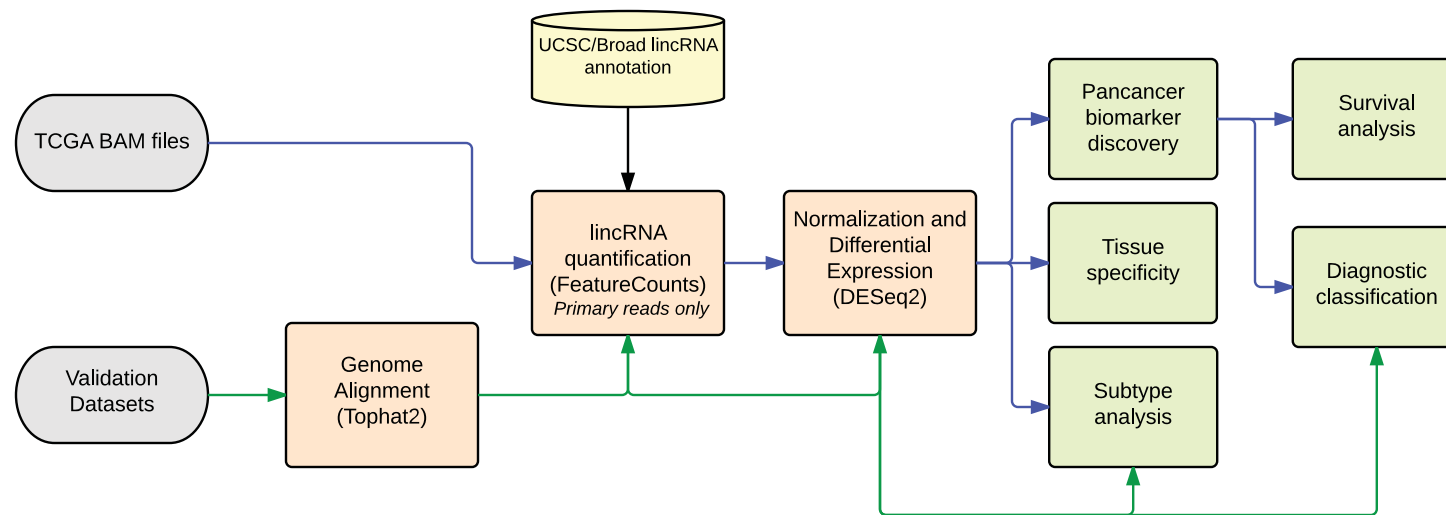
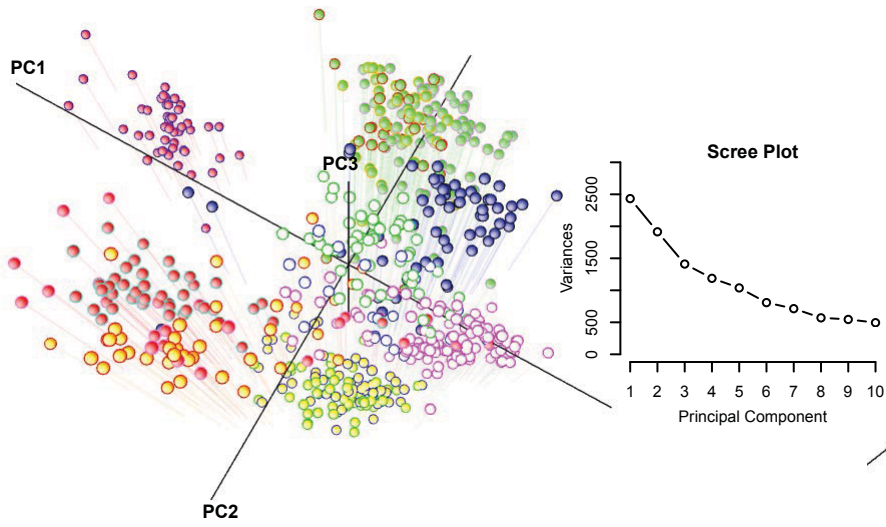


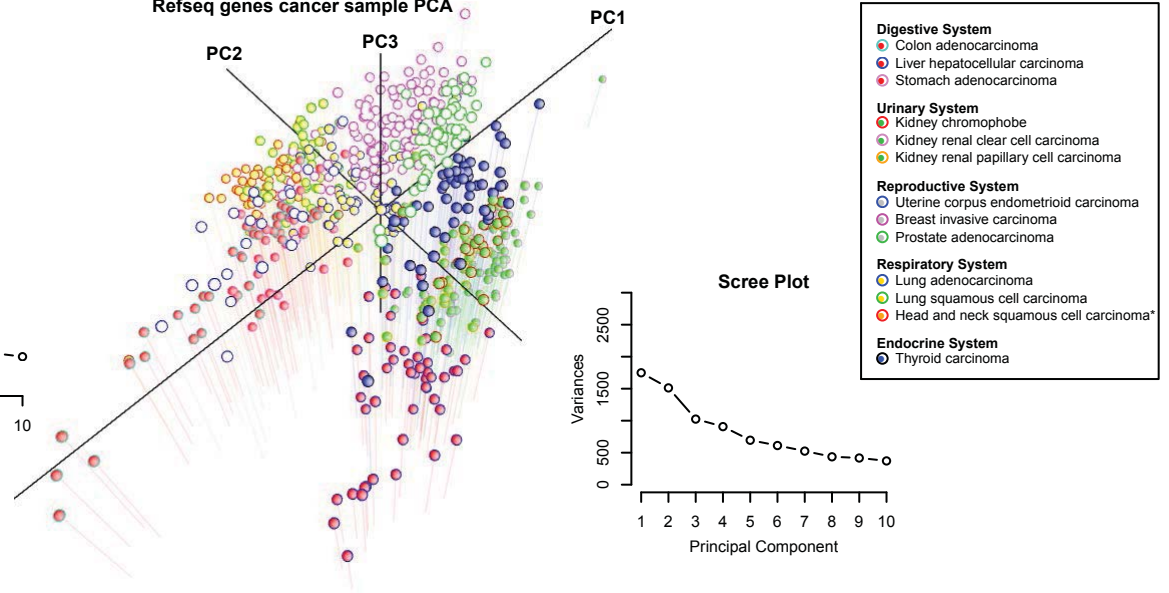
Supplementary Figure 1



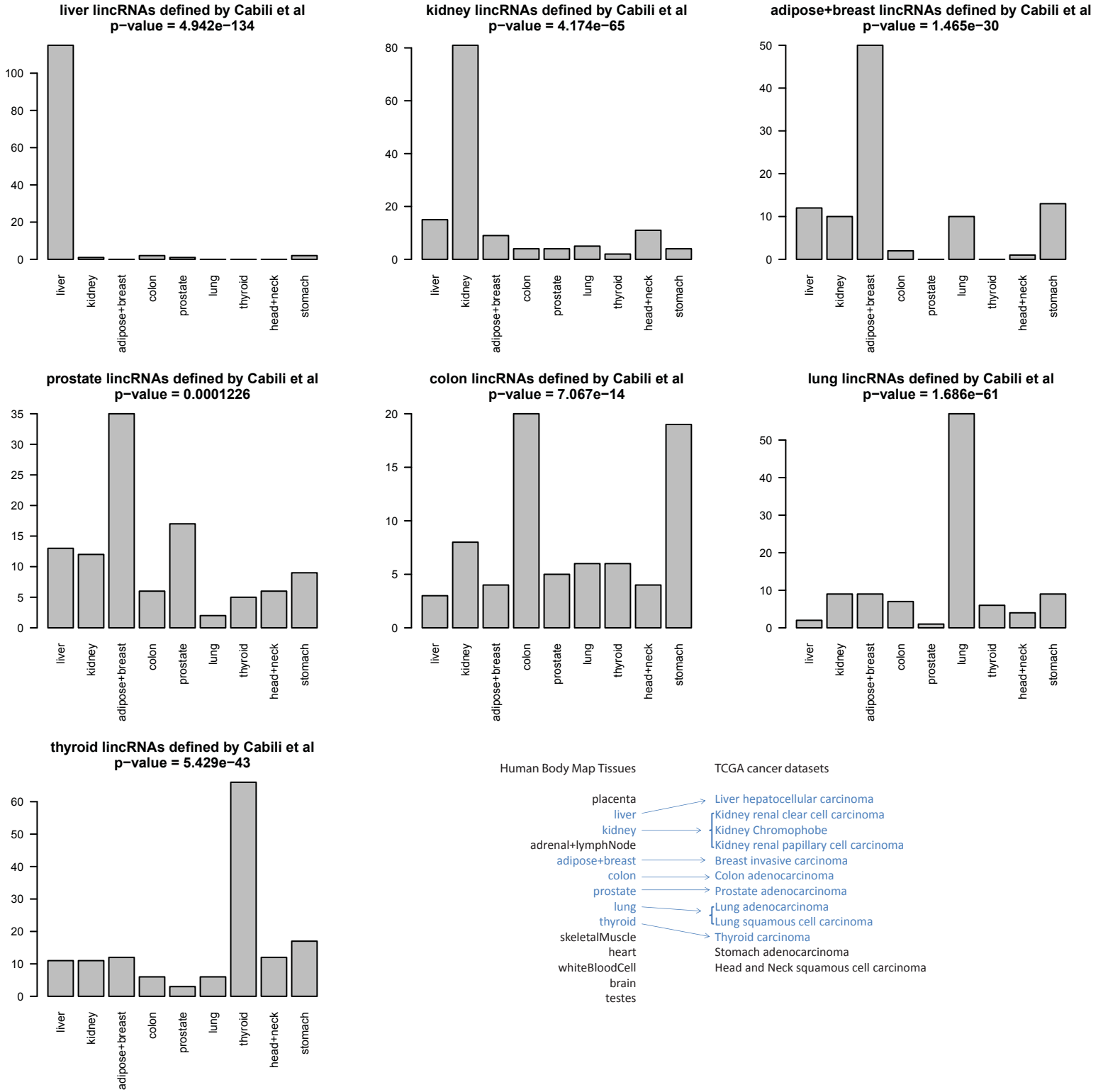
Refseq genes adjacent normal sample PCA



Refseq genes cancer sample PCA



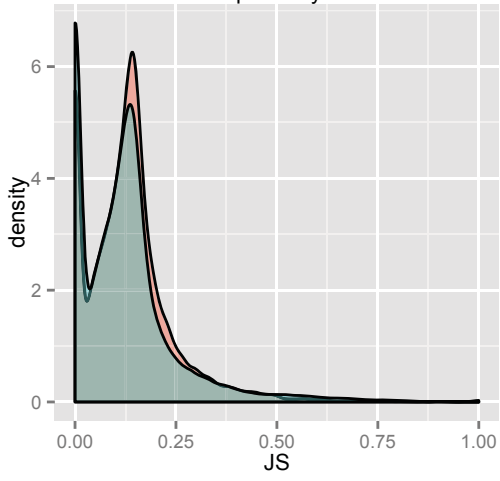
- Digestive System**
  - Colon adenocarcinoma
  - Liver hepatocellular carcinoma
  - Stomach adenocarcinoma
- Urinary System**
  - Kidney chromophobe
  - Kidney renal clear cell carcinoma
  - Kidney renal papillary cell carcinoma
- Reproductive System**
  - Uterine corpus endometrioid carcinoma
  - Breast invasive carcinoma
  - Prostate adenocarcinoma
- Respiratory System**
  - Lung adenocarcinoma
  - Lung squamous cell carcinoma
  - Head and neck squamous cell carcinoma\*
- Endocrine System**
  - Thyroid carcinoma



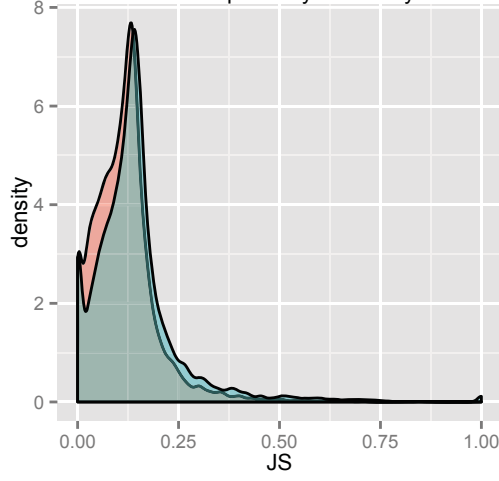
## JS scores by tissue

cancer normal

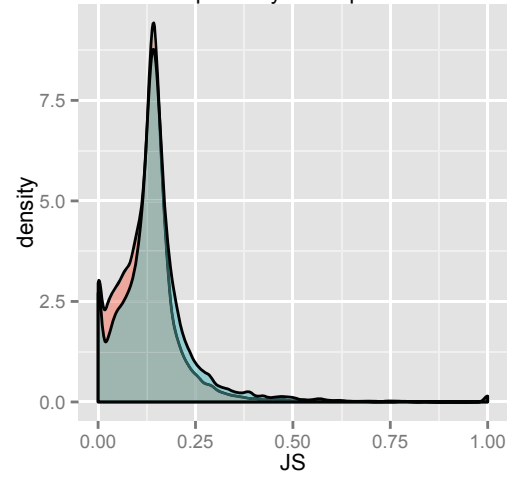
Tissue specificity for liver



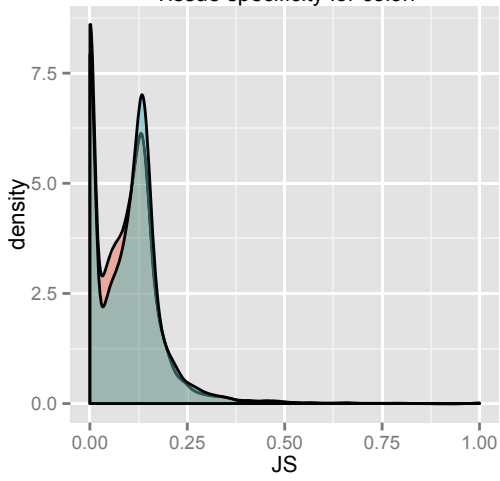
Tissue specificity for kidney



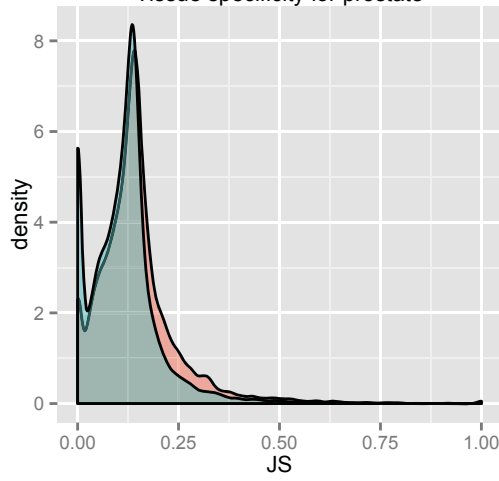
Tissue specificity for adipose+breast



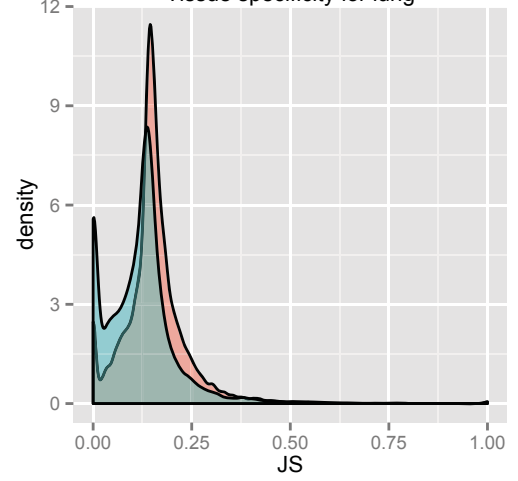
Tissue specificity for colon



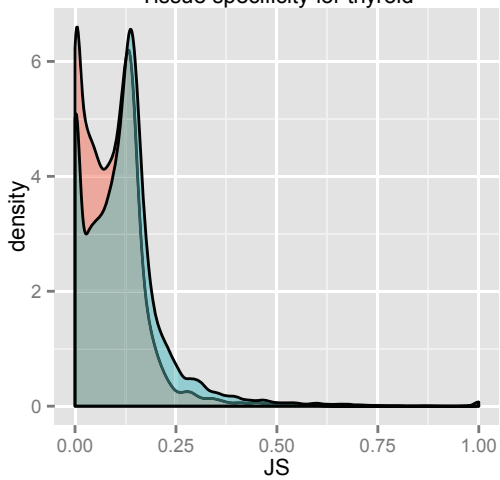
Tissue specificity for prostate



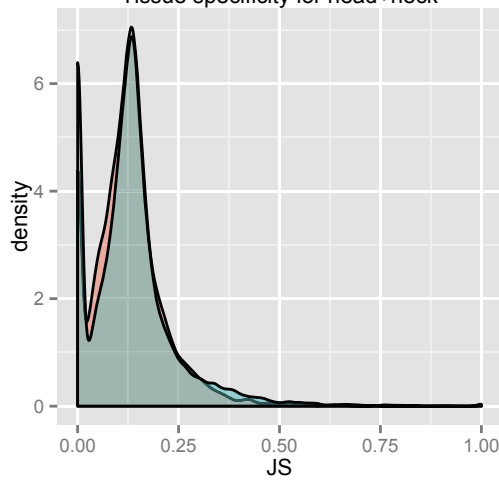
Tissue specificity for lung



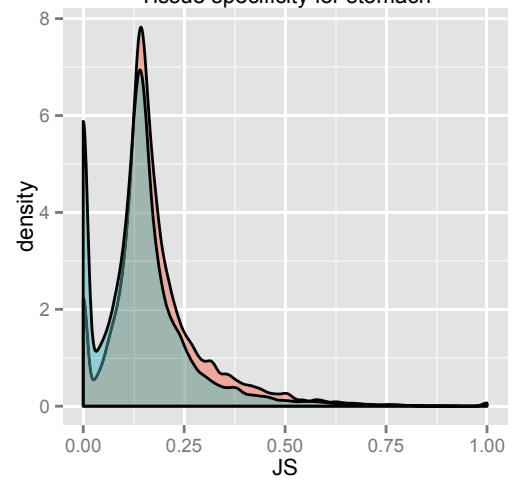
Tissue specificity for thyroid



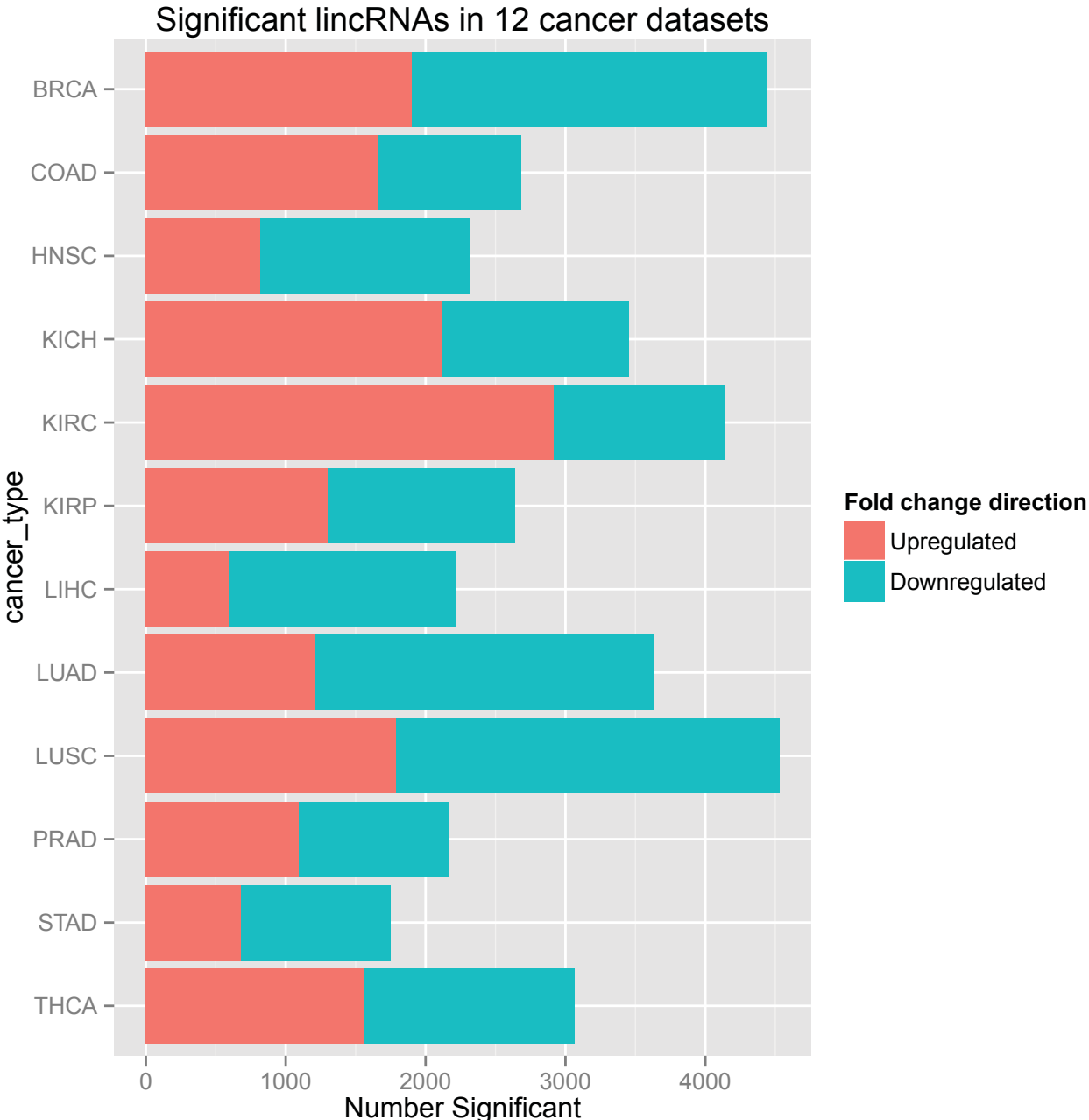
Tissue specificity for head+neck



Tissue specificity for stomach



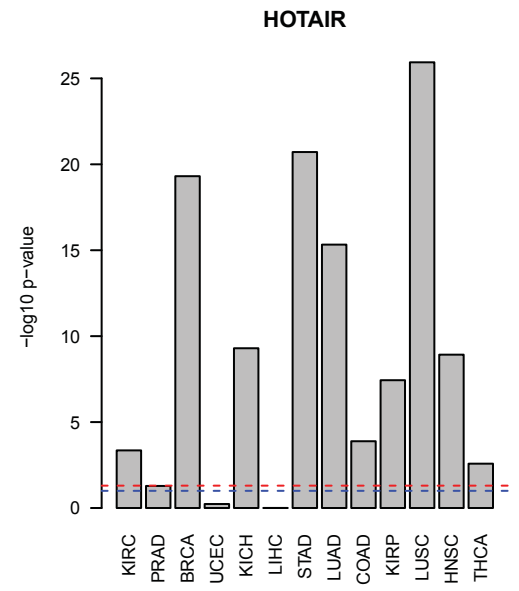
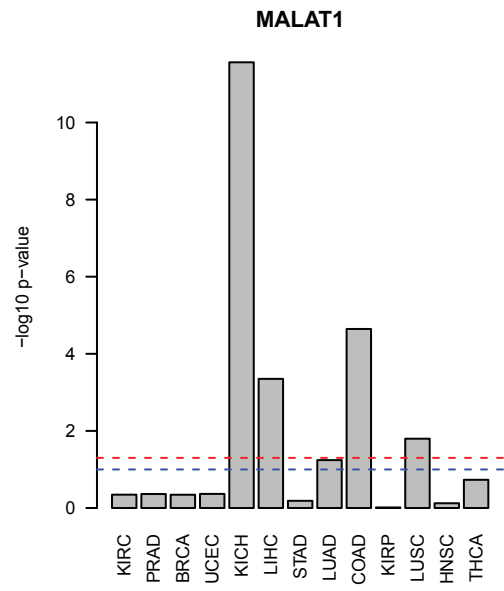
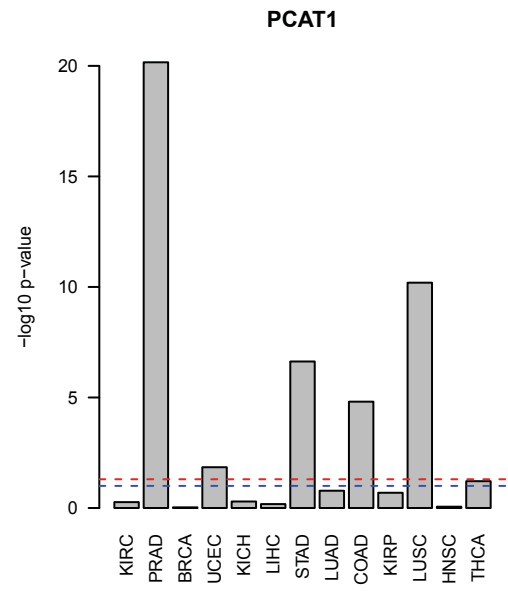
Supplementary Figure 5



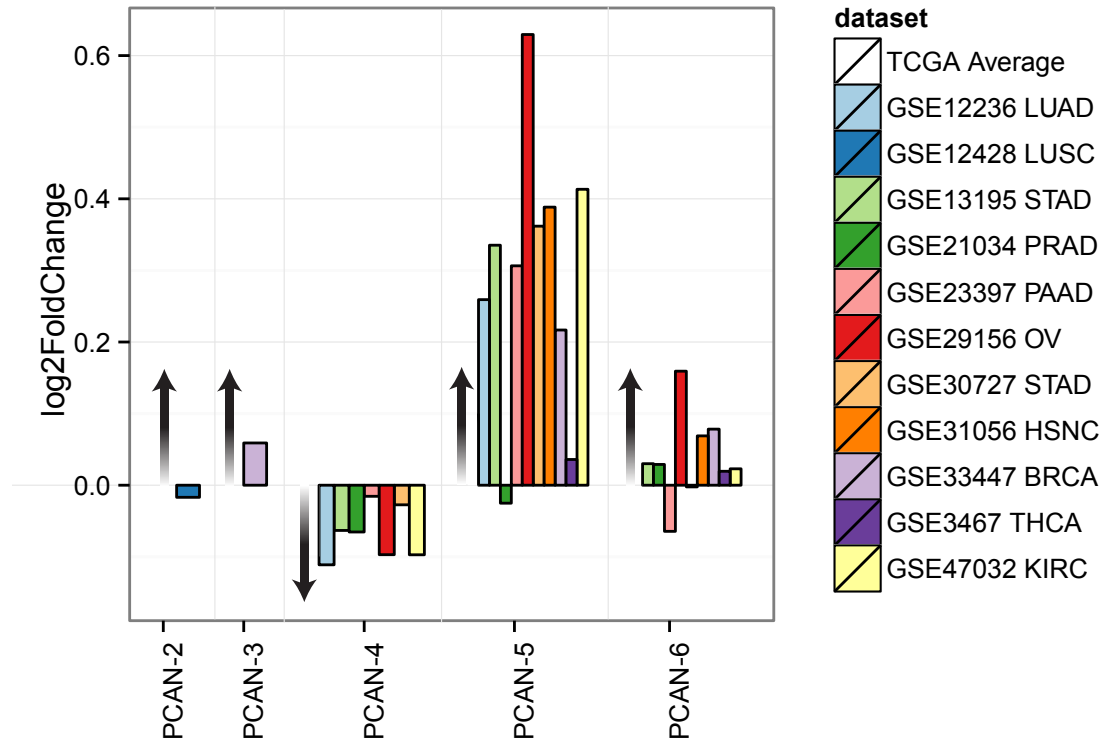
Supplementary Figure 6



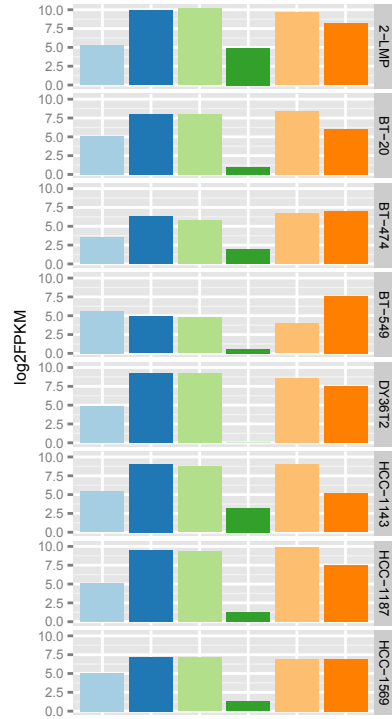
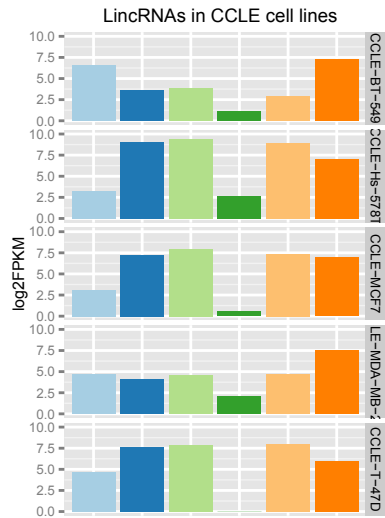
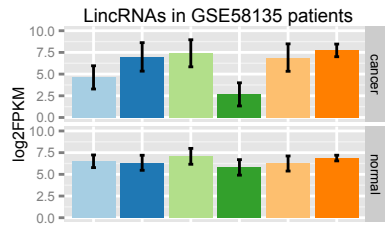
Supplementary Figure 7



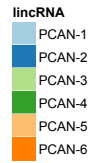
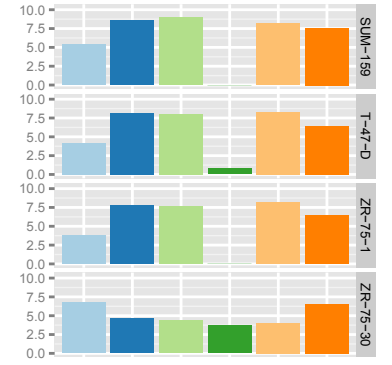
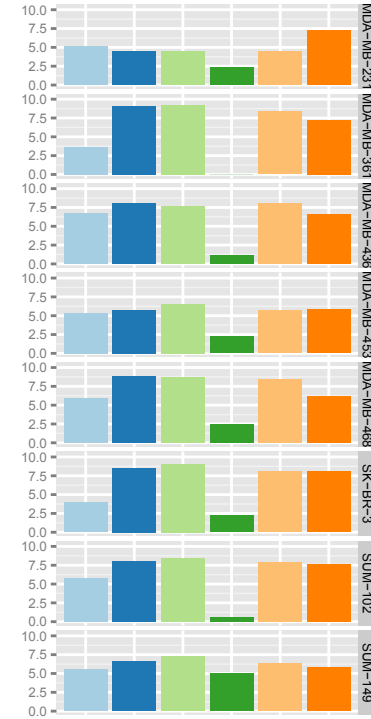
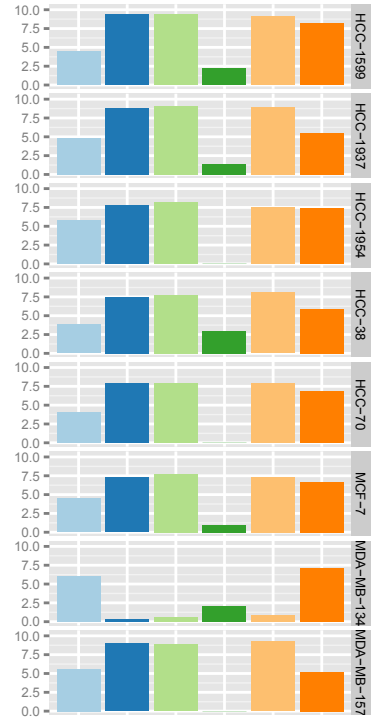
## Microarray Validation

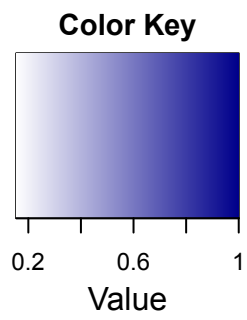




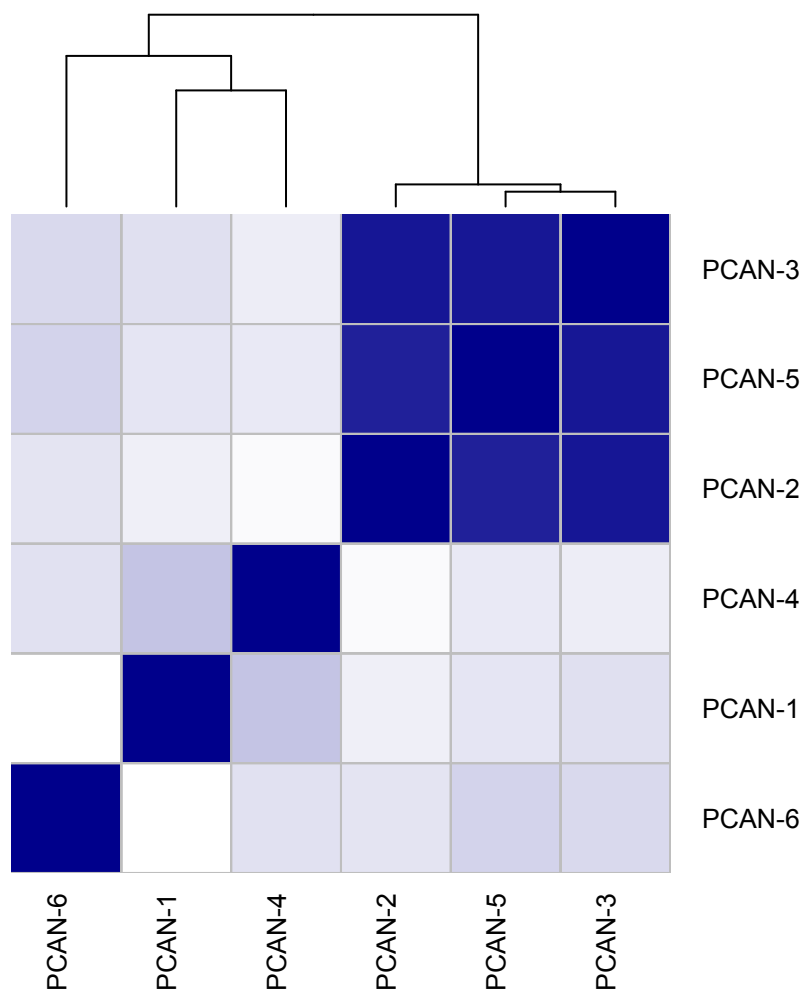


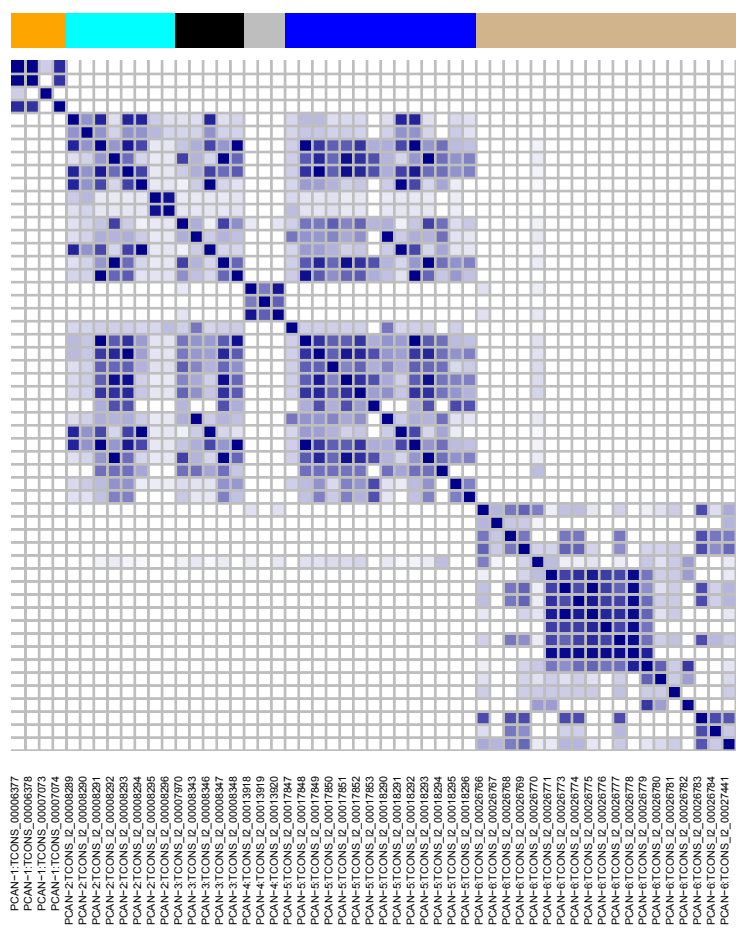
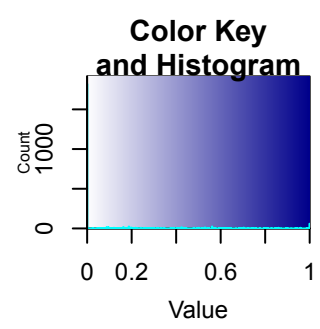
### LincRNAs in GSE58135 cell lines





# LincRNA expression correlation

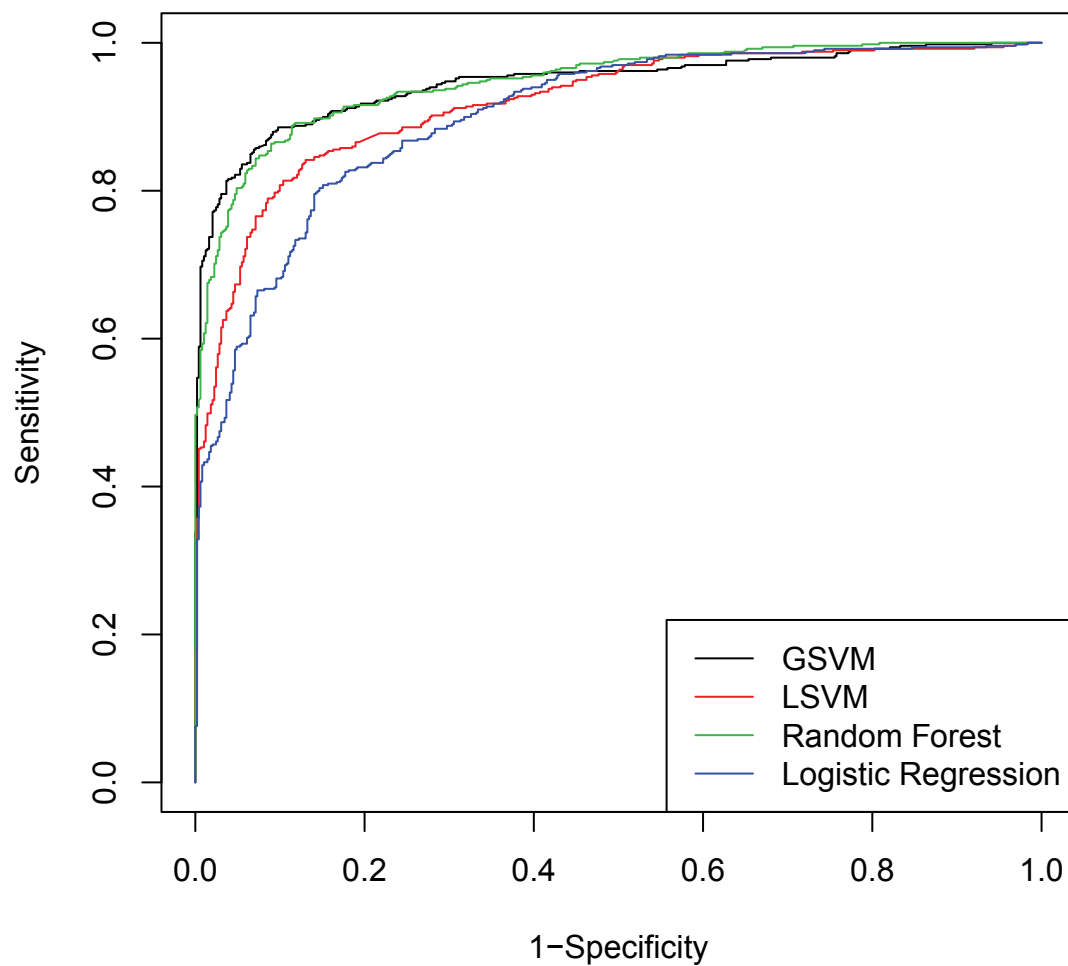




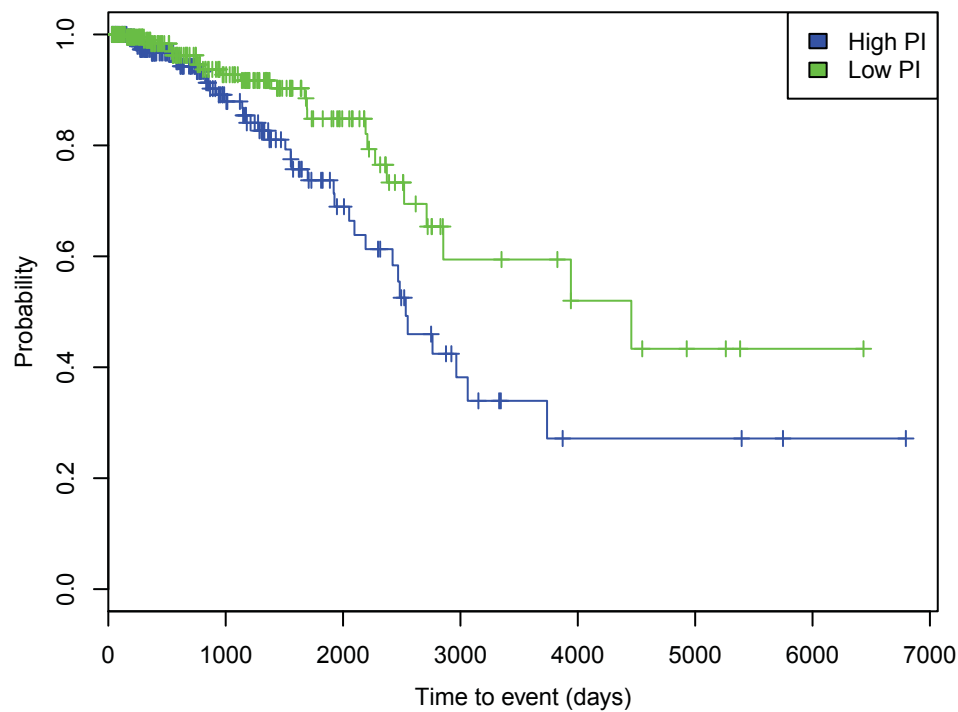
PCAN-1.TCONS\_00006377  
PCAN-1.TCONS\_00006378  
PCAN-1.TCONS\_00007073  
PCAN-1.TCONS\_00007074  
PCAN-2.TCONS\_00008289  
PCAN-2.TCONS\_00008290  
PCAN-2.TCONS\_00008291  
PCAN-2.TCONS\_00008292  
PCAN-2.TCONS\_00008293  
PCAN-2.TCONS\_00008294  
PCAN-2.TCONS\_00008295  
PCAN-2.TCONS\_00008296  
PCAN-3.TCONS\_00007970  
PCAN-3.TCONS\_00008343  
PCAN-3.TCONS\_00008346  
PCAN-3.TCONS\_00008347  
PCAN-3.TCONS\_00008348  
PCAN-4.TCONS\_00013918  
PCAN-4.TCONS\_00013919  
PCAN-4.TCONS\_00013920  
PCAN-5.TCONS\_00017847  
PCAN-5.TCONS\_00017848  
PCAN-5.TCONS\_00017849  
PCAN-5.TCONS\_00017850  
PCAN-5.TCONS\_00017851  
PCAN-5.TCONS\_00017852  
PCAN-5.TCONS\_00017853  
PCAN-5.TCONS\_00018290  
PCAN-5.TCONS\_00018291  
PCAN-5.TCONS\_00018292  
PCAN-5.TCONS\_00018293  
PCAN-5.TCONS\_00018294  
PCAN-5.TCONS\_00018295  
PCAN-5.TCONS\_00018296  
PCAN-6.TCONS\_00026766  
PCAN-6.TCONS\_00026767  
PCAN-6.TCONS\_00026768  
PCAN-6.TCONS\_00026769  
PCAN-6.TCONS\_00026770  
PCAN-6.TCONS\_00026771  
PCAN-6.TCONS\_00026773  
PCAN-6.TCONS\_00026774  
PCAN-6.TCONS\_00026775  
PCAN-6.TCONS\_00026776  
PCAN-6.TCONS\_00026777  
PCAN-6.TCONS\_00026778  
PCAN-6.TCONS\_00026779  
PCAN-6.TCONS\_00026780  
PCAN-6.TCONS\_00026781  
PCAN-6.TCONS\_00026782  
PCAN-6.TCONS\_00026783  
PCAN-6.TCONS\_00026784  
PCAN-6.TCONS\_00027441

PCAN-1.TCONS\_00006377  
PCAN-1.TCONS\_00006378  
PCAN-1.TCONS\_00007073  
PCAN-1.TCONS\_00007074  
PCAN-2.TCONS\_00008289  
PCAN-2.TCONS\_00008290  
PCAN-2.TCONS\_00008291  
PCAN-2.TCONS\_00008292  
PCAN-2.TCONS\_00008293  
PCAN-2.TCONS\_00008294  
PCAN-2.TCONS\_00008295  
PCAN-2.TCONS\_00008296  
PCAN-3.TCONS\_00007970  
PCAN-3.TCONS\_00008343  
PCAN-3.TCONS\_00008346  
PCAN-3.TCONS\_00008347  
PCAN-3.TCONS\_00008348  
PCAN-4.TCONS\_00013918  
PCAN-4.TCONS\_00013919  
PCAN-4.TCONS\_00013920  
PCAN-5.TCONS\_00017847  
PCAN-5.TCONS\_00017848  
PCAN-5.TCONS\_00017849  
PCAN-5.TCONS\_00017850  
PCAN-5.TCONS\_00017851  
PCAN-5.TCONS\_00017852  
PCAN-5.TCONS\_00017853  
PCAN-5.TCONS\_00018290  
PCAN-5.TCONS\_00018291  
PCAN-5.TCONS\_00018292  
PCAN-5.TCONS\_00018293  
PCAN-5.TCONS\_00018294  
PCAN-5.TCONS\_00018295  
PCAN-5.TCONS\_00018296  
PCAN-6.TCONS\_00026766  
PCAN-6.TCONS\_00026767  
PCAN-6.TCONS\_00026768  
PCAN-6.TCONS\_00026769  
PCAN-6.TCONS\_00026770  
PCAN-6.TCONS\_00026771  
PCAN-6.TCONS\_00026773  
PCAN-6.TCONS\_00026774  
PCAN-6.TCONS\_00026775  
PCAN-6.TCONS\_00026776  
PCAN-6.TCONS\_00026777  
PCAN-6.TCONS\_00026778  
PCAN-6.TCONS\_00026779  
PCAN-6.TCONS\_00026780  
PCAN-6.TCONS\_00026781  
PCAN-6.TCONS\_00026782  
PCAN-6.TCONS\_00026783  
PCAN-6.TCONS\_00026784  
PCAN-6.TCONS\_00027441

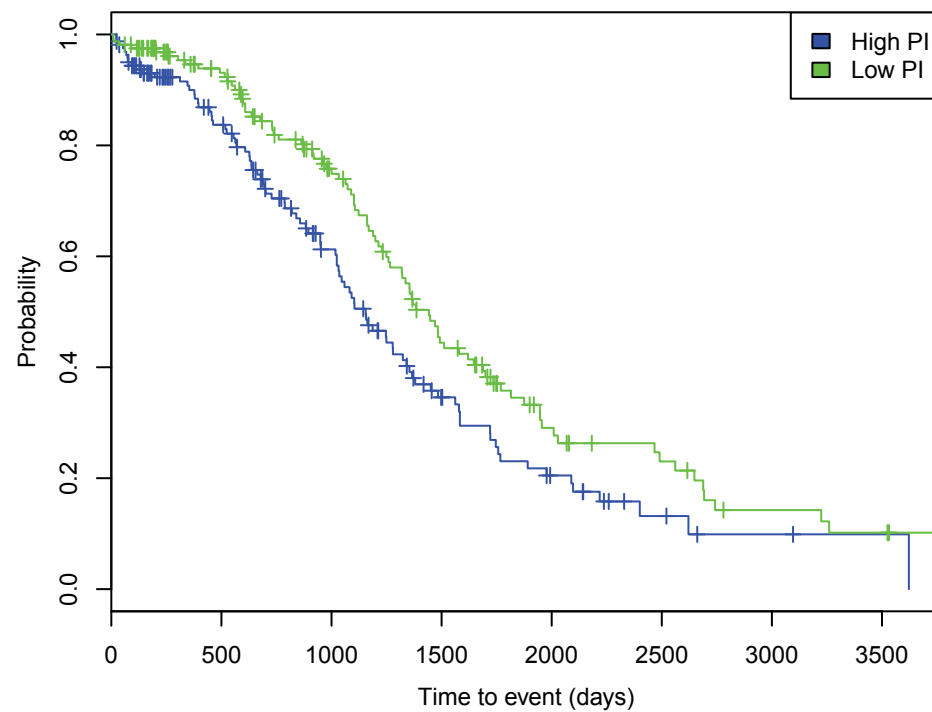
### ROC TCGA training data



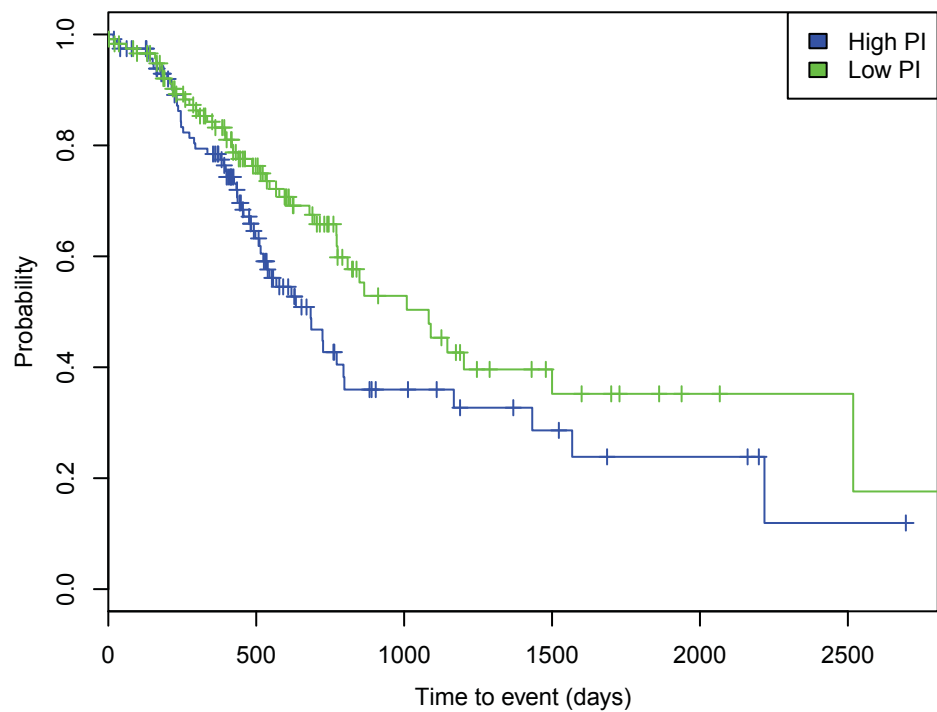
A

BRCA overall survival lincRNA model  $p = 0.01248$ 

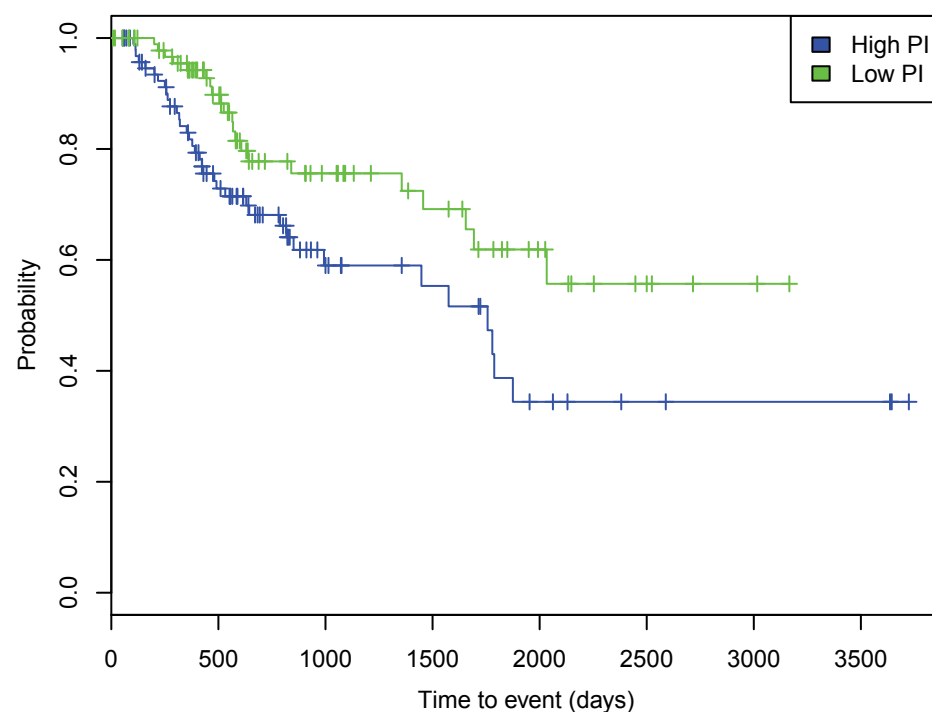
B

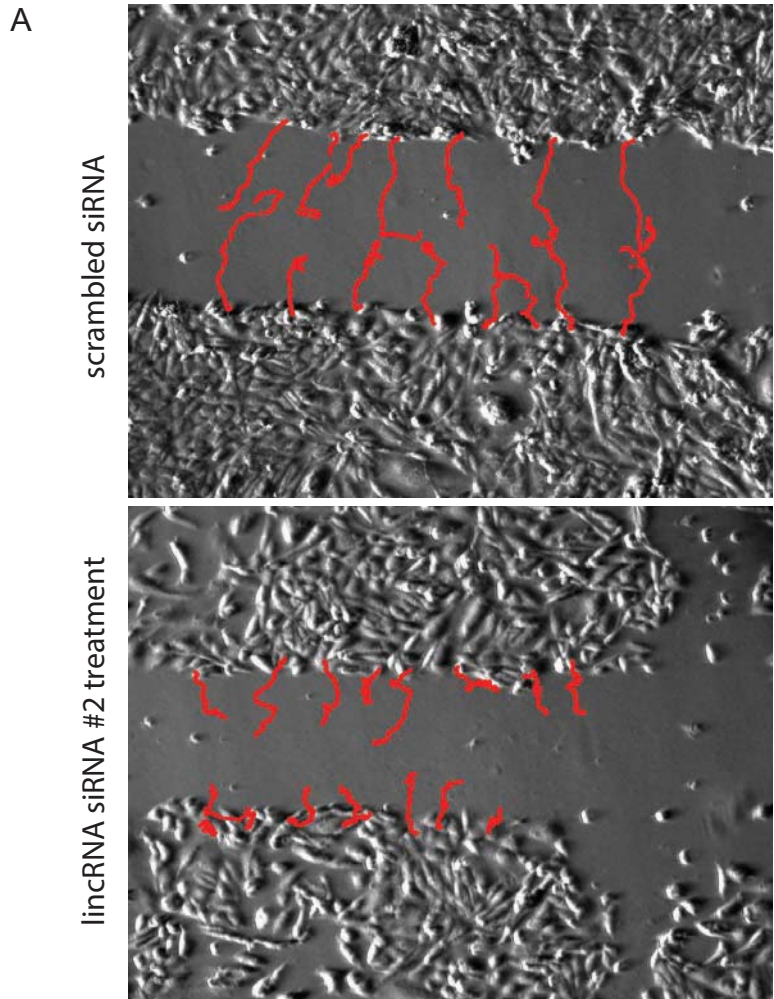
OV overall survival (Stage 3) lincRNA model  $p = 0.01012$ 

C

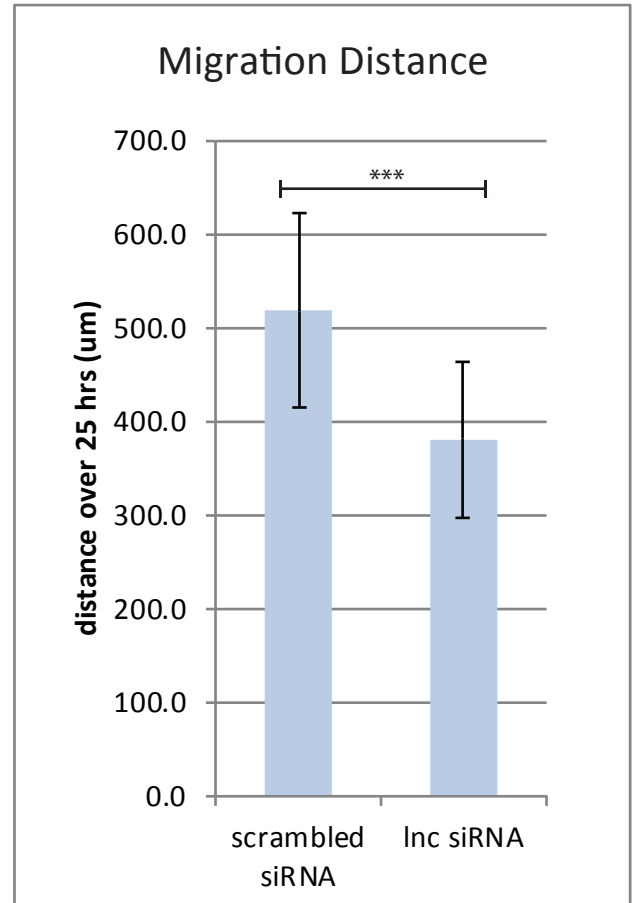
LUAD relapse free survival lincRNA model  $p = 0.04159$ 

D

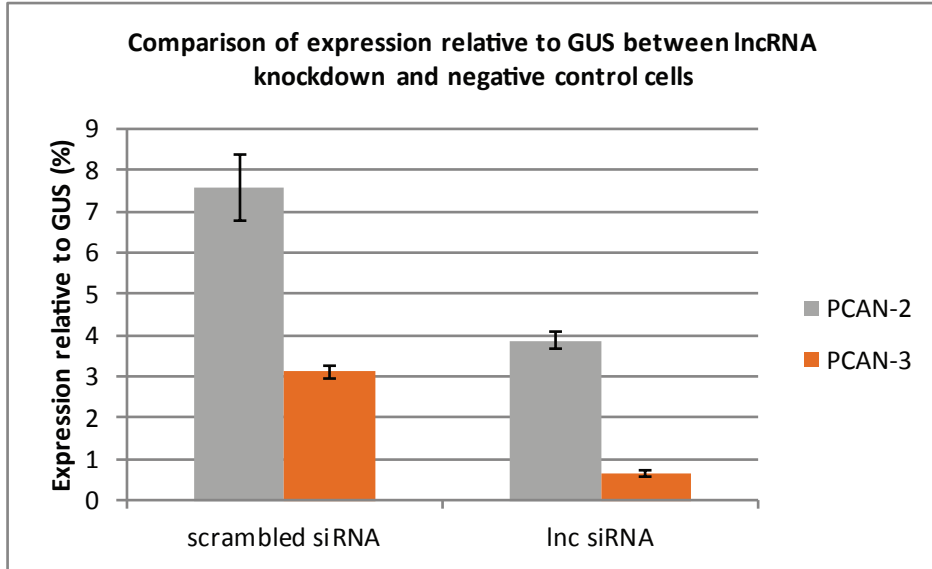
LUSC relapse free survival lincRNA model  $p = 0.01348$ 



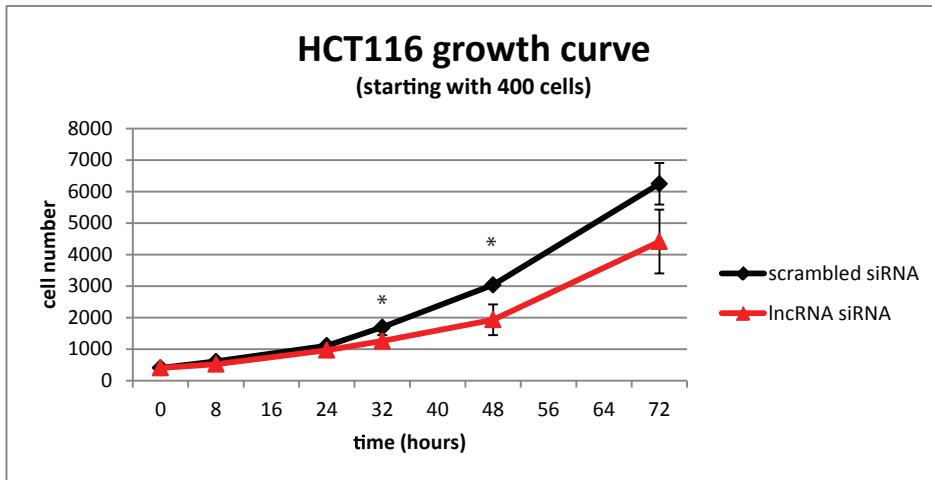
B



A



B



C

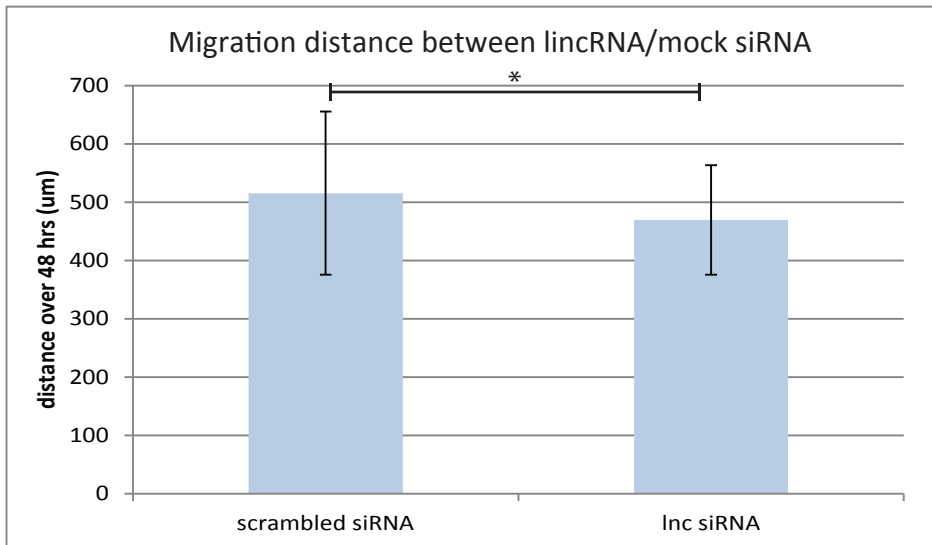


Table S1: Tabulation of samples used in study

TCGA datasets				
Dataset	Primary	Adjacent	Paired	# mapped reads
Breast invasive carcinoma	1059	111	111	193560468
Colon adenocarcinoma	444	41	41	44872617
Head and Neck squamous cell carcinoma	498	43	41	57516555
Kidney Chromophobe	66	25	25	42982445
Kidney renal clear cell carcinoma	531	72	72	94701140
Kidney renal papillary cell carcinoma	226	32	32	48960010
Liver hepatocellular carcinoma	210	50	50	46819687
Lung adenocarcinoma	489	58	57	69621821
Lung squamous cell carcinoma	489	50	50	86702630
Prostate adenocarcinoma	419	52	52	98002347
Stomach adenocarcinoma	285	32	30	65146435
Thyroid carcinoma	498	59	59	127022122
<b>Subtotal</b>	<b>5214</b>	<b>625</b>	<b>620</b>	<b>975908277</b>
<b>Samples used for survival analysis</b>				
<b>Cancer</b>				
Breast invasive carcinoma	463			
Ovarian serous adenocarcinoma	406			
Lung adenocarcinoma	193			
Lung squamous cell carcinoma	139			
<b>Samples used for subtype analysis</b>				
<b>Cancer</b>				
Breast invasive carcinoma	521			
<b>Validation Datasets</b>				
<b>Accssion</b>	<b>Cancer</b>	<b>Non-cancer</b>	<b>Additional Details</b>	
GSE25599	10	10	Single end, 36 bp, Liver Cancer (HBV)	
GSE58135	84	56	Paired, 100 bp, Breast Cancer	
GSE50760	36	36	Paired, 100 bp, Colon cancer	
Our BRCA dataset	10	6	Single end, 100 bp, BRCA; normal samples: GSE52194, GSE45326, GSE30611	
<b>Whole Human Genome Oligo Microarray G4112A</b>				
GSE12428	Cancer 34	Non-cancer 28	<b>Additional Details</b> Microarray, lung squamous cell carcinoma	
<b>Affymetrix Human Exon 1.0 ST Array</b>				
GSE30727	Cancer 30	Non-cancer 30	<b>Additional Details</b> Microarray, stomach cancer	
GSE47032	20	20	Microarray, kidney clear cell cancer	
GSE23397	15	6	Microarray, pancreatic cancer	
GSE21034	150	29	Microarray, prostate cancer	
GSE13195	25	25	Microarray, stomach cancer	
GSE12236	16	24	Microarray, lung cancer	
GSE29156	9	11	Microarray, ovarian cancer	
<b>Affymetrix Human Genome U133 Plus 2.0 Array</b>				
GSE3467	Cancer 9	Non-cancer 9	<b>Additional Details</b> Microarray, thyroid samples	
GSE31056	23	72	Microarray, oral carcinoma	
<b>SurePrint G3 Human GE 8x60K Microarray</b>				
GSE33447	Cancer 8	Non-cancer 8	<b>Additional Details</b> Microarray, breast cancer samples	
<b>GSE58135 subtype analysis</b>				
	ER+/HER2- 42	Triple Negative 42		
<b>Total unique biological samples</b>				
	<b>3354</b>			



Table S2

<u>lincRNA</u>	<u>JS Score</u>	<u>Associated Tissue</u>
XLOC_010690	0.766204217	kidney
XLOC_007596	0.857703592	liver
XLOC_003732	0.752100476	colon
XLOC_010399	0.768482652	kidney
XLOC_008645	0.758646776	lung
XLOC_011257	0.763064551	liver
XLOC_001387	0.77810803	liver
XLOC_007597	0.824089647	liver
XLOC_000947	0.903394197	adipose+breast
XLOC_011275	0.911613639	liver
XLOC_004719	0.941948946	kidney
XLOC_013795	0.781746482	adipose+breast
XLOC_004102	0.873732325	kidney
XLOC_004360	0.785572253	adipose+breast
XLOC_003239	0.775039257	lung
XLOC_008455	0.770565259	adipose+breast
XLOC_004177	1	lung
XLOC_001901	0.760514635	liver
XLOC_004836	0.754396243	adipose+breast
XLOC_004514	0.794355439	liver
XLOC_004801	0.906048522	adipose+breast
XLOC_004261	0.817042521	adipose+breast
XLOC_000224	0.799618236	kidney
XLOC_004269	0.867461193	kidney
XLOC_008454	0.792515074	adipose+breast
XLOC_004515	0.752963989	liver
XLOC_005515	0.767317503	thyroid
XLOC_008260	0.902906732	lung
XLOC_001704	0.848966352	liver
XLOC_013779	0.756769238	head+neck
XLOC_007883	0.926395141	liver
XLOC_006805	0.772086177	kidney
XLOC_008446	1	kidney
XLOC_003315	0.791869141	colon
XLOC_012693	0.755588854	liver
XLOC_000857	0.760980892	kidney
XLOC_009690	0.763803004	lung
XLOC_001425	0.900101922	kidney
XLOC_005607	0.808412113	liver
XLOC_010419	0.779716586	thyroid
XLOC_013465	0.754345459	prostate
XLOC_007017	0.773648455	adipose+breast
XLOC_008602	0.767872806	liver
XLOC_007405	0.804291982	prostate
XLOC_014355	0.762921632	prostate
XLOC_013037	0.768622905	prostate
XLOC_001378	0.862077396	kidney
XLOC_005775	0.817266832	kidney
XLOC_000430	0.936023257	liver
XLOC_I2_000357	0.767464299	lung
XLOC_I2_013883	0.914582666	prostate
XLOC_I2_001548	0.763547727	adipose+breast
XLOC_I2_004342	0.768931463	prostate

Table S3

lincRNA	Human Body Map ID	Chromosome	lincRNA strand	lincRNA start	lincRNA width	lincRNA end	Closest Gene	Distance	Gene start	Gene width	Gene end	Ensembl Annotation	Ensembl gene match
PCAN-1	XLOC_002996	chr3	+	195367106	10516	195377622	APOD	56029	195295573	15504	195311077	LincRNA, Unannotated	N/A
PCAN-2	XLOC_I2_004121	chr14	+	19650018	45163	19695181	POTEG	65075	19553365	31578	19584943	LincRNA, Unannotated	AL589743.1
PCAN-3	XLOC_I2_004340	chr14	-	19856361	68973	19925334	POTEM	58620	19983954	36319	20020273	LincRNA, Unannotated	CTD-2314B22.3
PCAN-4	XLOC_I2_007509	chr2	+	114298969	29137	114328106	FOXD4L1	40241	114256661	2067	114258728	Processed Pseudogene, Unannotated	N/A
PCAN-5	XLOC_I2_009441	chr22	-	16101370	91857	16193227	POTEH	63105	16256332	31606	16287938	LincRNA, Processed Transcript, Unannotated	AP000525.9
PCAN-6	XLOC_I2_013931	chr7	-	97503667	98000	97601667	ASNS	1812	97481429	20426	97501855	Processed Pseudogene, Unannotated	N/A

\*Ensembl v75 (GRCh37), annotated exons

Table S4

Assay Label	LincRNA target	Primers
4121_1:	PCAN-2 (XLOC_I2_004121)	F: 5'-AGCTTCGGAGAAGCAGTGGT-3' R: 5'-TTCTTTCCGCGGAGACCT-3'
4340_4	PCAN-3 (XLOC_I2_004340)	F: 5'-ACAGATGAACCGCGGAGAC-3' R: 5'-AGCTTCGGAGAAGCAGTGGT-3'
2996_3	PCAN-1 (XLOC_I2_002996)	F: 5'-TAAGGGTCATGGAGCTGGAG-3' R: 5'-ATCAGCTCCTCCCCGAGTAT-3'
7509_4	PCAN-4 (XLOC_I2_007509)	F: 5'-GAAGTTTAATGTTGCCAATGGA-3' R: 5'-GCCTTTGCACAGACTGACCT-3'
13931_6	PCAN-6 (XLOC_I2_013931)	F: 5'-ATCCAGAAGTGCAGCCAGTC-3' R: 5'-AGAAGTACATGGGGGTGTGG-3'

Supplemental Table V part 1

[http://lilab.research.bcm.edu/cpat/calculator\\_sub.php](http://lilab.research.bcm.edu/cpat/calculator_sub.php)

Reference: CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acid Research

**PCAN-1 (XLOC\_002996)**

Sequence Name	RNA Size	ORF Size	Ficket Score	Hexamer Score	Coding Probability	Coding Label
TCONS_00006377		6659	633	0.5291	0.173947044	0.760142576 yes
TCONS_00006378		1716	135	1.0111	0.171563902	0.080769211 no
TCONS_00007073		211	96	0.6124	-0.376596856	0.000483433 no
TCONS_00007074		1586	153	1.1382	0.213026381	0.17417223 no

**PCAN-2 (XLOC\_I2\_004121)**

Sequence Name	RNA Size	ORF Size	Ficket Score	Hexamer Score	Coding Probability	Coding Label
TCONS_L2_00008289		768	117	0.7464	0.041918727	0.013971842 no
TCONS_L2_00008290		549	192	0.8298	0.048534387	0.042017289 no
TCONS_L2_00008291		2102	276	0.484	-0.012649119	0.021810528 no
TCONS_L2_00008292		1406	327	0.639	-0.180683866	0.021204847 no
TCONS_L2_00008293		2391	267	0.42	-0.077650944	0.010473072 no
TCONS_L2_00008294		4116	267	0.42	-0.077650944	0.009441453 no
TCONS_L2_00008295		1060	342	0.5027	-0.119244657	0.02480705 no
TCONS_L2_00008296		990	342	0.5027	-0.119244657	0.024910079 no

**PCAN-3 (XLOC\_I2\_004340)**

Sequence Name	RNA Size	ORF Size	Ficket Score	Hexamer Score	Coding Probability	Coding Label
TCONS_L2_00007970		5950	459	0.6599	0.112352222	0.336164826 no
TCONS_L2_00008343		1169	144	0.6051	-0.107657632	0.004404995 no
TCONS_L2_00008346		4142	267	0.42	-0.077650944	0.009426701 no
TCONS_L2_00008347		1405	327	0.639	-0.180683866	0.021206108 no
TCONS_L2_00008348		570	168	0.904	-0.068320034	0.019342915 no

**PCAN-4 (XLOC\_I2\_007509)**

Sequence Name	RNA Size	ORF Size	Ficket Score	Hexamer Score	Coding Probability	Coding Label
TCONS_L2_00013918		2879	270	0.6819	-0.343879747	0.004149952 no
TCONS_L2_00013919		547	192	1.0539	0.129150962	0.131216094 no
TCONS_L2_00013920		2179	270	0.6819	-0.343879747	0.00432936 no

**PCAN-5 (XLOC\_I2\_009441)**

Sequence Name	RNA Size	ORF Size	Ficket Score	Hexamer Score	Coding Probability	Coding Label
TCONS_L2_00017847		609	150	0.8864	-0.136597695	0.009670352 no
TCONS_L2_00017848		880	168	0.868	-0.07522081	0.016248197 no
TCONS_L2_00017849		913	168	0.868	-0.07522081	0.016216201 no
TCONS_L2_00017850		763	168	0.868	-0.07522081	0.016362141 no
TCONS_L2_00017851		757	144	0.6203	-0.079415711	0.005701195 no
TCONS_L2_00017852		631	168	0.868	-0.07522081	0.016491637 no
TCONS_L2_00017853		359	93	0.9437	-0.087655495	0.008776281 no
TCONS_L2_00018290		1174	144	0.6051	-0.107657632	0.004403664 no
TCONS_L2_00018291		4136	267	0.42	-0.077650944	0.009430103 no
TCONS_L2_00018292		2102	276	0.484	-0.012649119	0.021810528 no
TCONS_L2_00018293		1394	327	0.6533	-0.197474682	0.019892986 no
TCONS_L2_00018294		413	144	0.6051	-0.107657632	0.004610957 no
TCONS_L2_00018295		871	102	0.8213	-0.089455015	0.006323578 no
TCONS_L2_00018296		551	258	0.6655	0.081717204	0.062360503 no

**PCAN-6 (XLOC\_I2\_013931)**

Sequence Name	RNA Size	ORF Size	Ficket Score	Hexamer Score	Coding Probability	Coding Label
TCONS_L2_00026766		1676	459	0.484	0.100992688	0.259495691 no
TCONS_L2_00026767		809	591	0.702	0.137974424	0.795810077 yes
TCONS_L2_00026768		376	252	1.047	0.176429194	0.280760562 no
TCONS_L2_00026769		453	153	1.0189	0.316429065	0.23511127 no
TCONS_L2_00026770		4110	396	0.618	-0.361689179	0.010864527 no
TCONS_L2_00026771		1215	267	1.1441	0.363689952	0.67084369 yes
TCONS_L2_00026773		1150	267	1.1441	0.363689952	0.671714583 yes
TCONS_L2_00026774		1091	267	1.1441	0.363689952	0.672504065 yes
TCONS_L2_00026775		829	267	1.1441	0.363689952	0.675998075 yes
TCONS_L2_00026776		930	327	1.2265	0.349269552	0.823711195 yes
TCONS_L2_00026777		1119	267	1.1441	0.363689952	0.672129517 yes
TCONS_L2_00026778		582	267	1.1441	0.363689952	0.679274184 yes
TCONS_L2_00026779		1167	267	1.1441	0.363689952	0.671486925 yes
TCONS_L2_00026780		659	135	0.3685	-0.40494765	0.00027646 no
TCONS_L2_00026781		400	156	0.6087	-0.431469675	0.000628962 no
TCONS_L2_00026782		255	246	0.4729	-0.16219555	0.006446424 no
TCONS_L2_00026783		287	204	0.9136	0.316628344	0.279341985 no
TCONS_L2_00026784		1549	210	1.0844	0.305066245	0.377734015 yes
TCONS_L2_00027441		929	426	0.4914	0.202855179	0.339790484 no

**Protein coding gene controls**

**GAPDH**

Sequence Name	RNA Size	ORF Size	Ficket Score	Hexamer Score	Coding Probability	Coding Label
ENST00000229239.5		1875	1008	1.2926	0.519960341	0.999962146 yes
ENST00000396861.1		1348	1008	1.2926	0.519960341	0.999963338 yes
ENST00000396858.1		1292	882	1.2952	0.533919114	0.999870952 yes
ENST00000396859.1		1256	1008	1.2926	0.519960341	0.999963542 yes
ENST00000396856.1		1266	783	1.315	0.548545855	0.999679465 yes

**GUSB**

Sequence Name	RNA Size	ORF Size	Ficket Score	Hexamer Score	Coding Probability	Coding Label
ENST00000304895.4		2300	339	1.2668	0.277977644	0.776289336 yes
ENST00000421103.1		1742	339	1.2668	0.277977644	0.782118123 yes
ENST00000345660.6		2027	285	1.1446	0.333726267	0.659495687 yes

Supplemental Table V part 2  
 Results from iSeeRNA: <http://137.189.133.71/iSeeRNA/>  
**PCNA-1 (XLOC\_002996)**

ID	C/NC	NONCODING SCORE
TCONS_00007074	noncoding	0.9837
TCONS_00007073	noncoding	0.996
TCONS_00006377	noncoding	0.9873
TCONS_00006378	noncoding	0.9828

**PCAN-2 (XLOC\_I2\_004121)**

ID	C/NC	NONCODING SCORE
TCONS_I2_00008295	noncoding	0.8142
TCONS_I2_00008296	noncoding	0.7723
TCONS_I2_00008290	noncoding	0.9384
TCONS_I2_00008289	noncoding	0.9267
TCONS_I2_00008292	noncoding	0.9812
TCONS_I2_00008294	noncoding	0.9769
TCONS_I2_00008291	noncoding	0.9582
TCONS_I2_00008293	noncoding	0.9681

**PCAN-3 (XLOC\_I2\_004340)**

ID	C/NC	NONCODING SCORE
TCONS_I2_00008348	noncoding	0.9764
TCONS_I2_00008347	noncoding	0.9819
TCONS_I2_00008343	noncoding	0.8167
TCONS_I2_00008346	noncoding	0.9733
TCONS_I2_00007970	noncoding	0.9729

**PCAN-4 (XLOC\_I2\_007509)**

ID	C/NC	NONCODING SCORE
TCONS_I2_00013919	coding	0.208638
TCONS_I2_00013920	noncoding	0.9405
TCONS_I2_00013918	noncoding	0.6351

**PCAN-5 (XLOC\_I2\_009441)**

ID	C/NC	NONCODING SCORE
TCONS_I2_00017847	noncoding	0.6722
TCONS_I2_00017848	noncoding	0.9813
TCONS_I2_00017849	noncoding	0.9852
TCONS_I2_00017850	noncoding	0.9341
TCONS_I2_00017851	noncoding	0.9674
TCONS_I2_00017852	noncoding	0.8795
TCONS_I2_00017853	noncoding	0.984
TCONS_I2_00018290	noncoding	0.8125
TCONS_I2_00018291	noncoding	0.9676
TCONS_I2_00018292	noncoding	0.9702
TCONS_I2_00018293	noncoding	0.9803
TCONS_I2_00018294	noncoding	0.9135
TCONS_I2_00018295	noncoding	0.9988
TCONS_I2_00018296	noncoding	0.9808

**PCAN-6 (XLOC\_I2\_013931)**

ID	C/NC	NONCODING SCORE
TCONS_I2_00026774	noncoding	0.6841
TCONS_I2_00026769	coding	0.0722
TCONS_I2_00026766	coding	0.2569
TCONS_I2_00026779	noncoding	0.8778
TCONS_I2_00026780	noncoding	0.8494
TCONS_I2_00026781	noncoding	0.9237
TCONS_I2_00026767	noncoding	0.7907
TCONS_I2_00026771	noncoding	0.872
TCONS_I2_00026778	noncoding	0.6522
TCONS_I2_00026776	noncoding	0.7908
TCONS_I2_00026777	noncoding	0.605
TCONS_I2_00027441	coding	0.141
TCONS_I2_00026784	noncoding	0.8035
TCONS_I2_00026768	noncoding	0.8508
TCONS_I2_00026773	noncoding	0.8681
TCONS_I2_00026770	noncoding	0.6034
TCONS_I2_00026783	noncoding	0.7266
TCONS_I2_00026782	noncoding	0.5388
TCONS_I2_00026775	noncoding	0.8405

**Protein coding gene controls**

**GAPDH**

ID	C/NC	NONCODING SCORE
ENST00000229239.5	coding	0
ENST00000396861.1	coding	0
ENST00000396858.1	coding	0.002
ENST00000396859.1	coding	0
ENST00000396856.1	coding	0.0038

**GUSB**

ID	C/NC	NONCODING SCORE
ENST00000304895.4	coding	0
ENST00000421103.1	coding	0
ENST00000345660.6	coding	0

Table S6

<b>Area under curve</b>							
	TCGA_training	TCGA_testing	GSE58135_breast_cancer	GSE50760_colon_cancer	GSE25599_liver_cancer	Our BRCA dataset	
GSVM	0.946	0.939	0.545	0.827	0.910	0.750	
LSVM	0.918	0.917	0.968	0.796	0.885	1.000	
Random Forest	0.947	0.942	0.972	0.841	0.970	0.950	
Logistic Regression	0.901	0.905	0.646	0.410	0.390	0.550	
<b>Max accuracy</b>							
	TCGA_training	TCGA_testing	GSE58135_breast_cancer	GSE50760_colon_cancer	GSE25599_liver_cancer	Our BRCA dataset	
GSVM	0.894	0.903	0.743	0.861	0.850	0.773	
LSVM	0.856	0.851	0.921	0.833	0.850	1.000	
Random Forest	0.887	0.895	0.936	0.861	0.900	0.909	
Logistic Regression	0.828	0.831	0.679	0.639	0.600	0.636	
<b>F-score</b>							
	TCGA_training	TCGA_testing	GSE58135_breast_cancer	GSE50760_colon_cancer	GSE25599_liver_cancer	Our BRCA dataset	
GSVM	0.788	0.808	0.489	0.732	0.734	0.624	
LSVM	0.712	0.708	0.836	0.684	0.734	1.000	
Random Forest	0.775	0.790	0.873	0.732	0.816	0.817	
Logistic Regression	0.658	0.665	0.418	0.302	0.250	0.346	
<b>Matthew's correlation coefficient</b>							
	TCGA_training	TCGA_testing	GSE58135_breast_cancer	GSE50760_colon_cancer	GSE25599_liver_cancer	Our BRCA dataset	
GSVM	0.894	0.897	0.822	0.848	0.870	0.800	
LSVM	0.855	0.843	0.936	0.824	0.842	1.000	
Random Forest	0.888	0.890	0.944	0.848	0.909	0.900	
Logistic Regression	0.826	0.835	0.750	0.667	0.667	0.645	