

LncmiRSRN: identification and analysis of long non-coding RNA related miRNA sponge regulatory network in human cancer

Junpeng Zhang^{1,*}, Lin Liu², Jiuyong Li² and Thuc Duy Le^{2,*}

¹School of Engineering, Dali University, Dali, Yunnan, 671003, P. R. China.

²School of Information Technology and Mathematical Sciences, University of South Australia, Mawson Lakes, SA 5095, Australia.

*To whom correspondence should be addressed.

Contact: zhangjunpeng_411@yahoo.com, Thuc.Le@unisa.edu.au

In this file, we provide supplementary materials discussed in ‘Results’ section.

1. Cancer related genes of conserved sponge lncRNA-mRNA regulatory relationships in four human cancers

We collect a list of lncRNAs related to the four human cancers from LncRNADisease v2015 (Chen, *et al.*, 2013), Lnc2Cancer v2016 (Ning, *et al.*, 2016) and MNDR v2013 (Wang, *et al.*, 2013). The list of mRNAs associated with the four human cancers is from DisGeNET v4.0 (Piñero, *et al.*, 2017). The cancer lncRNAs and mRNAs associated with at least one of the four cancers are shown in Table S1.

Table S1. Cancer related genes including lncRNAs and mRNAs.

Cancer genes		GBM	LSCC	OvCa	PrCa
lncRNAs	CTB-89H12.4				√
	GAS5		√	√	√
	MEG3		√		√
	NEAT1			√	√
	TUG1	√			√
mRNAs	ADAM10	√		√	√
	APC	√	√	√	√
	ATM	√	√	√	√
	AURKB	√	√	√	√
	BRCA1	√		√	√
	CDC25A			√	√
	CDC42	√	√	√	√

CTNND1		√	√	√
DICER1		√	√	
DNMT3A	√	√	√	√
FAS	√	√	√	√
FOXO3	√		√	√
GAS5				√
LATS2		√	√	√
MCL1	√	√	√	√
MYB	√			√
NCOA3			√	√
NEAT1				√
NOTCH2	√	√		
PODXL	√			√
PTEN	√	√	√	√
PURA				√
RAB5A			√	
RAF1	√	√	√	√
RPS6KB1	√	√	√	√
SETD2	√		√	√
SIRT1	√	√	√	√
SMAD1		√	√	√
SP1	√	√	√	√
STAT3	√	√	√	√
TDG		√		
TET2				√
TIMP3	√	√	√	√
TNFAIP3	√	√		√
ZEB2			√	√

2. Functional enrichment analysis of differential and conserved LncmiRSRN network modules

We use the R-package *clusterProfiler* (Yu *et al.*, 2012) to conduct functional enrichment analysis. We are interested in the Gene Ontology (GO) (Ashburner, *et al.*, 2000) biological processes and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) pathways with adjusted p -value < 0.05 (adjusted by Benjamini-Hochberg method). The differential and conserved LncmiRSRN network modules associated with at least one GO biological process or KEGG pathway are listed in Table S2.

Table S2. Functional enrichment analysis of differential and conserved LncmiRSRN network modules between GBM, LSCC, OvCa and PrCa.

Modules	Module ID	#Genes	#GO Terms	#KEGG Terms
Differential	1	5	325	0
	3	23	120	2
	4	6	43	0
	5	5	43	0
	6	10	81	2
	7	23	227	32
	8	5	165	4
	9	11	69	1
	10	4	52	0
	11	13	15	1
	12	9	158	2
	13	12	48	1
	14	28	153	0
	15	14	45	2
	16	8	41	0
	17	23	94	7
	18	14	81	0
	19	4	53	0
	20	5	63	0
	21	5	406	0
	23	28	38	2
	24	17	56	3
	25	27	113	0
	27	11	11	5
	28	19	22	0
	29	6	520	6
	30	18	270	0
	31	5	75	1
	32	11	228	0
	33	18	110	8
	34	4	312	1
	35	34	173	3
	36	6	205	15
	37	4	358	0
	38	12	173	6
	39	8	69	12
	40	9	489	3
	41	9	174	1
	42	15	18	0
	43	10	723	30

	44	5	373	2
	45	12	188	0
	46	6	349	7
	47	11	256	4
	48	5	358	0
	49	4	261	9
	50	7	20	0
	51	9	491	1
	52	4	404	0
	53	8	136	0
	54	6	17	2
	55	6	159	8
Conserved	1	5	34	13
	2	11	4	3
	3	10	0	1
	5	6	227	22
	6	10	10	1
	7	5	14	1
	8	6	255	0
	9	7	115	0
	10	4	136	3
	11	40	181	23
	12	5	49	1
	13	5	251	1
	14	9	71	2
	15	4	39	2
	16	6	78	2
	17	5	117	0
	18	7	230	0
	19	5	90	4
	20	6	332	20
	21	6	160	23
	22	4	61	0
	23	5	54	0
	24	4	163	0
	25	4	71	4
	26	5	0	2
	27	4	246	1
	28	4	437	2
	29	4	249	25

The adjusted p -value cutoff is set to 0.05 (adjusted by Benjamini-Hochberg method).

3. Cancer gene enrichment of differential and conserved

LncmiRSRN network modules

Based on existing gene-disease associations, the differential and conserved LncmiRSRN network modules related to at least one cancer are listed in Table S3.

Table S3. Cancer gene enrichment of differential and conserved LncmiRSRN network modules between GBM, LSCC, OvCa and PrCa.

Modules	Module ID	#GBM	#LSCC	#OvCa	#PrCa
Differential	1	1	2	2	3
	2	0	1	1	1
	3	11	11	12	13
	4	1	0	1	1
	5	0	0	2	0
	6	2	1	2	3
	7	10	11	13	12
	8	1	3	3	3
	9	3	1	3	3
	10	1	0	1	1
	11	2	2	3	3
	12	6	4	5	6
	13	1	1	1	1
	14	9	7	8	9
	15	4	3	4	4
	16	2	1	1	4
	17	6	9	11	13
	18	2	3	2	3
	19	1	2	0	0
	20	1	0	1	0
	21	3	3	3	3
	22	8	8	8	9
	23	10	8	12	13
	24	5	7	7	10
	25	7	9	12	16
	26	4	3	6	7
	27	2	3	5	4
	28	5	3	6	9
	29	2	3	2	2
	30	4	3	3	3
	31	1	0	1	1
	32	2	3	2	2

	33	1	2	2	1
	34	2	1	2	2
	35	5	9	7	9
	36	2	0	3	3
	37	1	2	2	2
	38	4	4	4	4
	39	1	0	1	0
	40	2	2	2	3
	41	1	1	1	2
	42	2	3	3	3
	43	3	3	3	3
	44	1	2	2	3
	45	3	3	5	4
	46	1	2	2	3
	47	3	3	4	4
	48	2	2	2	2
	49	1	2	2	2
	50	0	0	1	0
	51	2	3	3	3
	52	2	2	2	2
	53	2	1	1	2
	54	1	1	2	1
	55	2	1	1	2
Conserved	1	2	0	0	2
	2	5	3	1	5
	3	1	3	5	3
	4	1	1	2	1
	5	1	1	2	3
	6	4	5	5	4
	7	2	2	2	2
	8	2	1	2	2
	9	3	2	3	3
	10	1	0	0	1
	11	11	10	14	20
	12	2	1	2	1
	13	2	2	2	2
	14	1	1	2	1
	15	1	0	1	2
	16	1	1	1	1
	17	1	1	2	1
	18	1	0	1	2
	19	0	1	2	2
	20	2	2	2	2
	21	1	2	2	2

22	1	1	0	1
23	1	1	1	1
24	2	1	2	1
25	0	0	1	1
26	2	2	2	2
27	2	1	1	2
28	2	1	2	2
29	2	2	2	2

4. Survival analysis of differential and conserved LncmiRSRN network modules

We perform survival analysis of differential and conserved LncmiRSRN network modules using R-packages *survival* (Therneau and Lumley, 2017) and *survcomp* (Schröder *et al.*, 2011). We calculate Hazard Ratio (HR) between the high and the low risk groups, and the Log-rank test as well as the Kaplan Meier curve is also performed. In GBM, LSCC, OvCa and PrCa datasets, the differential and conserved LncmiRSRN network modules acting as module biomarkers are listed in Table S4-S7, respectively.

Table S4. Survival analysis of differential and conserved LncmiRSRN network modules in GBM dataset.

Modules	Module ID	Hazard Ratio (95% CI)	Log-rank <i>p</i> -value
Differential	17	1.80 (1.46-2.22)	5.08E-09
	22	1.63 (1.32-2.01)	1.28E-06
	23	1.83 (1.48-2.26)	9.12E-10
	24	1.54 (1.25-1.89)	2.25E-05
	25	1.96 (1.58-2.42)	1.38E-11
	26	1.62 (1.31-2.00)	1.61E-06
	28	1.71 (1.39-2.12)	5.78E-08
	30	1.64 (1.33-2.03)	8.37E-07
	33	1.92 (1.56-2.38)	4.11E-11
	35	1.66 (1.35-2.04)	5.30E-07
	38	1.51 (1.22-1.85)	6.06E-05
	45	1.59 (1.29-1.96)	4.41E-06
	46	1.74 (1.41-2.14)	2.88E-08
	50	1.58 (1.28-1.94)	6.69E-06
	Conserved	11	1.93 (1.56-2.39)
13		1.54 (1.25-1.89)	2.47E-05
23		1.55 (1.26-1.91)	1.31E-05

The Hazard Ratio is at least 1.5 with Log-rank p -value < 0.05 .

Table S5. Survival analysis of differential and conserved LncmiRSRN network modules in LSCC dataset.

Modules	Module ID	Hazard Ratio (95% CI)	Log-rank p -value
Differential	17	3.71 (1.94-7.07)	2.22E-05
	18	2.03 (1.09-3.77)	2.92E-02
	20	1.88 (1.01-3.50)	4.58E-02
	21	3.46 (1.84-6.51)	9.30E-05
	22	3.49 (1.88-6.50)	2.56E-04
	23	5.21 (2.74-9.93)	2.90E-07
	24	3.37 (1.81-6.30)	2.60E-04
	25	3.14 (1.69-5.85)	5.62E-04
	26	2.21 (1.19-4.10)	1.55E-02
	27	2.11 (1.13-3.96)	1.70E-02
	28	4.03 (2.09-7.74)	4.77E-06
	29	1.90 (1.02-3.53)	4.54E-02
	30	1.95 (1.04-3.65)	3.24E-02
	31	2.12 (1.13-3.97)	0.016934
	33	3.35 (1.77-6.33)	8.14E-05
	34	2.28 (1.21-4.29)	8.54E-03
	35	3.16 (1.69-5.91)	4.01E-04
	36	1.99 (1.07-3.70)	3.36E-02
	39	2.00 (1.07-3.77)	2.60E-02
	40	1.91 (1.02-3.59)	3.74E-02
	42	3.92 (2.07-7.44)	1.15E-05
	43	2.92 (1.56-5.48)	8.39E-04
	45	2.32 (1.24-4.32)	8.69E-03
	46	1.96 (1.05-3.65)	3.54E-02
	47	3.74 (1.99-7.01)	5.30E-05
	52	2.44 (1.31-4.54)	6.27E-03
	53	2.30 (1.24-4.30)	9.30E-03
55	2.29 (1.22-4.32)	7.91E-03	
Conserved	1	2.27 (1.21-4.28)	8.12E-03
	2	2.27 (1.21-4.25)	8.87E-03
	3	3.20 (1.70-6.04)	1.96E-04
	4	2.98 (1.60-5.57)	7.92E-04
	8	2.63 (1.40-4.93)	2.36E-03
	10	2.84 (1.51-5.35)	8.18E-04
	11	6.04 (3.20-11.37)	6.65E-08
	13	2.04 (1.08-3.83)	2.09E-02
	17	2.23 (1.19-4.18)	1.13E-02
	20	3.59 (1.92-6.71)	1.04E-04
21	2.30 (1.23-4.31)	8.16E-03	

23	1.87 (1.00-3.48)	4.98E-02
26	2.85 (1.51-5.36)	7.95E-04
27	2.19 (1.17-4.09)	1.36E-02

The Hazard Ratio is at least 1.5 with Log-rank p -value < 0.05 .

Table S6. Survival analysis of differential and conserved LncmiRSRN network modules in OvCa dataset.

Modules	Module ID	Hazard Ratio (95% CI)	Log-rank p -value
Differential	17	1.72 (1.37-2.16)	1.65E-06
	18	1.80 (1.43-2.26)	2.02E-07
	23	1.88 (1.50-2.36)	1.89E-08
	25	1.82 (1.45-2.29)	8.24E-08
	26	1.66 (1.33-2.08)	7.81E-06
	27	1.56 (1.24-1.95)	1.10E-04
	28	1.56 (1.24-1.96)	1.04E-04
	30	1.55 (1.24-1.95)	1.19E-04
	33	1.70 (1.35-2.13)	3.02E-06
	36	1.51 (1.21-1.90)	3.01E-04
	38	1.65 (1.32-2.07)	1.08E-05
	40	1.72 (1.37-2.16)	1.75E-06
	42	1.69 (1.35-2.12)	4.12E-06
Conserved	11	1.84 (1.47-2.31)	9.16E-08
	21	1.64 (1.31-2.05)	1.72E-05

The Hazard Ratio is at least 1.5 with Log-rank p -value < 0.05 .

Table S7. Survival analysis of differential and conserved LncmiRSRN network modules in PrCa dataset.

Modules	Module ID	Hazard Ratio (95% CI)	Log-rank p -value
Differential	16	3.91 (1.09-14.03)	1.06E-02
	18	6.52 (1.95-21.84)	2.53E-02
	20	10.72 (3.29-34.98)	2.79E-03
	24	5.89 (1.79-19.39)	4.59E-03
	25	9.84 (2.75-35.14)	8.66E-05
	26	4.25 (1.26-14.31)	1.54E-02
	27	3.29 (1.00-10.80)	4.78E-02
	30	4.46 (1.37-14.53)	1.67E-02
	39	6.03 (1.83-19.93)	5.77E-03
	43	7.88 (2.40-25.90)	1.43E-02
	45	6.50 (1.95-21.61)	2.56E-03
	49	3.46 (1.05-11.41)	3.96E-02
	50	3.54 (1.01-12.42)	1.79E-02
	53	7.21 (2.18-23.89)	2.14E-02

	54	5.00 (1.53-16.34)	1.35E-02
Conserved	5	3.86 (1.08-13.82)	1.34E-02
	6	9.26 (2.84-30.23)	6.06E-03
	7	4.18 (1.24-14.05)	1.68E-02
	11	13.50 (4.09-44.61)	6.95E-04
	12	3.30 (1.01-10.85)	4.85E-02
	14	4.45 (1.36-14.51)	2.40E-02
	17	3.44 (1.05-11.33)	4.12E-02
	18	3.30 (1.01-10.84)	4.87E-02
	20	7.42 (2.25-24.53)	1.88E-02
	21	4.11 (1.13-14.96)	1.08E-02

The Hazard Ratio is at least 1.5 with Log-rank p -value < 0.05.

References

- Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. *Nat Genet.*, **25**, 25-9.
- Chen, G. *et al.* (2013) LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.*, **41**, D983-6.
- Csardi, G. and Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal, Complex Systems*. **1695**,1-9.
- Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27-30.
- Ning, S. *et al.* (2016) Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. *Nucleic Acids Res.*, **44**, D980-5.
- Piñero, J. *et al.* (2017) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.*, **45**, D833-D839.
- Schröder, M.S. *et al.* (2011) survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics*. **27**, 3206-8.
- Therneau, T.M. and Lumley, T. (2017) Survival Analysis. R package version 2.41-3.
- Wang, Y. *et al.* (2013) Mammalian ncRNA-disease repository: a global view of ncRNA-mediated disease network. *Cell Death Dis.*, **4**, e765.
- Yu, G. *et al.* (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. **16**, 284-7.