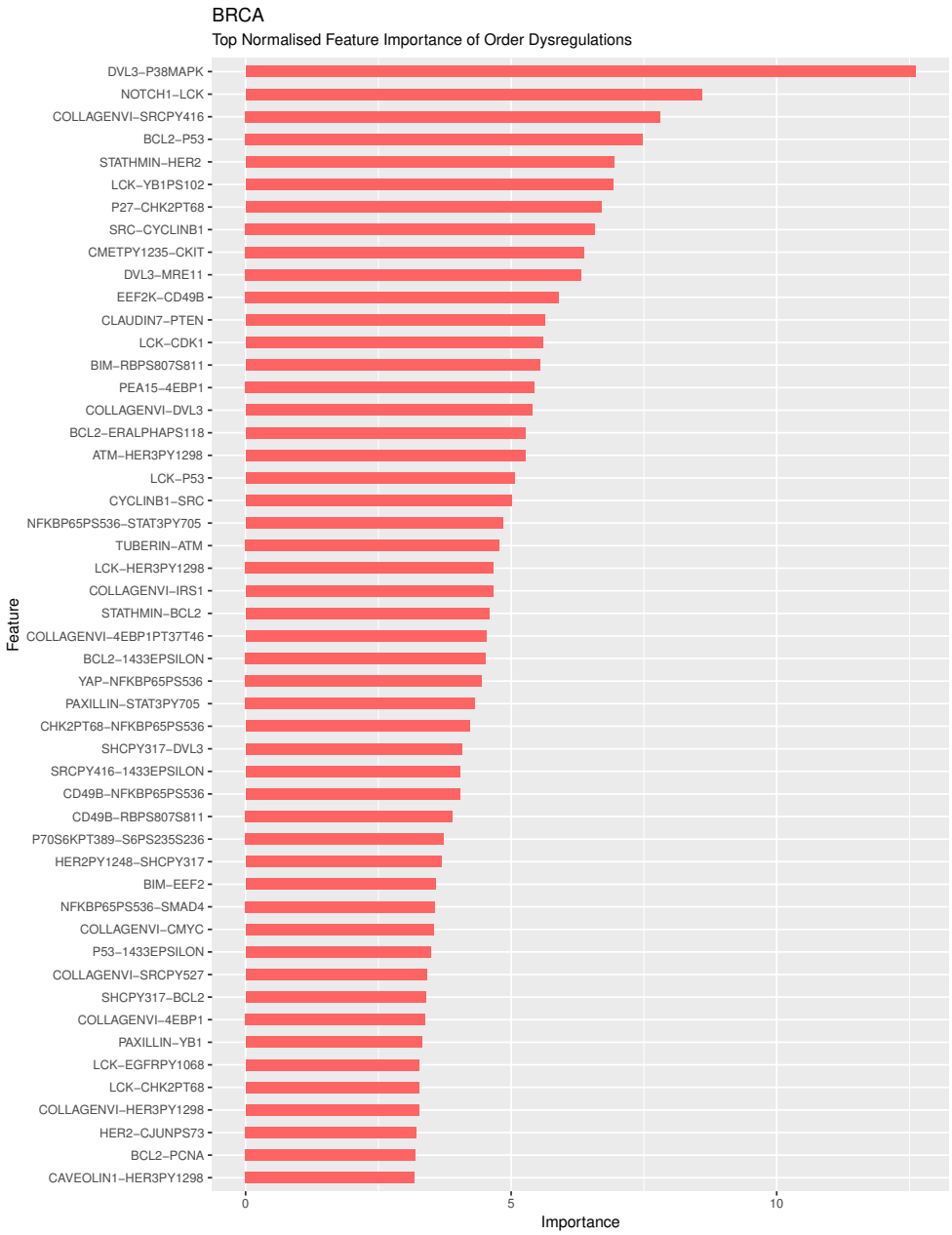


Figure 2: Variable importance of significant pairwise ranking embeddings for breast invasive carcinoma

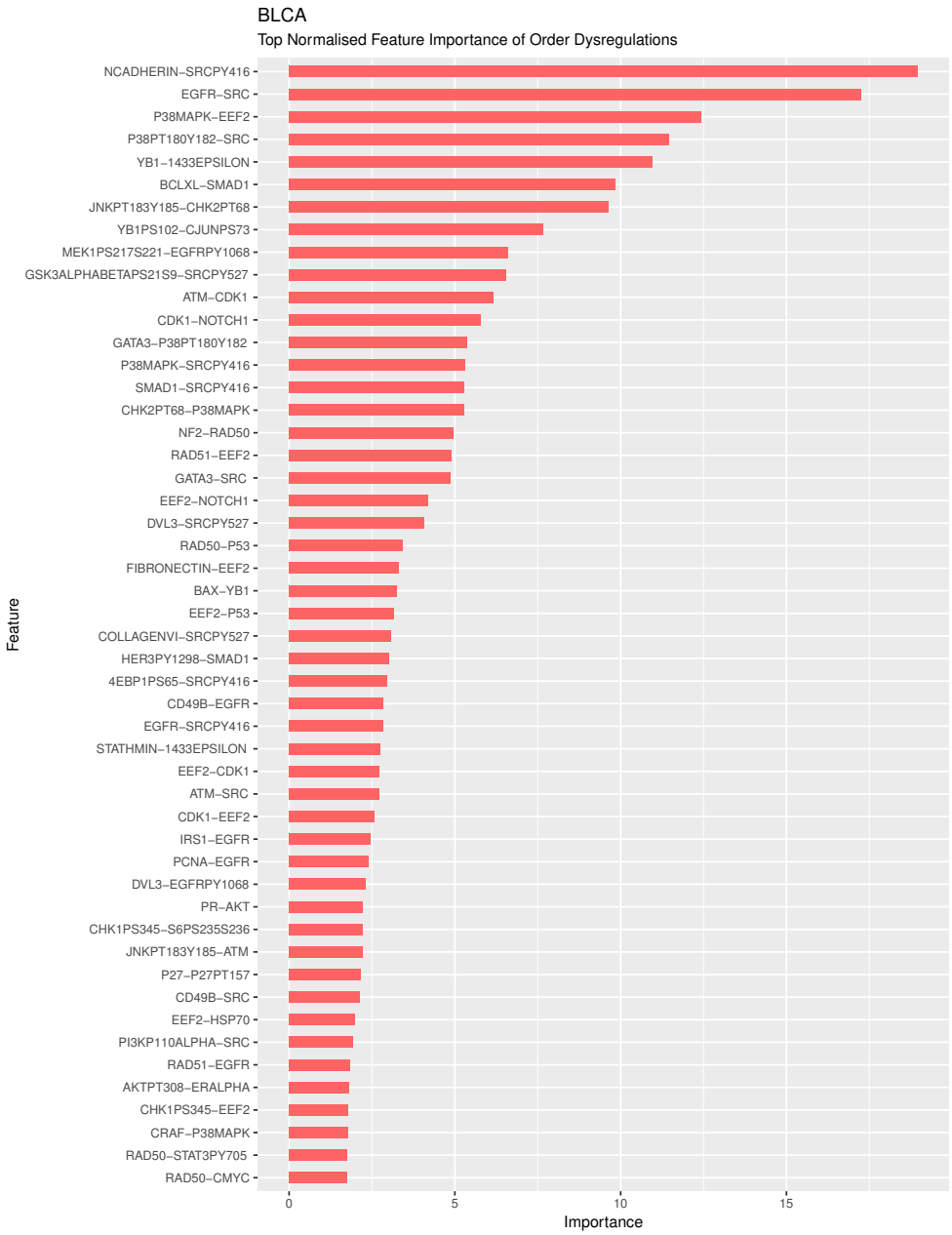


Figure 3: Variable importance of significant pairwise ranking embeddings for bladder urothelial carcinoma

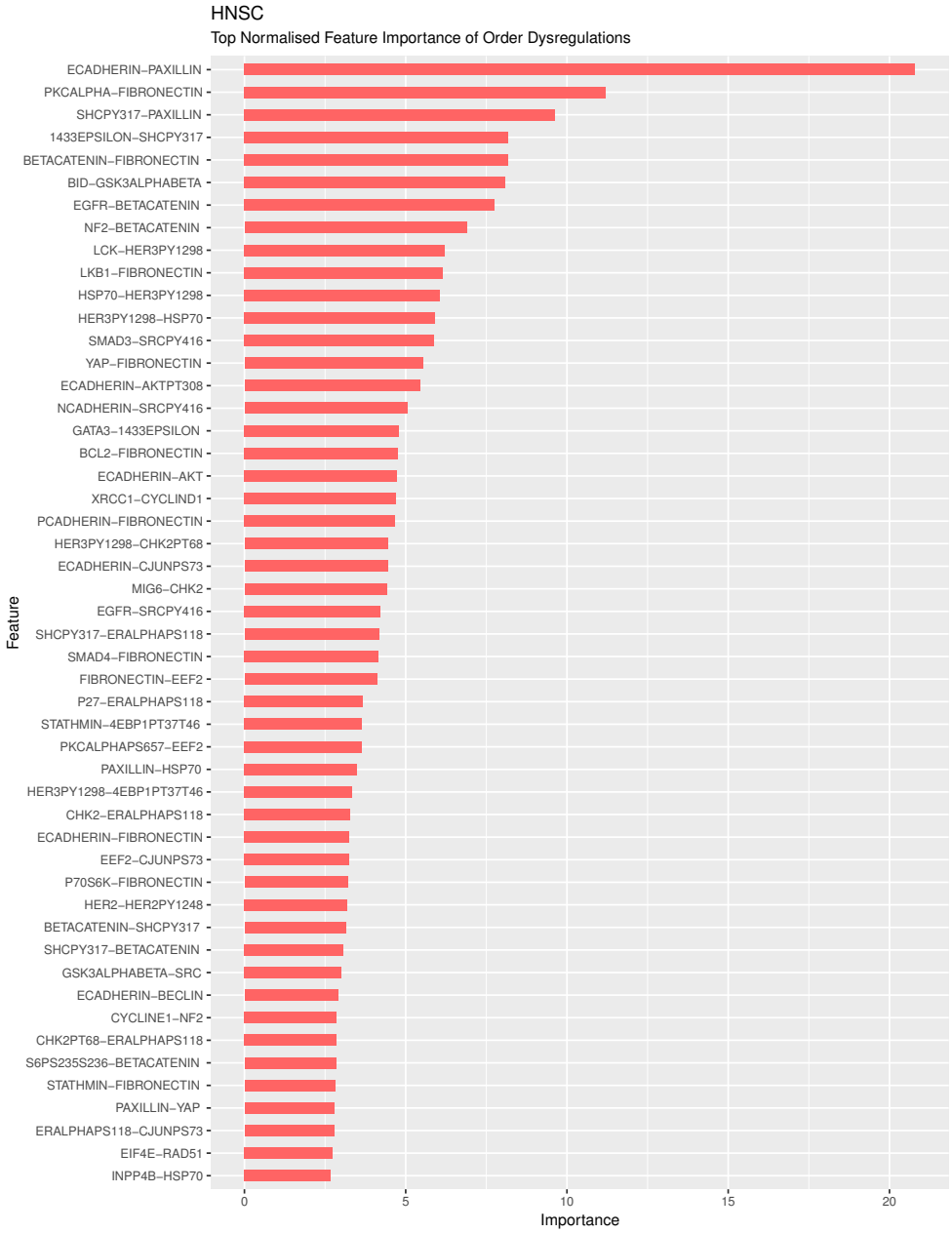


Figure 4: Variable importance of significant pairwise ranking embeddings for head and neck squamous cell carcinoma

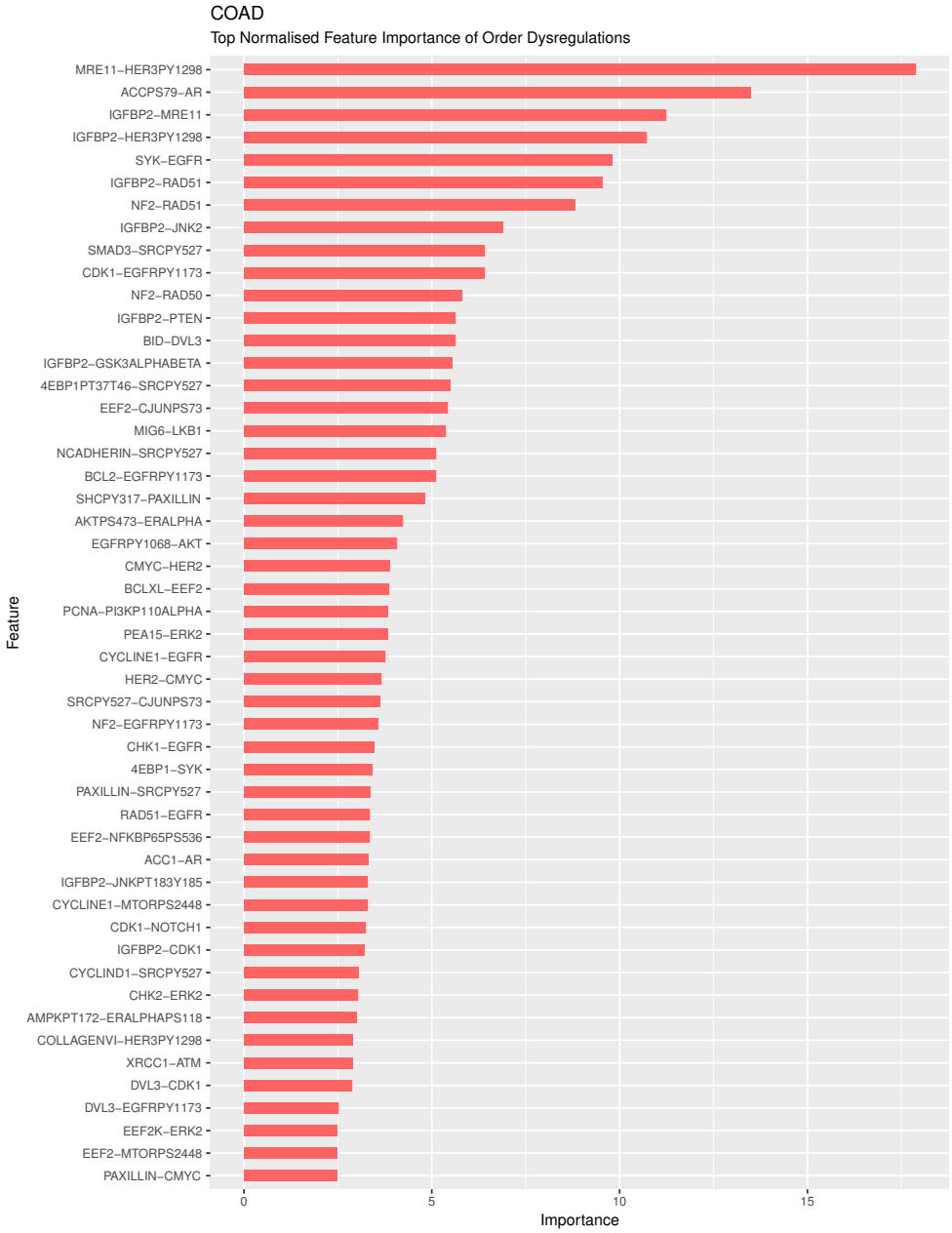


Figure 5: Variable importance of significant pairwise ranking embeddings for colon adenocarcinoma

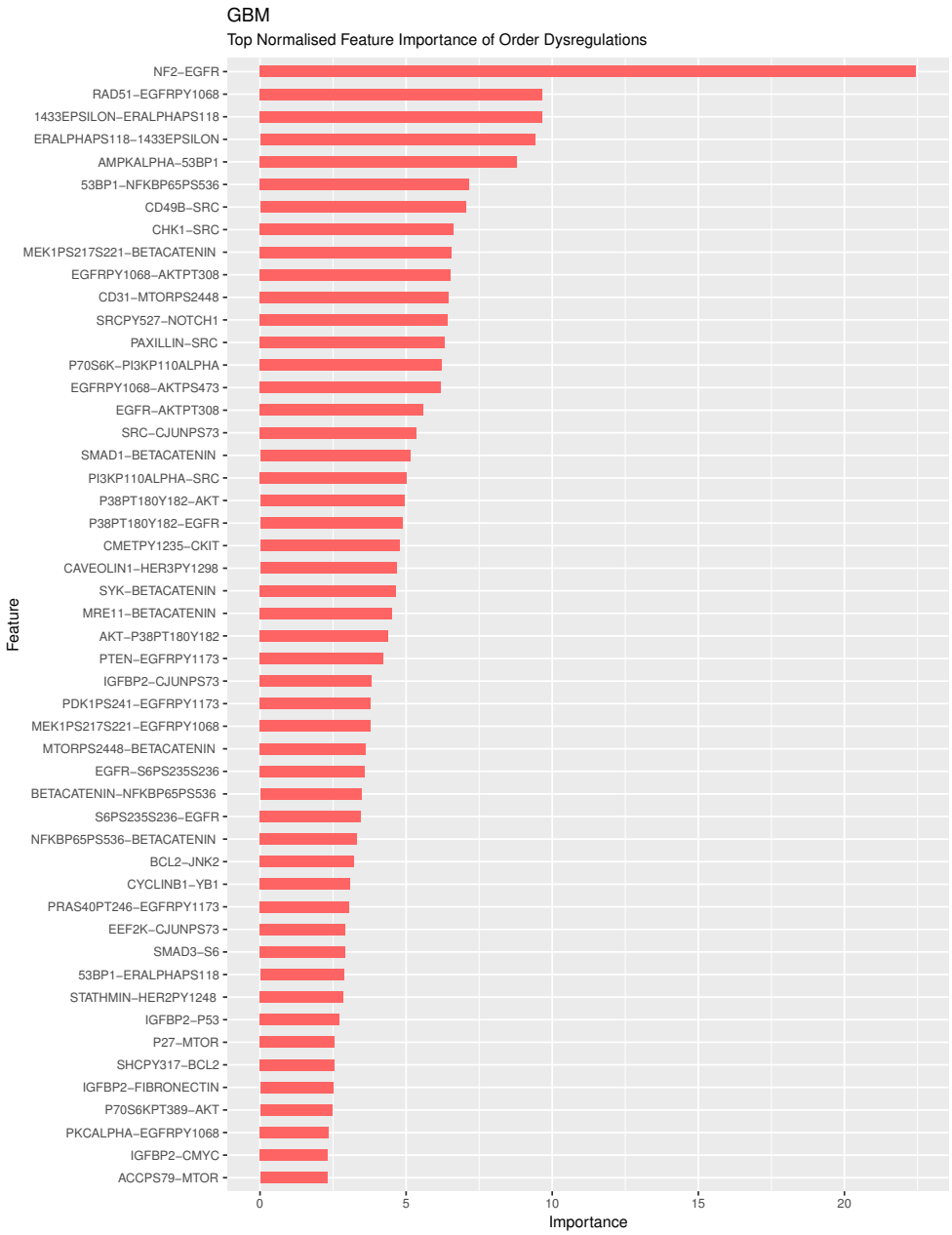


Figure 6: Variable importance of significant pairwise ranking embeddings for glioblastoma multiforme

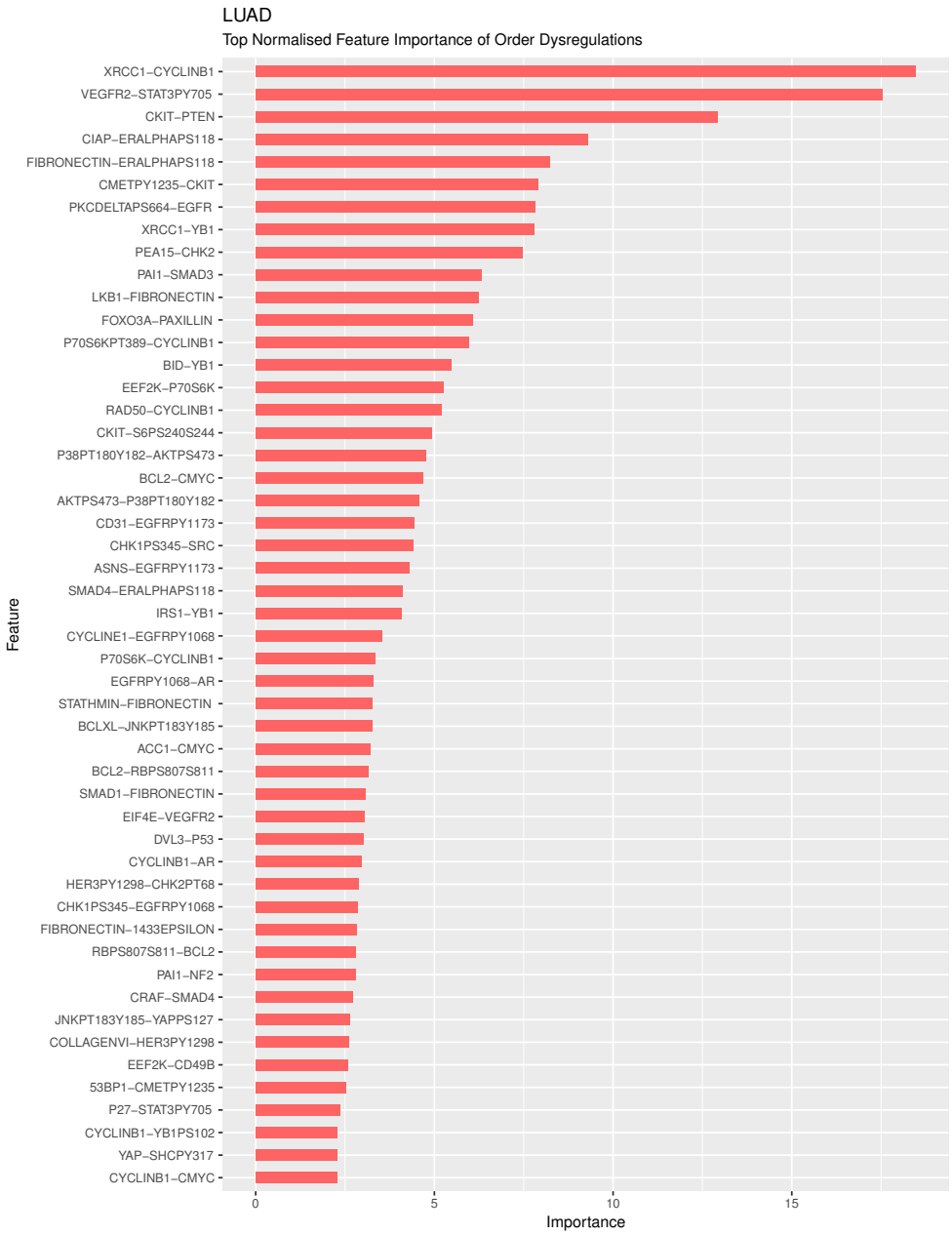


Figure 7: Variable importance of significant pairwise ranking embeddings for lung adenocarcinoma

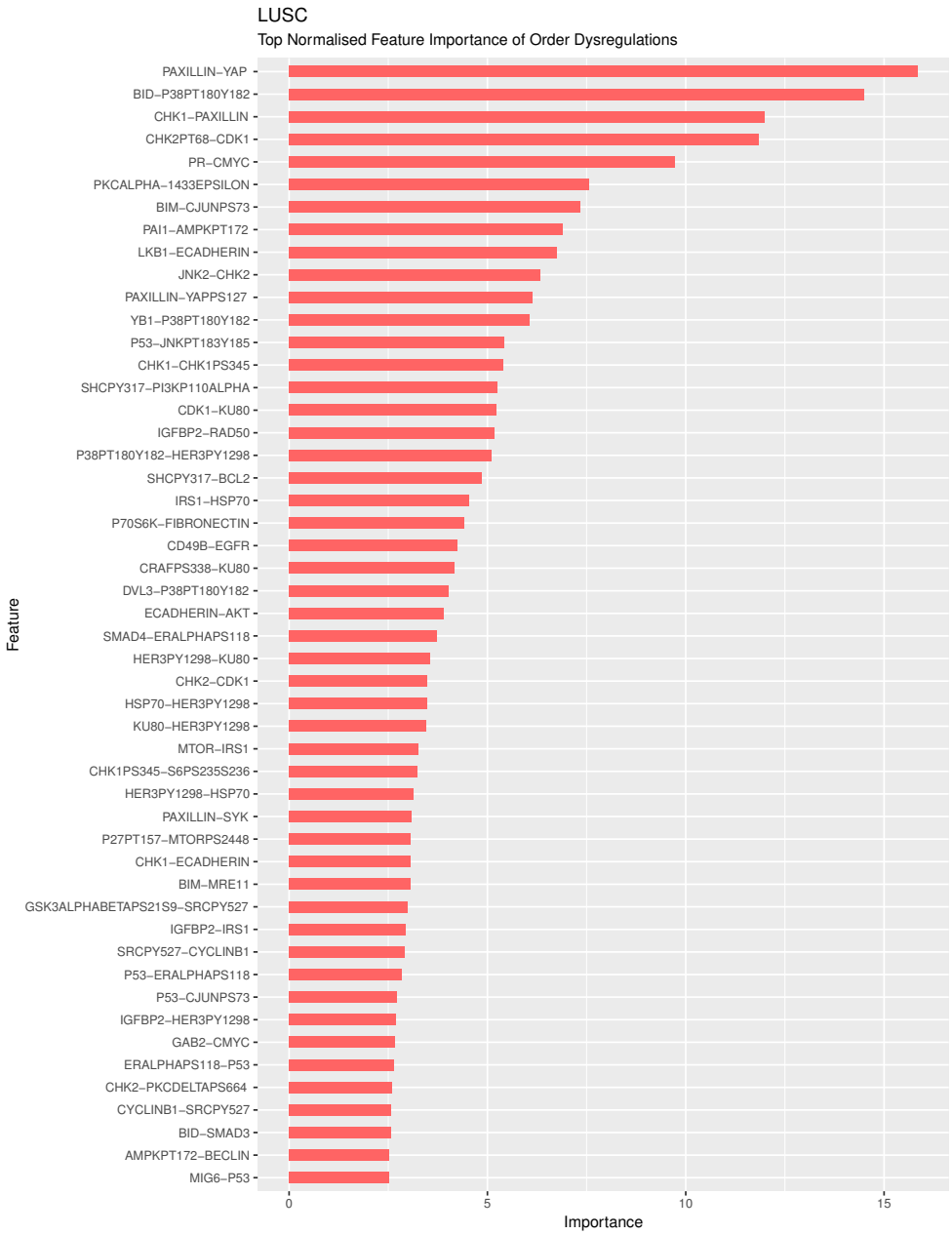


Figure 8: Variable importance of significant pairwise ranking embeddings for lung squamous cell carcinoma

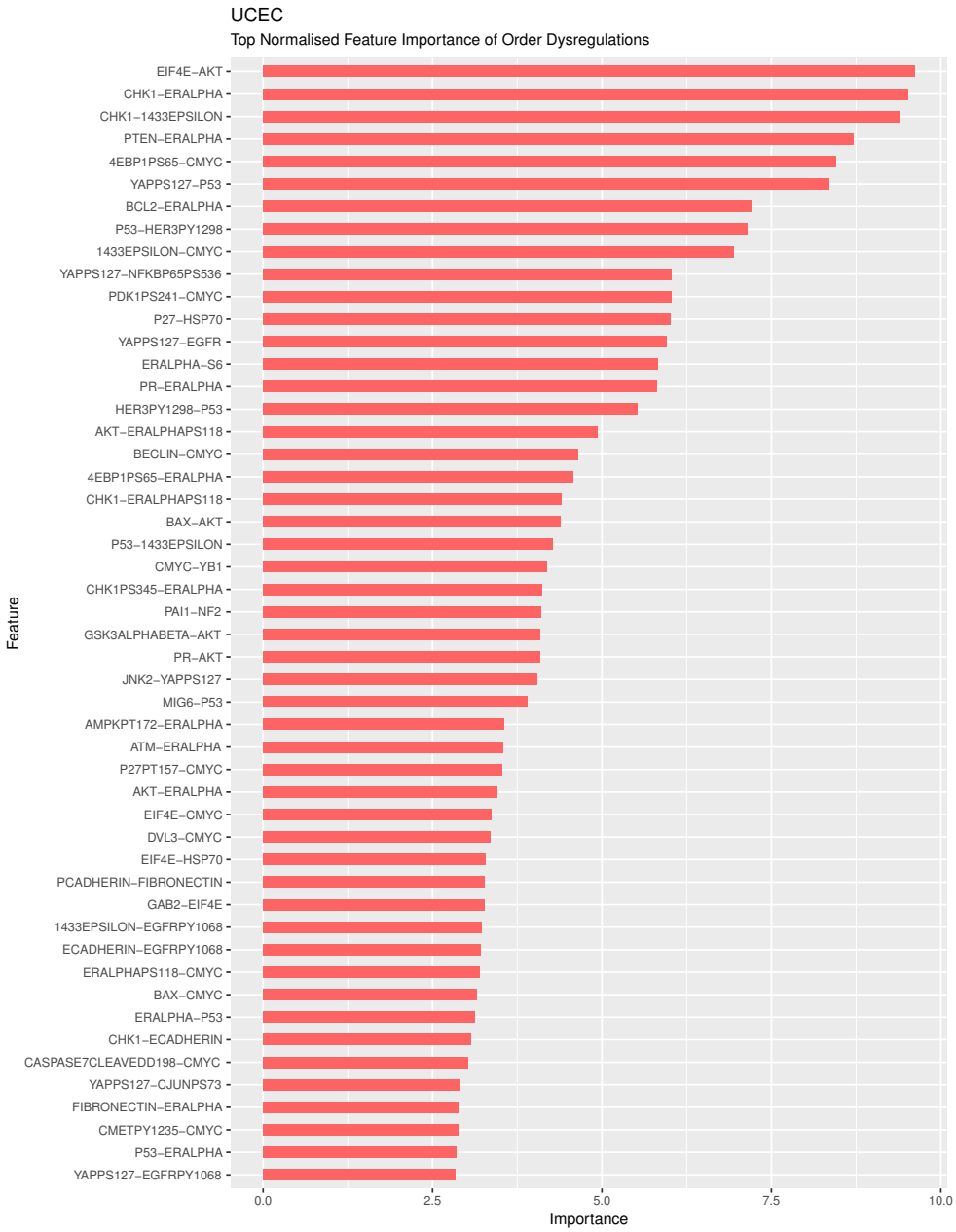


Figure 9: Variable importance of significant pairwise ranking embeddings for uterine corpus endometrial carcinoma

2.2 PRER Networks

KIRC

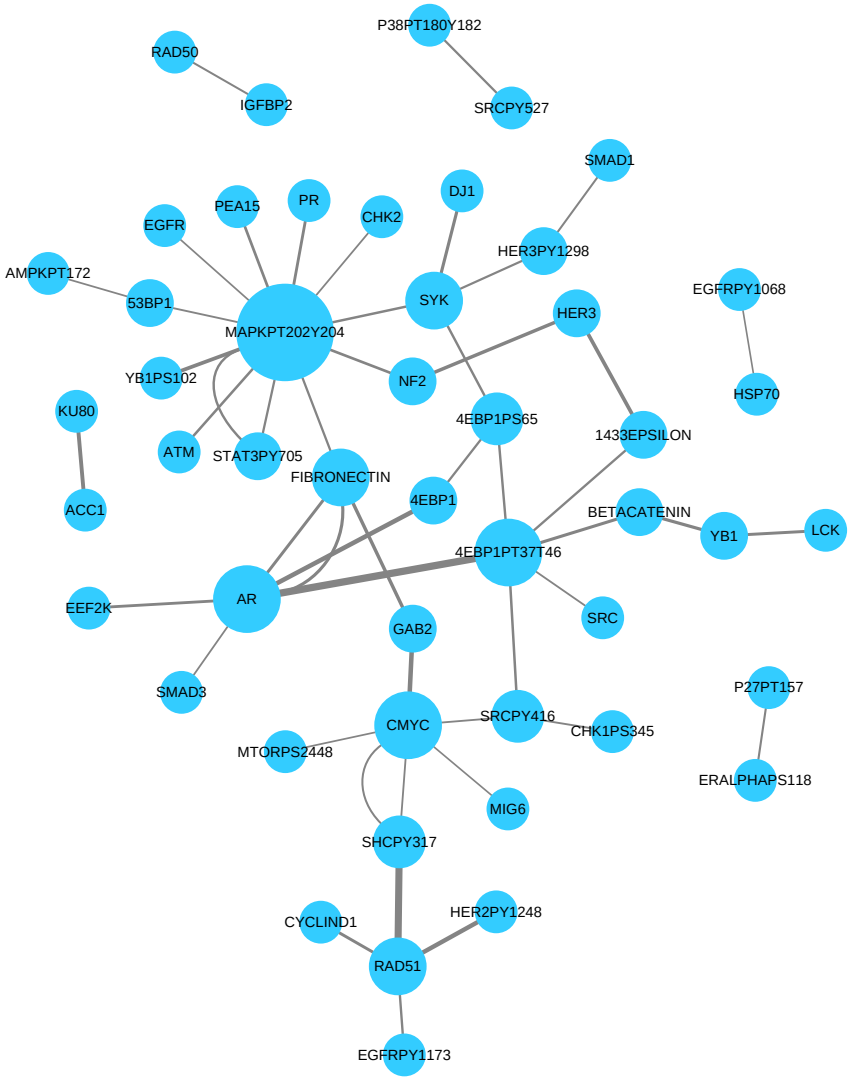


Figure 10: Nodes represent proteins that appear in the top 50 pairwise ranking embeddings for kidney renal clear cell carcinoma; edges represent that two proteins participate in a pairwise rank order feature together

BRCA

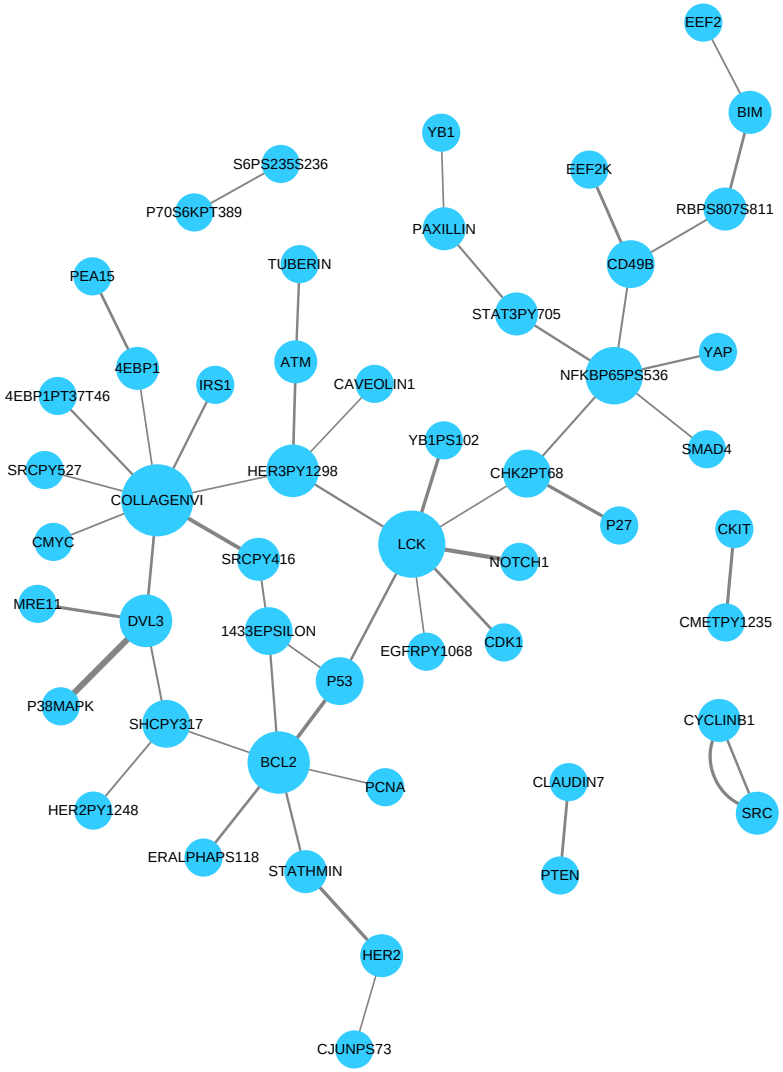


Figure 11: Nodes represent proteins that appear in the top 50 pairwise ranking embeddings for breast invasive carcinoma; edges represent that two proteins participate in a pairwise rank order feature together

BLCA

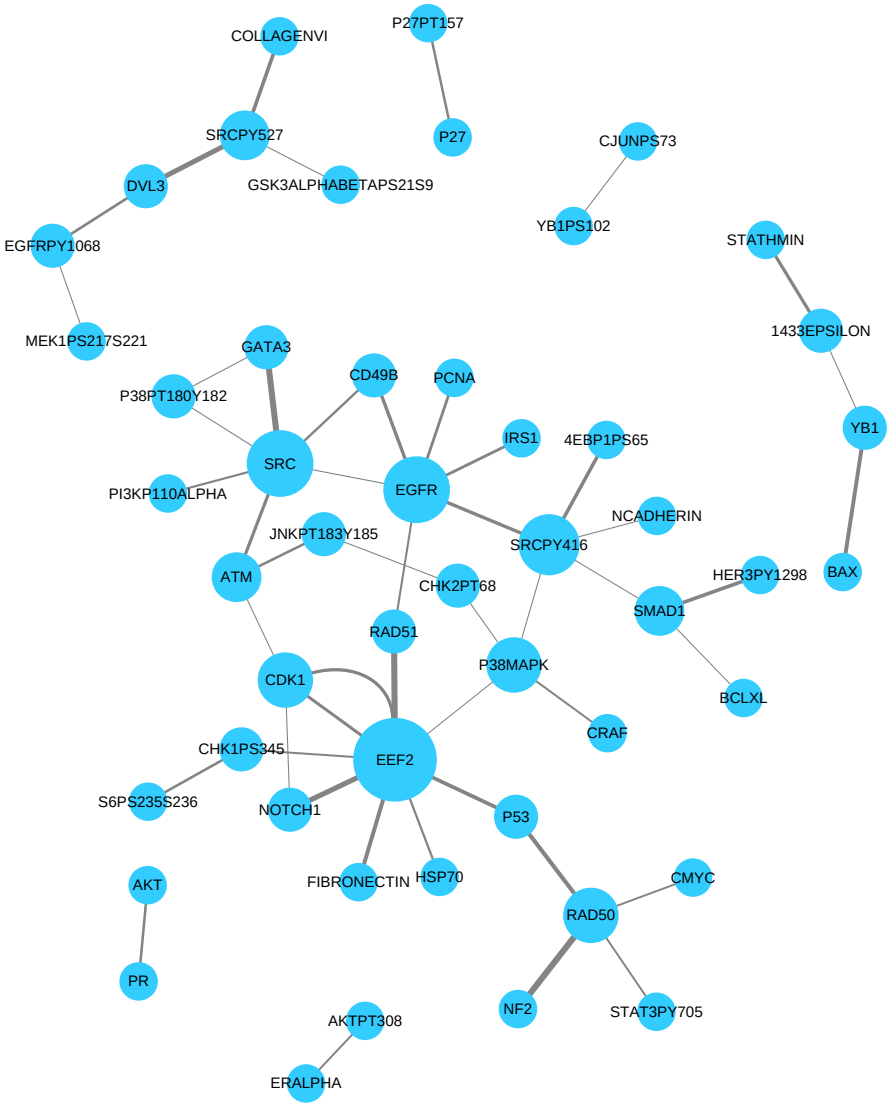


Figure 12: Nodes represent proteins that appear in the top 50 pairwise ranking embeddings for bladder urothelial carcinoma; edges represent that two proteins participate in a pairwise rank order feature together

HNSC

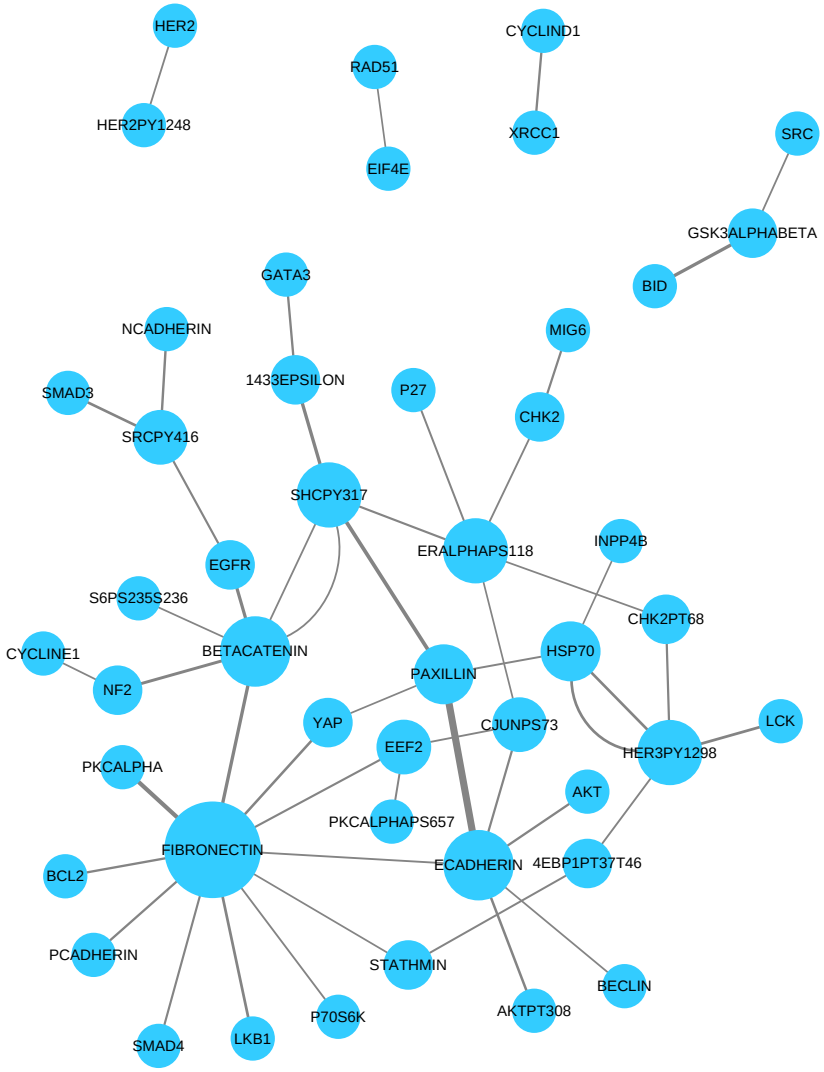


Figure 13: Nodes represent proteins that appear in the top 50 pairwise ranking embeddings for head and neck squamous cell carcinoma; edges represent that two proteins participate in a pairwise rank order feature together

COAD

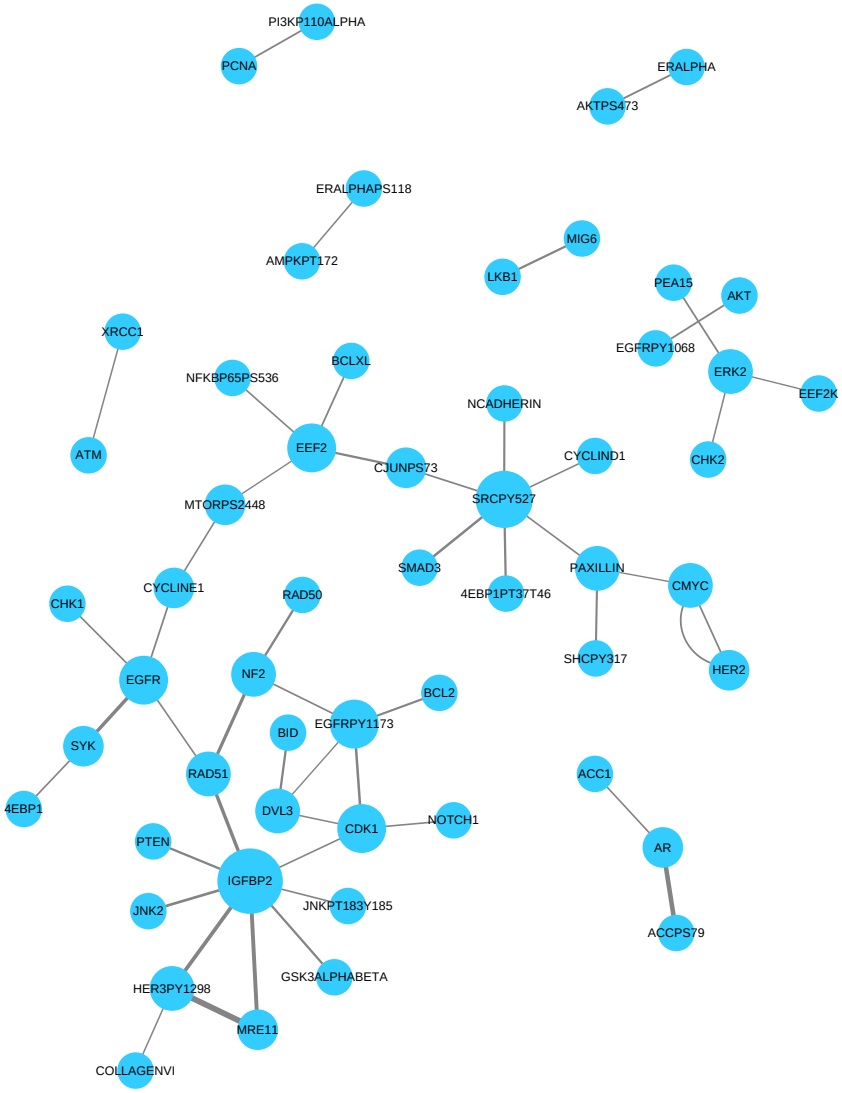


Figure 14: Nodes represent proteins that appear in the top 50 pairwise ranking embeddings for colon adenocarcinoma; edges represent that two proteins participate in a pairwise rank order feature together

GBM

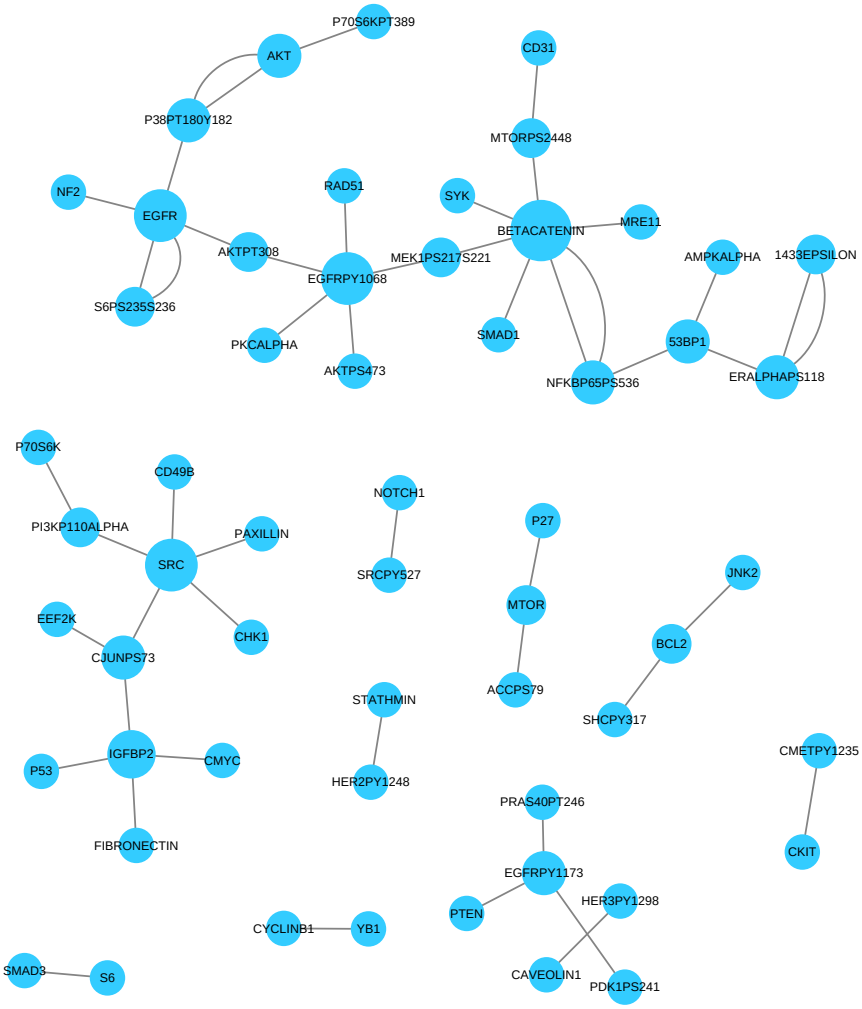


Figure 15: Nodes represent proteins that appear in the top 50 pairwise ranking embeddings for glioblastoma multiforme; edges represent that two proteins participate in a pairwise rank order feature together

LUAD

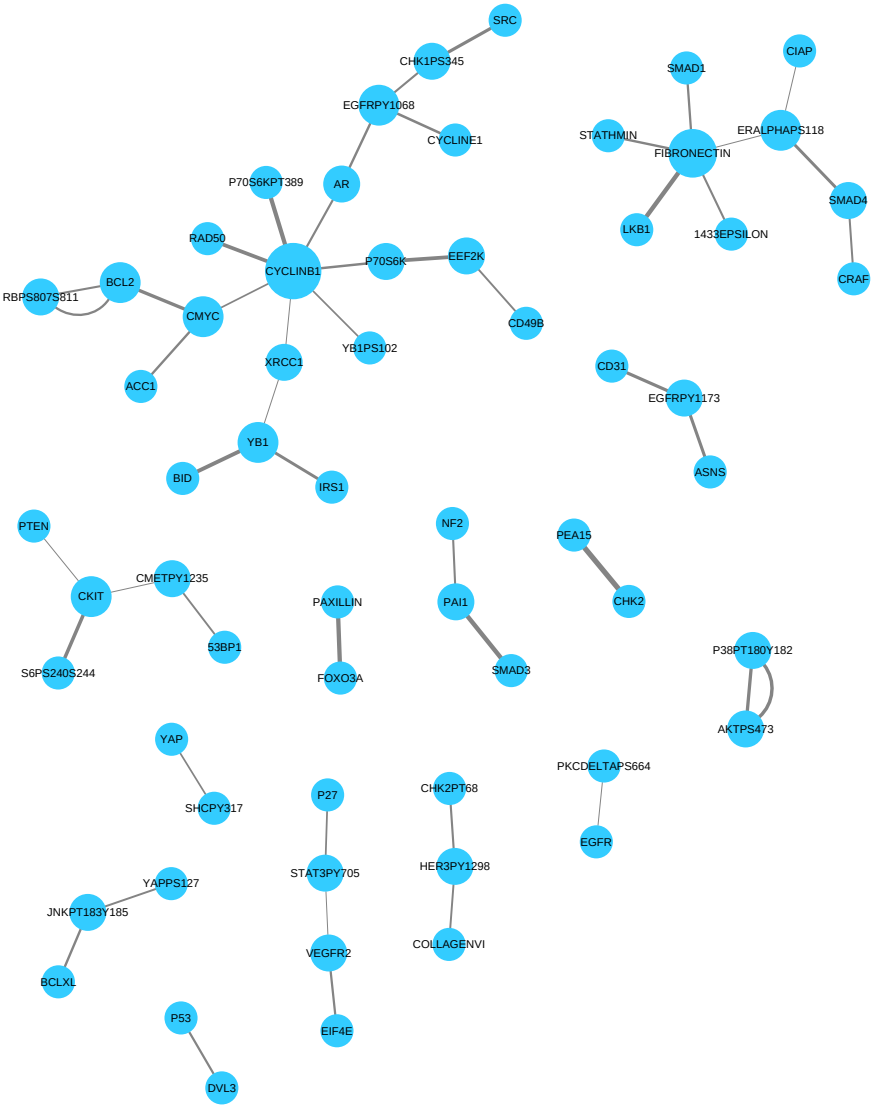


Figure 16: Nodes represent proteins that appear in the top 50 pairwise ranking embeddings for lung adenocarcinoma; edges represent that two proteins participate in a pairwise rank order feature together

LUSC

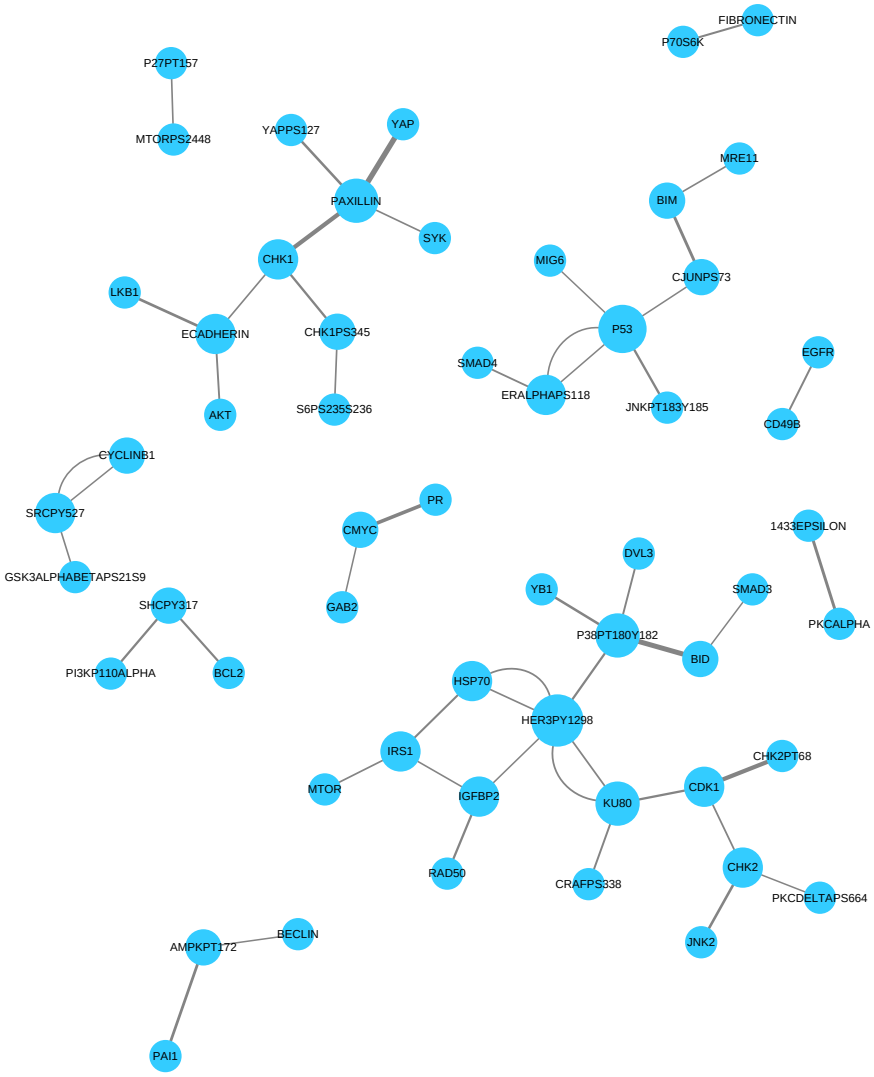


Figure 17: Nodes represent proteins that appear in the top 50 pairwise ranking embeddings for lung squamous cell carcinoma; edges represent that two proteins participate in a pairwise rank order feature together

UCEC

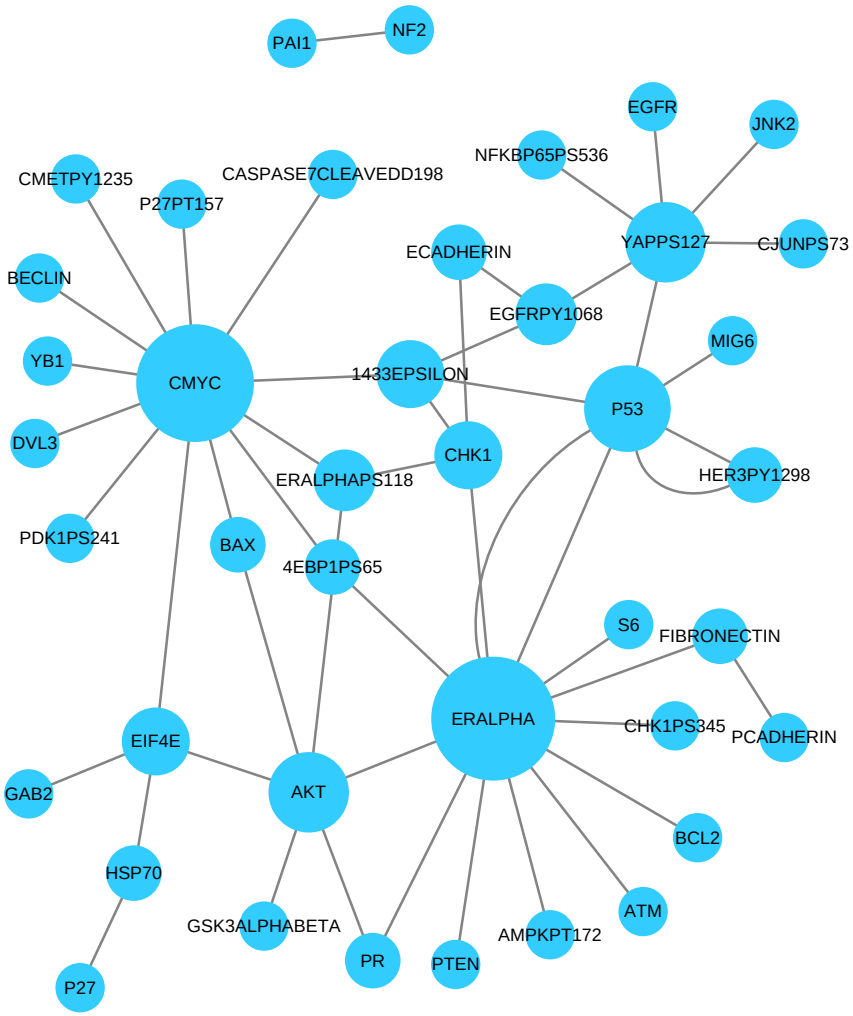
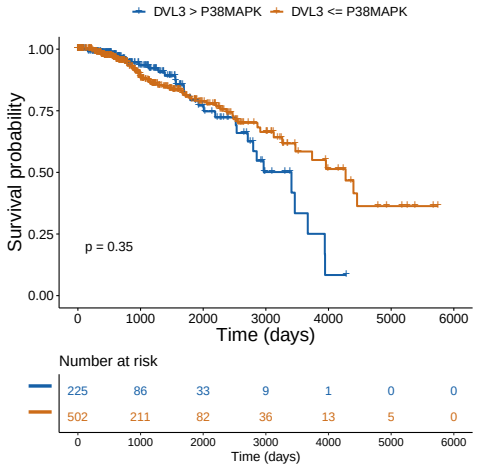


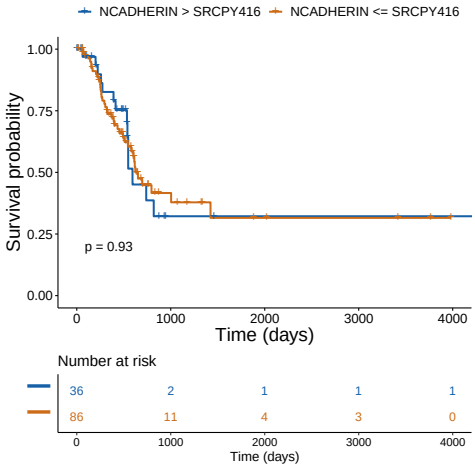
Figure 18: Nodes represent proteins that appear in the top 50 pairwise ranking embeddings for uterine corpus endometrial carcinoma; edges represent that two proteins participate in a pairwise rank order feature together

2.3 Top Predictive PRER are Prognostic Biomarkers

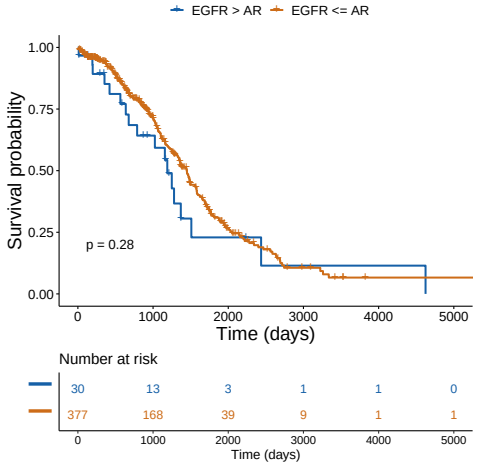
To check the relevance of the PRER genes, we check whether the patients with different feature values are different in terms of their survival distribution. For a PRER protein pair, A-B, we group patients such that those patients with protein expression values A>B are in one group and the rest are in the other group. We then check if the survival distribution of these groups is different using the log-rank test. The Figure S 19 shows the KM plots for each cancer type and the top-ranked feature.



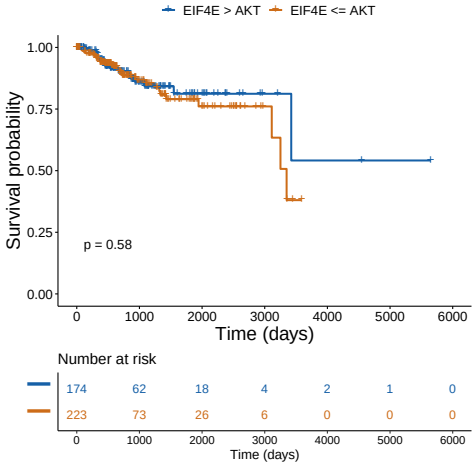
(a) BRCA



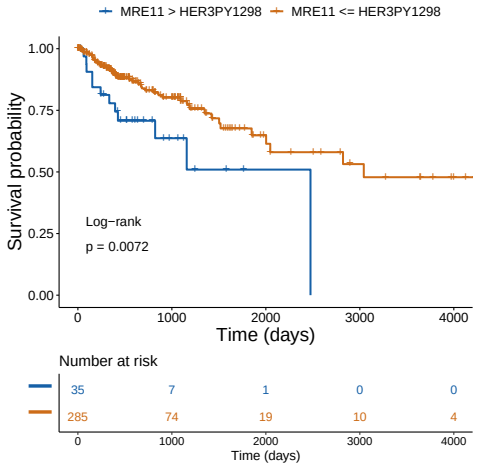
(b) BLCA



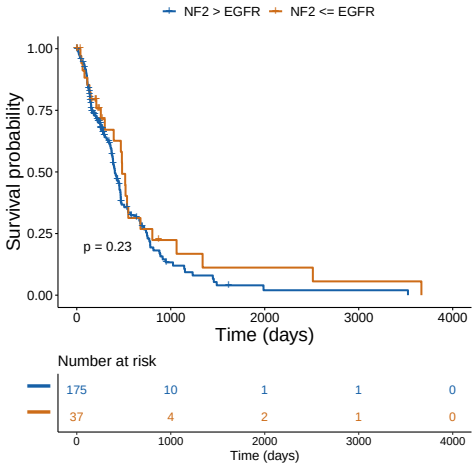
(c) OV



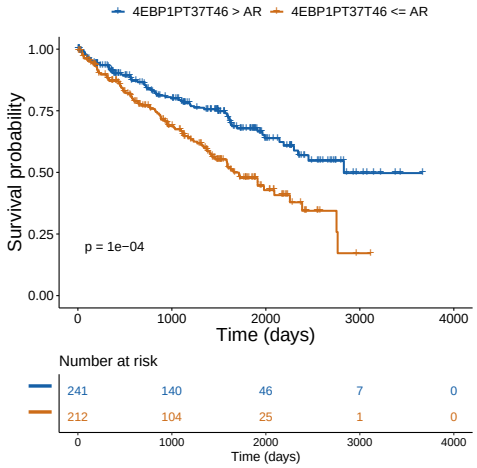
(d) UCEC



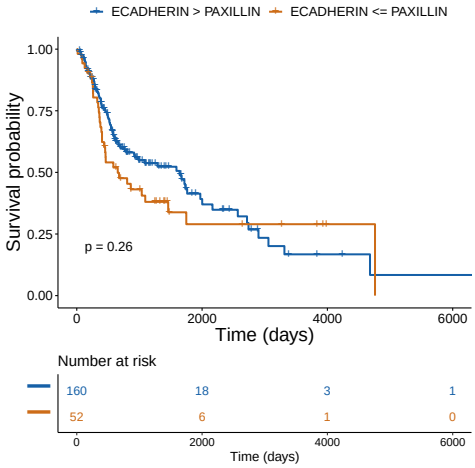
(e) COAD



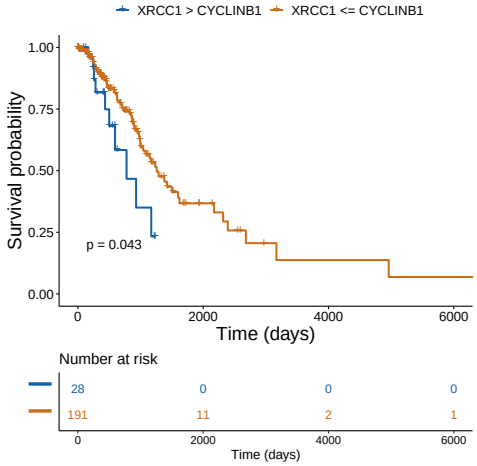
(f) GBM



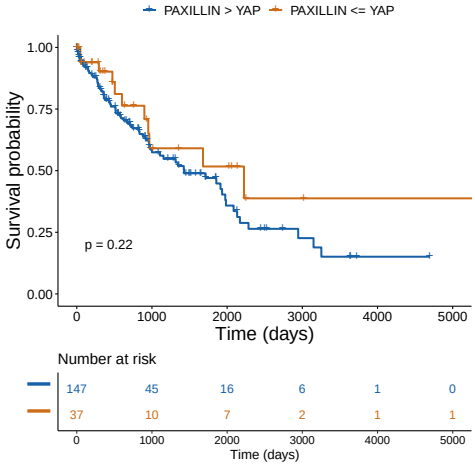
(g) KIRC



(h) HNSC



(i) LUAD



(j) LUSC

Figure 19: Kaplan-Meier plot for each cancer type based on overall survival. Number at risk denotes the number of patients at risk at a given time, and p-value is calculated with the log-rank test.