Supplemental Material

Systematic Characterization of A-to-I RNA Editing Hotspots in MicroRNAs across Human Cancers

Table of contents:

Supplemental Figures

Supplemental Fig S1. Over expression of ADAR1 and ADAR2 in Hs578T, 786O and HeyA8 cell lines **2**

Supplemental Fig S2. The editing levels of 15 validated A-to-I editing hotspots in 20 cancer types **3**

Supplemental Fig S3. The correlations of miRNA editing level with WT miRNA expression across different cancer types **4**

Supplemental Fig S4. The correlations of miRNA editing levels with significantly mutated genes and copy number alterations in different cancer types **5**

Supplemental Fig S5. Correlation of ADAR mRNA expression with patient survival times **6**

Supplemental Fig S6. Quantitative assessment of transfected miR-200b expression amount in cell lines **7**

Supplemental Fig S7. Effects of miR-200b mimics on cell viability in MCF10A, MDAMB-231 and SLR25 cells **8**

Supplemental Fig S8. Effect of wild-type miR-200b on target gene ZEB2 ..... **9**

Supplemental Fig S9. Correlations of LIFR with patient survival times in different cancer types **10**

Supplemental Fig S10. Correlation between edited miR-200b expression level and LIFR expression across cancer types **11**

Supplemental Tables

Supplemental Table S1. Overview of TCGA miRNA sequencing data **12**

Supplemental Table S2. Summary of high-confidence RNA editing events and RNA editing hotspots **13**

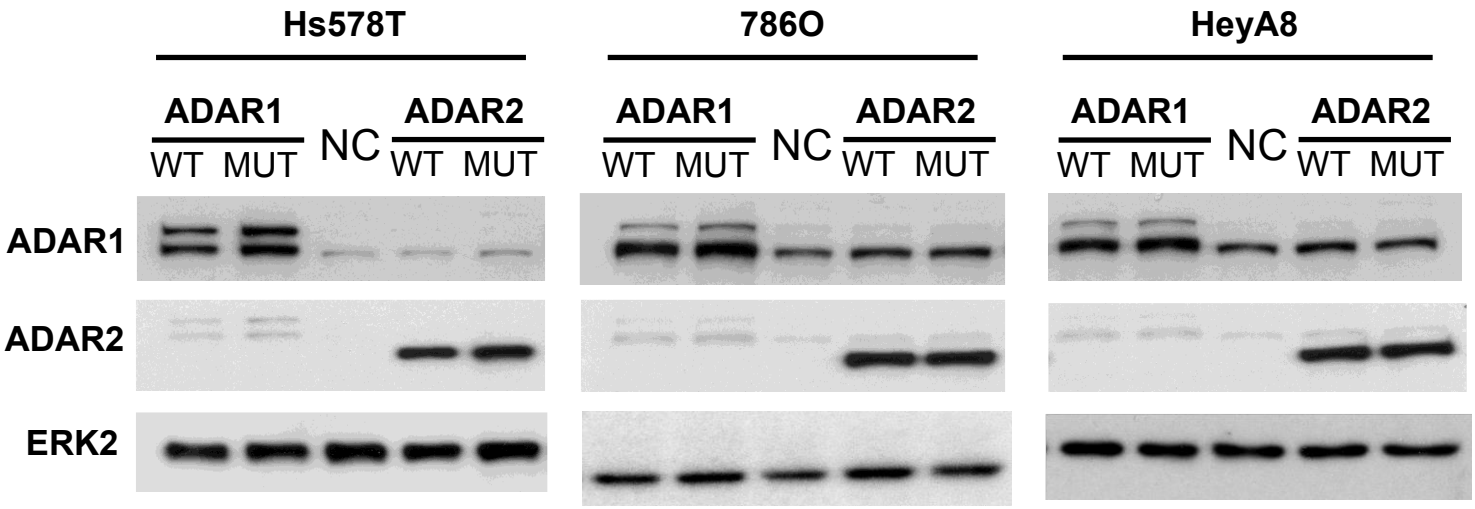
Supplemental Table S3. Information about 19 A-to-I RNA editing hotspots identified in miRNAs **15**

Supplemental Table S4. Validation of the correlations of miRNA editing level with ADAR2 enzymes using a miRNA-seq dataset of ADAR2-perturbed cell lines **16**

Supplemental Table S5. List of potential new targets of edited miR-200b **17**

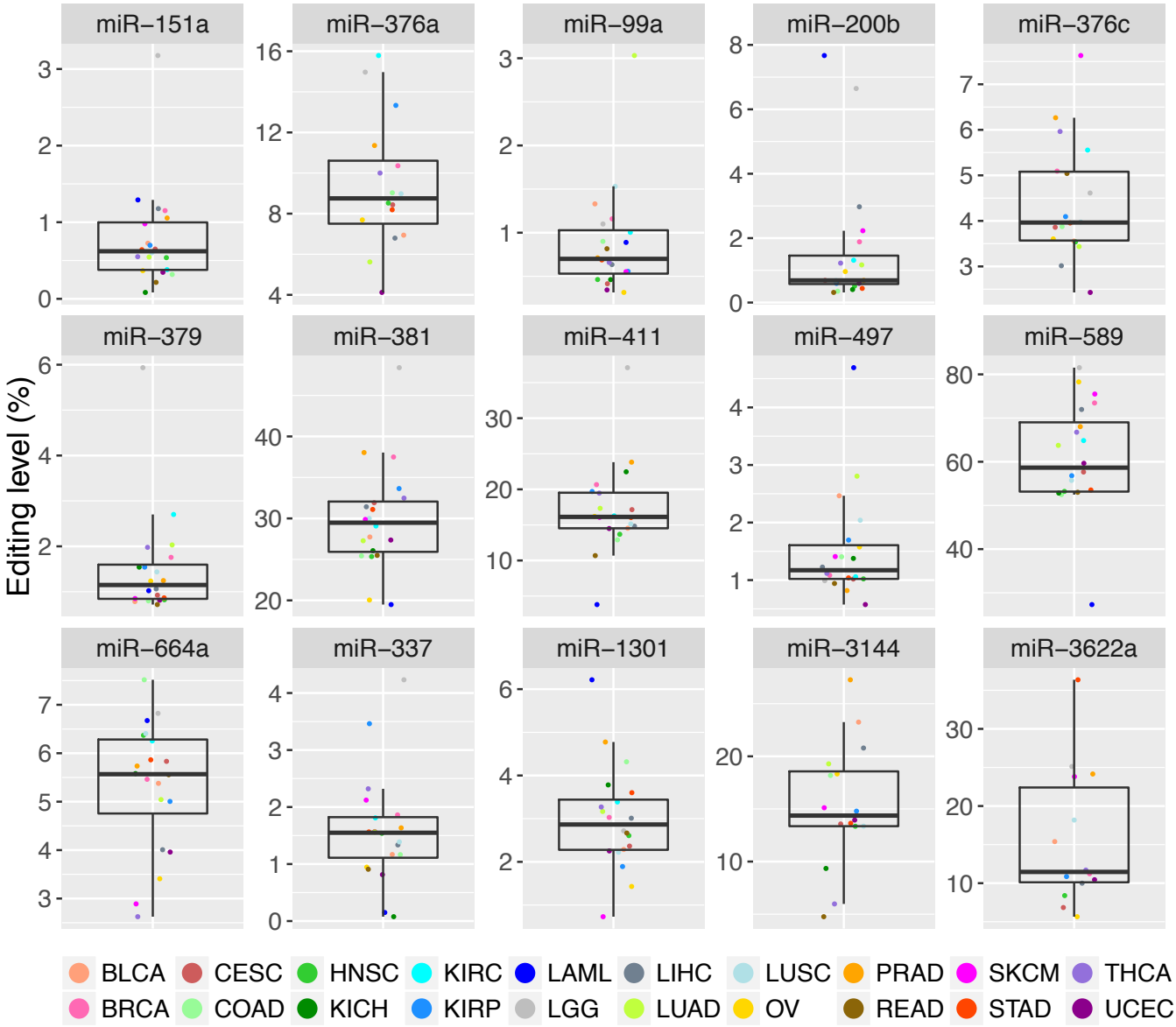
Supplemental Table S6. List of 3'UTR binding sites of wild-type and edited miR-200b to their target genes ZEB1, ZEB2, and LIFR **18**

Supplemental Fig S1



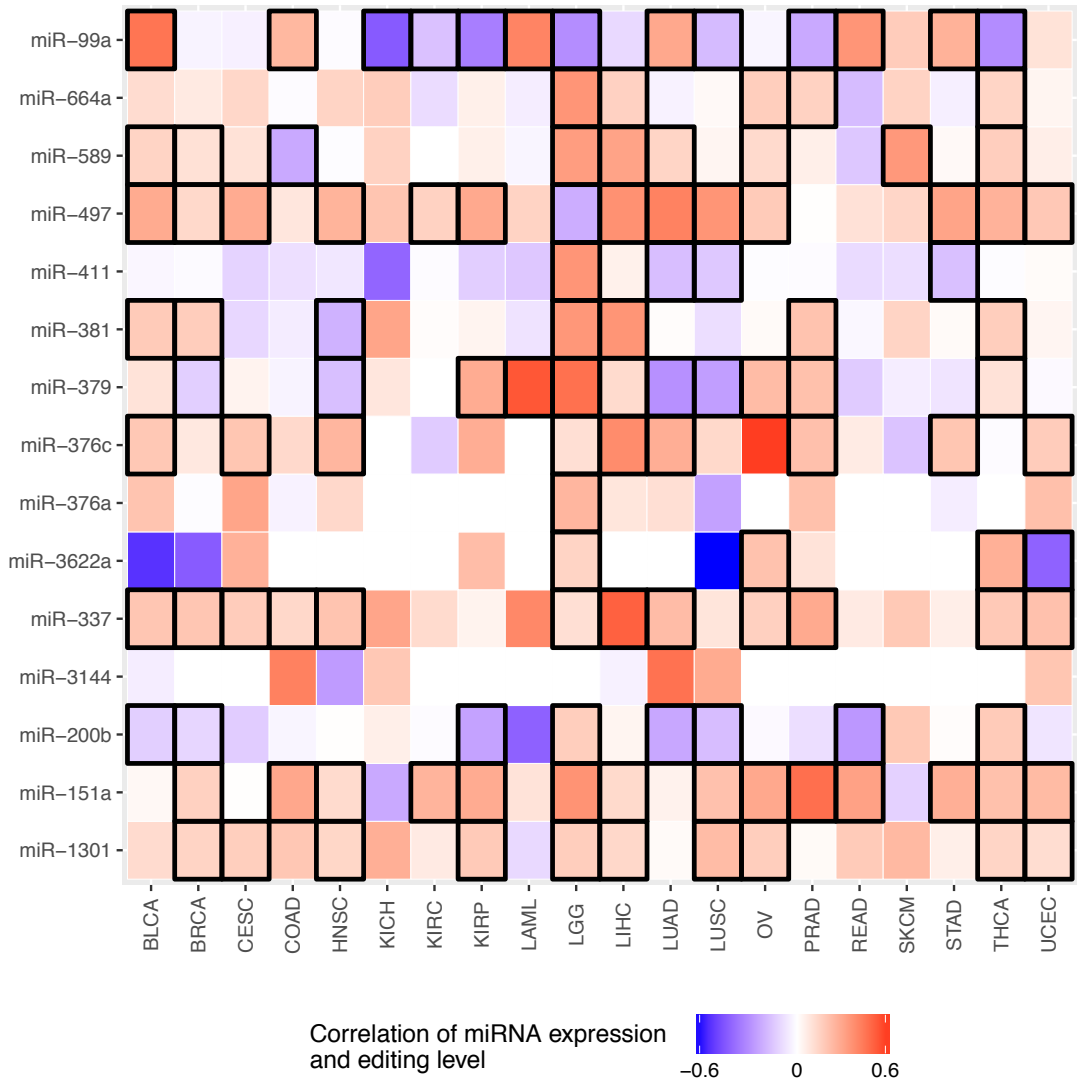
Supplemental Fig S1. Overexpression of ADAR1 and ADAR2 in 3 cell lines.
Western blot of ADAR enzymes overexpression in Hs578T, 786O and HeyA8 cell lines (NC: negative control, WT: wild-type, MUT: mutated).

Supplemental Fig S2



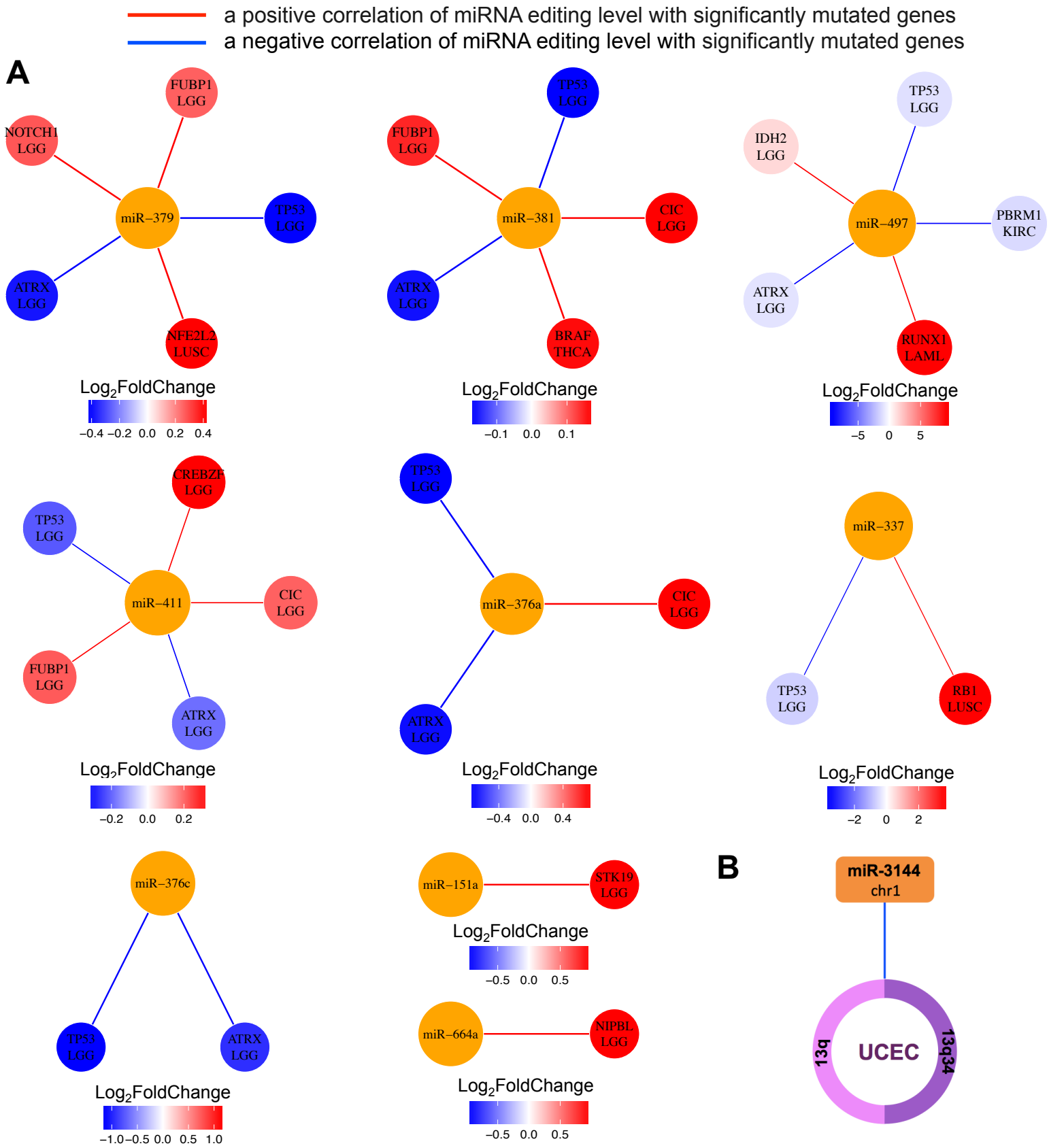
Supplemental Fig S2. The editing levels of 15 validated A-to-I editing hotspots in 20 cancer types.
Each box plot represents the average editing level of a miRNA editing hotspot across the edited samples within a cancer type.

Supplemental Fig S3



Supplemental Fig S3. The correlations of miRNA editing level with WT miRNA expression across different cancer types.
Color depicts the correlation coefficient, and black box highlights correlations that are significant (*FDR* < 0.05).

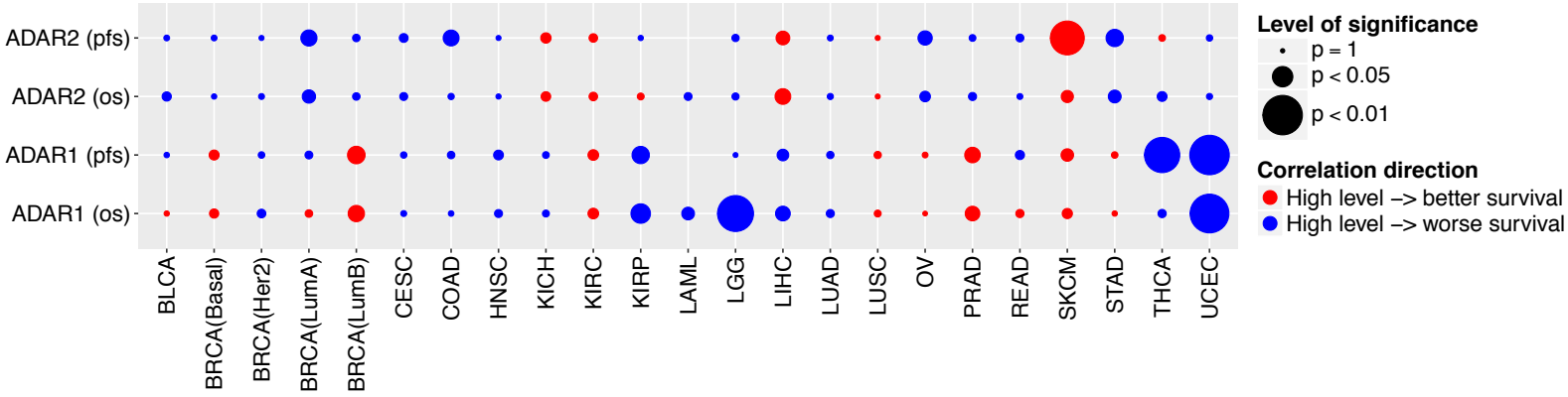
Supplemental Fig S4



Supplemental Fig S4. The correlations of miRNA editing levels with significantly mutated genes and copy number alterations in different cancer types

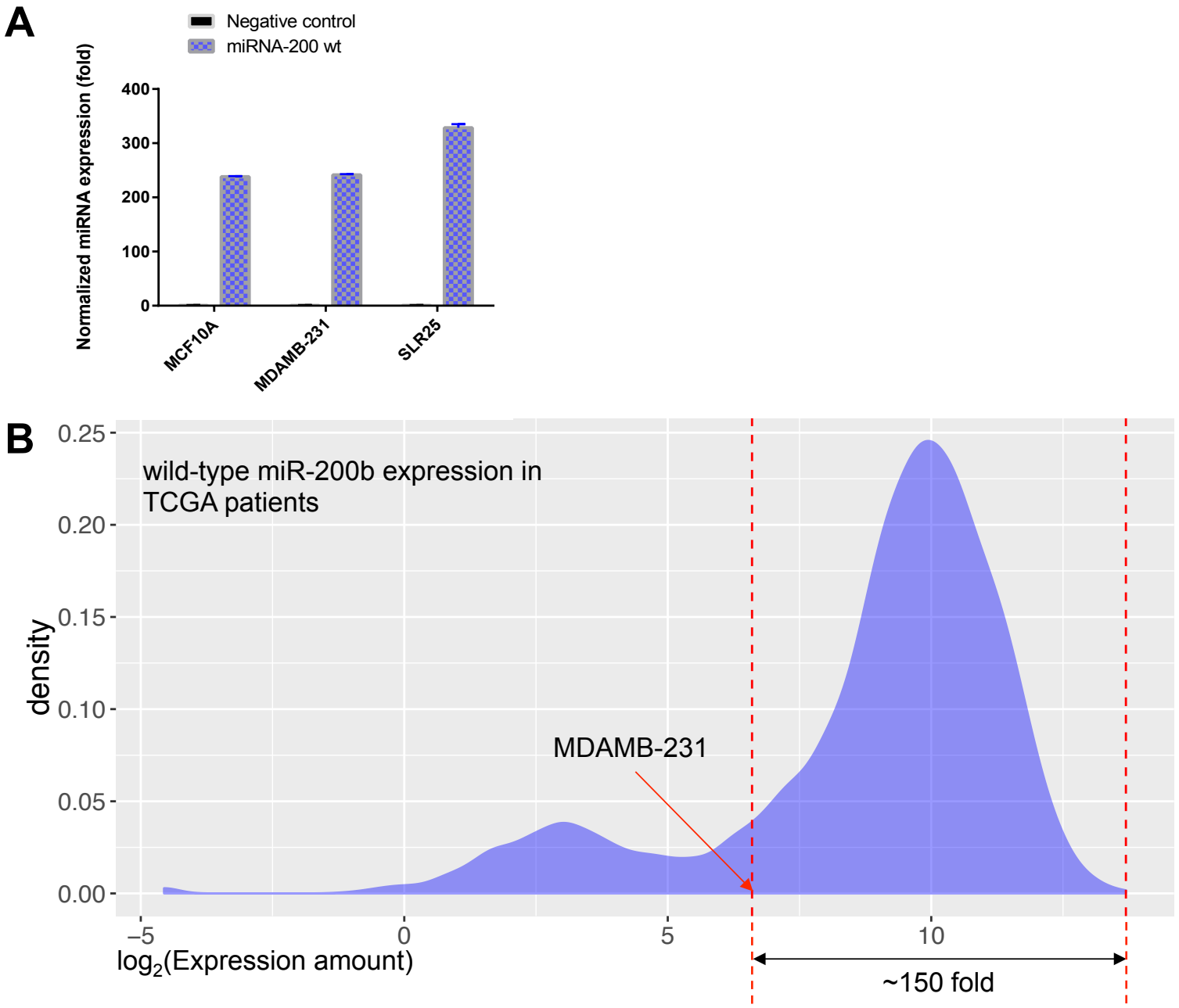
(A) miRNAs are depicted in orange, and mutated genes are in red or blue. Given a mutated gene, the edge color reflects the fold-change direction of the mean miRNA editing level in the mutated sample group relative to the wild-type sample group: red, increase; and blue, decrease. Color of each gene node reflects the fold change (log transformed). The specific cancer types showing the significant correlations are listed below gene names. (B) correlations of miRNA editing hotspots with frequent SCNAs. Red for positive correlations and blue for negative correlations.

Supplemental Fig S5



Supplemental Fig S5. Correlations of ADAR mRNA expression with patient survival times
Dot size depicts log-rank P value, and color represents the correlation direction (os = overall survival; pfs = progression free survival).

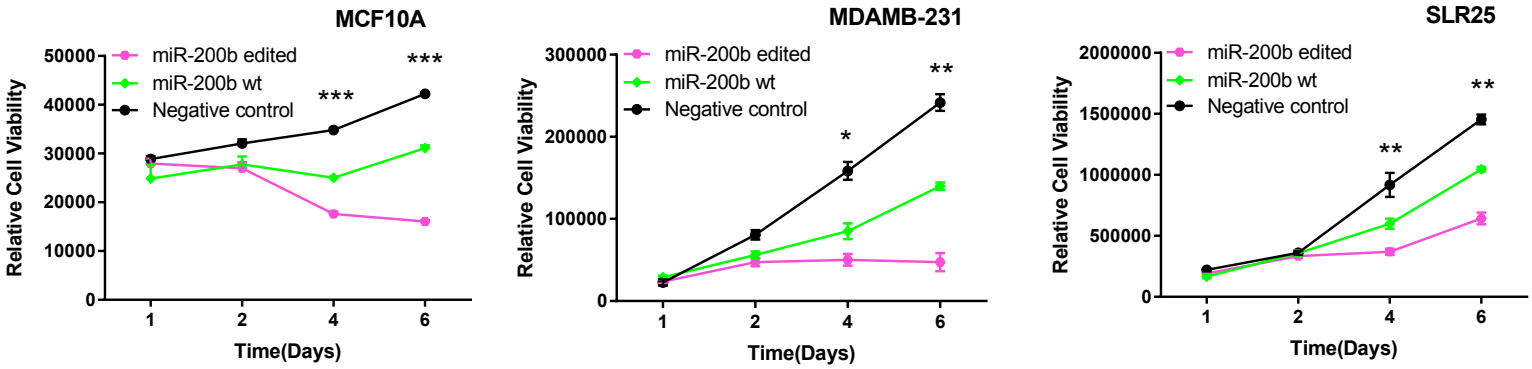
Supplemental Fig S6



Supplemental Fig S6. Quantitative assessment of transfected miR-200b expression amount in cell lines

(A) The over-expression amounts of miR-200b upon 24-hr transfection with 50nm wide-type miR-200b mimics in MCF10A, MDAMB-231 and SLR25 cell lines by qRT-PCR. The expression amounts transfected with negative controls were set as 1. (B) The expression amount distribution of wide-type miR-200b (RPM) in all TCGA patient cancer samples. The expression amount of miR-200b in MDAMB-231 was calculated based on a miRNA sequencing dataset (NCBI SRA SRX004030) using the same analytic procedure. The arrow indicates the fold increase of miR-200b expression in MDAMB-231 relative to the highest expression level observed among patients.

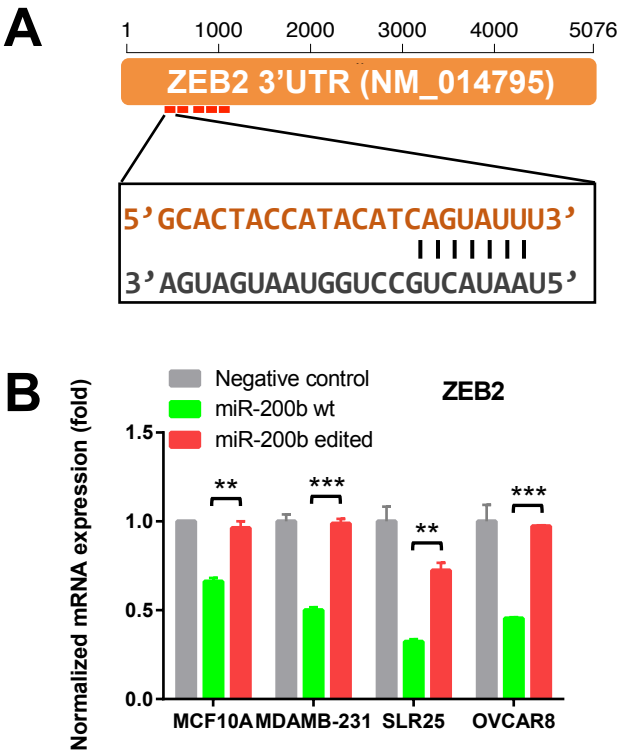
Supplemental Fig S7



Supplemental Fig S7. Effects of miR-200b mimics on cell viability in MCF10A, MDAMB-231 and SLR25 cells

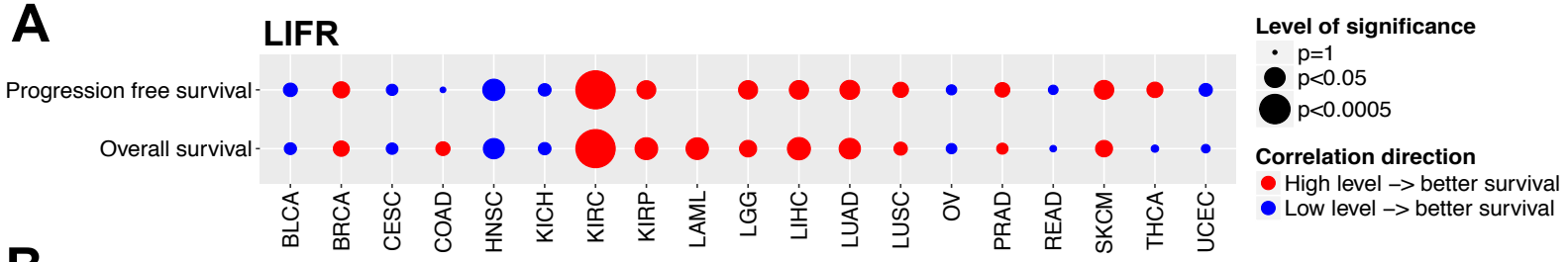
Two-sided *t*-test was used to assess the difference. Error bars denote +/- SEM; **p* < 0.05; ***p* <0.01; ****p* < 0.001.

Supplemental Fig S8

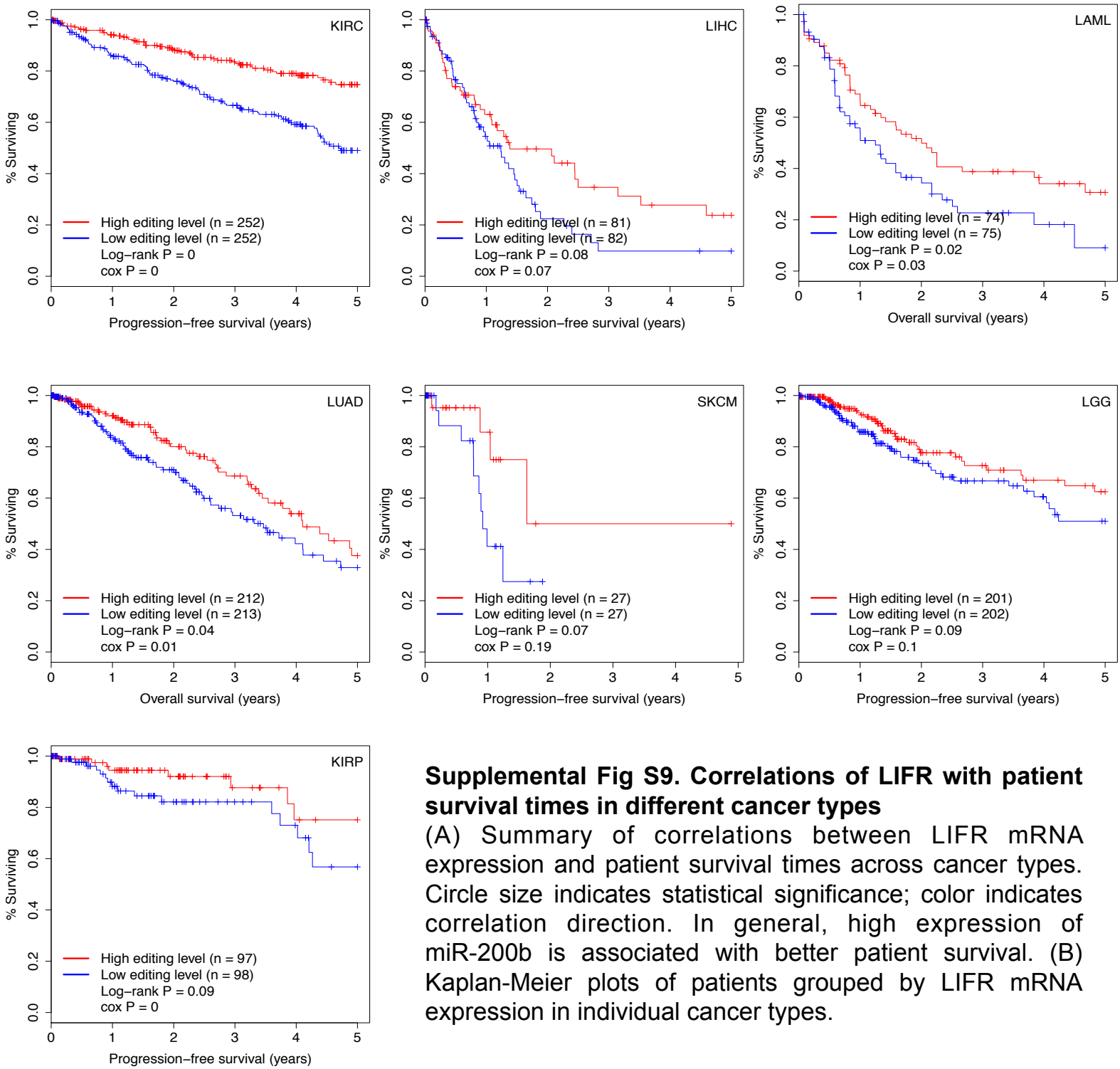


Supplemental Fig S8. Effect of wild-type miR-200b on target gene ZEB2
(A) 3′ UTR representations of WT miR-200b target genes ZEB1; (B) qRT-PCR of ZEB2 upon 24-hr transfection with WT miR-200b mimics in MCF10A, MDAMB-231 and SLR25 cells. Two-sided *t*-test was used to assess the difference. Error bars denote +/- SEM; *p < 0.05; **p < 0.01; ***p < 0.001.

Supplemental Fig S9



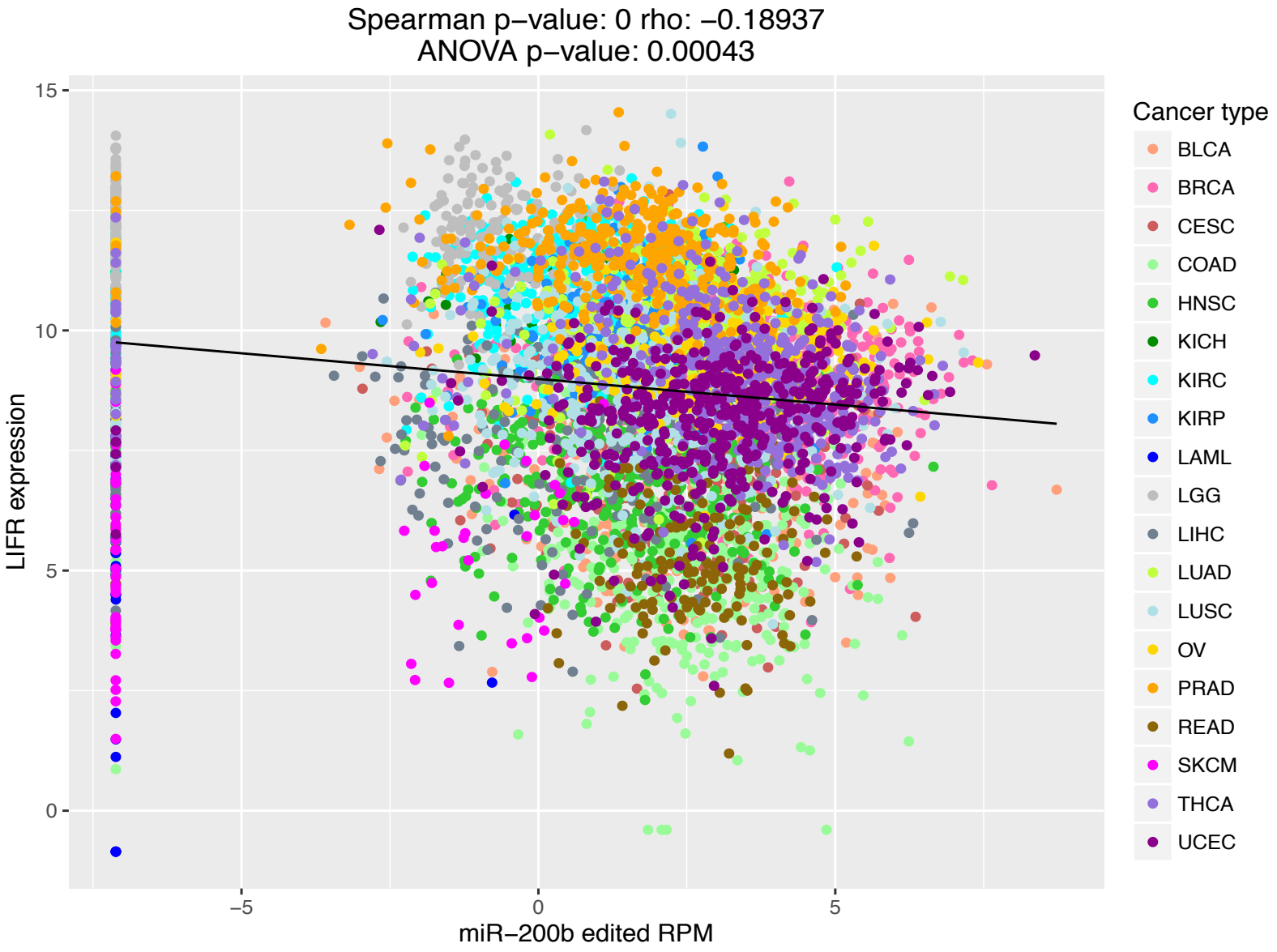
B



Supplemental Fig S9. Correlations of LIFR with patient survival times in different cancer types

(A) Summary of correlations between LIFR mRNA expression and patient survival times across cancer types. Circle size indicates statistical significance; color indicates correlation direction. In general, high expression of miR-200b is associated with better patient survival. (B) Kaplan-Meier plots of patients grouped by LIFR mRNA expression in individual cancer types.

Supplemental Fig S10



Supplemental Fig S10. Correlation between edited miR-200b expression level and LIFR expression across cancer types

We used edited miR-200b expression (log_2RPM) and the corresponding LIFR mRNA expression level (log_2RSEM) to evaluate their correlation. Each dot represents a sample; different colors depict different cancer types. A significant negative correlation was detected by both Spearman rank correlation (*R*s = -0.189, *p* = 0); and two-factor ANOVA test, with the cancer type being the fixed effect and miR-200b being an independent variable ($p = 4.3\times10^{-4}$).

Supplemental Table S1. Overview of TCGA miRNA sequencing data.

TCGA code	Cancer type	# Sample	Tumor sample	Normal sample	Average tumor mappable reads (millions)	Average normal mappable reads (millions)	# Confident A-to-I editing events
BLCA	**Bladder**	414	395	19	5.96 ± 3.85	15.39 ± 9.99	929
BRCA	**Breast**	890	801	89	3.76 ± 2.74	3.80 ± 2.49	2021
CESC	**Cervical**	301	298	3	5.33 ± 2.46	15.3 ± 1.74	682
COAD	**Colon**	399	391	8	4.48 ± 3.82	1.22 ± 0.36	883
HNSC	**Head and neck**	562	518	44	5.05 ± 2.34	6.34 ± 2.13	1189
KICH	**Kidney (chromophobe)**	90	65	25	6.35 ± 1.60	8.08 ± 2.25	103
KIRC	**Kidney (clear)**	587	516	71	3.57 ± 2.41	3.73 ± 1.39	1032
KIRP	**Kidney (papillary)**	325	291	34	6.74 ± 2.94	9.00 ± 2.64	551
LAML	**Leukemia**	188	188	0	0.85 ± 0.31	NA	311
LGG	**Low-grade glioma**	511	511	0	2.41 ± 1.08	NA	2620
LIHC	**Liver**	421	371	50	5.12 ± 2.27	5.53 ± 1.61	935
LUAD	**Lung (adeno)**	529	483	46	5.47 ± 2.73	5.99 ± 2.75	1250
LUSC	**Lung (squamus)**	519	474	45	3.83 ± 2.10	8.22 ± 2.88	1137
OV	**Ovarian**	489	489	0	4.01 ± 1.99	8.34 ± 4.70	976
PRAD	**Prostate**	545	493	52	6.71 ± 3.56	NA	1233
READ	**Rectum**	160	157	3	5.32 ± 4.17	1.10 ± 0.34	347
SKCM	**Melanoma**	100	98	2	4.20 ± 2.24	1.94 ± 0.11	219
STAD	**Stomach**	430	389	41	5.31 ± 4.16	9.70 ± 6.57	879
THCA	**Thyroid**	587	518	69	5.65 ± 2.09	7.16 ± 2.12	1057
UCEC	**Uterus**	548	515	33	5.12 ± 3.74	16.69 ± 7.57	1092
Total		**8595**	**7961**	**634**	**4.76**	**7.50**	**19446**

Supplemental Table S2. Summary of high-confidence RNA editing events and RNA editing hotspots

BLCA	#RNA editing events	#RNA editing hotspots
AG	929	5
AC	355	16
UA	153	1
GU	28	2
CU	25	1
UG	16	1

BRCA	#RNA editing events	#RNA editing hotspots
AG	2021	7
UA	81	1
AC	58	3
CU	23	1
GU	11	1

CESC	#RNA editing events	#RNA editing hotspots
AG	682	3
UA	107	1
AC	27	1

COAD	#RNA editing events	#RNA editing hotspots
AG	833	4
AC	25	2
UA	17	1

HNSC	#RNA editing events	#RNA editing hotspots
AG	1189	3
AC	801	20
GU	499	15
UA	66	1

KICH	#RNA editing events	#RNA editing hotspots
AG	103	2
GU	13	1

KIRC	#RNA editing events	#RNA editing hotspots
AG	1032	5
UA	24	1
CU	11	1
UA	301	1

KIRP	#RNA editing events	#RNA editing hotspots
AG	551	4
AC	239	12
GU	214	11
UA	96	1

LAML	#RNA editing events	#RNA editing hotspots
AG	311	3

LGG	#RNA editing events	#RNA editing hotspots
AG	2620	14
AC	760	44
UA	171	1
GC	32	1
UG	21	1
CU	18	1
UC	15	1

LIHC	#RNA editing events	#RNA editing hotspots
AG	935	7
GU	236	18
AC	199	11
UA	105	1

LUAD	#RNA editing events	#RNA editing hotspots
AG	1250	7
AC	1141	22
GU	591	15
UA	43	1

LUSC	#RNA editing events	#RNA editing hotspots
AG	1137	6
AC	535	15
GU	259	11
UA	49	1

OV	#RNA editing events	#RNA editing hotspots
AG	976	5
GC	282	1
CU	59	2
UG	30	1
UA	11	1

PRAD	#RNA editing events	#RNA editing hotspots
AG	1233	4
UA	301	1
AC	177	10
CU	139	2
READ	#RNA editing events	#RNA editing hotspots
AG	347	3
AC	28	1
SKCM	#RNA editing events	#RNA editing hotspots
AG	219	4
UA	33	1
UG	28	1
GC	17	1
GU	10	1
STAD	#RNA editing events	#RNA editing hotspots
AG	879	3
UA	152	1
AC	126	5
CA	54	3
UG	47	1
CU	13	1
THCA	#RNA editing events	#RNA editing hotspots
AG	1057	6
AC	537	23
UA	71	1
GU	36	2
UG	35	1
UCEC	#RNA editing events	#RNA editing hotspots
AG	1092	4
AC	106	7
UA	44	1
GU	20	2

Supplemental Table S3. Information about 19 A-to-I RNA editing hotspots identified in miRNAs

miRNA	Edited sequence	Cross mapping	Edited seed region	Shared seed region with miRNA
hsa-miR-376a-5p	guggauucuccuucuaugagua	No	uggauuc	No
hsa-miR-381-3p	uaugcaagggcaagcucucugu	No	augcaag	No
hsa-miR-411-5p	uaguggaccguauagcguacg	No	aguggac	No
hsa-miR-99a-5p	gacccguagauccgaucuugug	No	acccgua	not in seed region
hsa-miR-379-5p	ugguggacuauggaacguagg	No	gguggac	hsa-miR-8071
hsa-miR-376c-3p	aacauggaggaaauuccacgu	No	acaugga	hsa-miR-4802-3p
hsa-miR-589-3p	ucagagcaaaugccgguucccaga	No	cagagca	hsa-miR-6501-3p
hsa-miR-664a-5p	acuggcuggggaaaaugauuggau	No	cuggcug	hsa-miR-3064-5p
hsa-miR-497-5p	cggcagcacacuguggguuugu	No	ggcagca	No
hsa-miR-151a-3p	cuggacugaagcuccuugagg	No	uggacug	hsa-miR-1269a
hsa-miR-200b-3p	uaaugcugccugguaaugauga	No	aaugcug	No
hsa-miR-3144-3p	auguaccuguucggucucuuua	No	uguaccu	No
hsa-miR-1301-3p	uugcggcugccugggagugacuuc	No	ugcggcu	No
hsa-miR-1251-5p	acucuggcugccaaaggcgcu	No	cucuggc	No
hsa-miR-6503-3p	gggacugggaugcagaccucc	No	ggacugg	hsa-miR-4515
hsa-miR-1295b-3p	aauaggccgcggaucugggcaa	No	auaggcc	not in seed region
hsa-miR-337-3p	cuccuguaugaugccuuucuuc	No	uccugua	No
hsa-miR-1304-3p	ucucgcuguagccucgaacccc	No	cucgcug	No
hsa-miR-3622a-3p	ucgccugaccucccaugccugu	No	cgccuga	hsa-miR-6078

Supplemental Table S4. Validation of the correlations of miRNA editing level with ADAR2 enzymes using a miRNA-seq dataset of ADAR2-perturbed cell lines

miRNA	Editing site	TCGA LGG		U87 cell line		U118 cell line			
		Correlation Coefficient	FDR	control	ADAR2 over - expression	control	ADAR2 over - expression	ADAR2 E/A	siADAR2
miR-99a	1	0.64	0	0	0.786	0	13.815	0	6.913
miR-379	5	0.51	0	NE	NE	NE	NE	NE	NE
miR-497	2	0.42	8.36E-22	0	1.058	0	25.862	0	16.245
miR-411	5	0.37	2.53E-16	0	2	NE	13.26	NE	5.882
miR-1301	5	0.33	9.58E-14	0	NE	0	9.524	0	0
miR-1251	6	0.31	5.79E-11	NE	NE	NE	NE	NE	NE

The editing levels (%) in two glioblastoma cell lines U87 and U118 were presented for 6 miRNA editing hotspots that show significantly strong correlation (*FDR* < 0.05, spearman correlation *Rs* ≥ 0.3) with ADAR2 enzyme in TCGA LGG patient samples. ADAR2 E/A is the inactive form of ADAR2. “NE” depicts sites with not enough coverage (< 10) to quantify the editing level.

Supplemental Table S5. List of potential new targets of edited miR-200b

Gene	Transcript	# Match to wt miR-200b	# Match to edited miR-200b	FDR	Log_2 fold change
LIFR	NM_002310	0	2	2.05E-12	0.940541692
RAB5C	NM_004583	0	2	7.05E-12	1.045945875
MFAP3	NM_005927	0	2	1.34E-11	1.178306516
LIN9	NM_173083	0	2	1.28E-10	1.418951397
MARCH6	NM_005885	0	4	2.66E-10	0.966610739
GINS1	NM_021067	0	2	2.66E-10	0.823375314
SLC12A6	NM_005135	0	2	1.93E-09	0.976012095
BTBD3	NM_014962	0	2	2.02E-09	1.224580566
CSRP2	NM_001321	0	2	1.65E-08	1.173966876
C15orf41	NM_032499	0	3	1.02E-07	0.923243231
SUN1	NM_025154	0	2	1.49E-07	0.76629024
PDS5A	NM_001100399	0	3	1.97E-07	1.066848588
ARL5B	NM_178815	0	2	2.48E-07	0.880867339
SNX13	NM_015132	0	2	1.23E-06	0.753840947
IL13RA1	NM_001560	0	2	1.47E-06	0.691503253
ZSCAN31	NM_030899	0	2	3.54E-06	0.814574932
TRPC1	NM_003304	0	2	5.13E-06	1.08040322
FAR1	NM_032228	0	2	5.38E-06	0.960681537
RUNDC1	NM_173079	0	2	1.01E-05	0.697887552

Supplemental Table S6. List of 3'UTR binding sites of wild-type and edited miR-200b to their target genes ZEB1, ZEB2, and LIFR.

Target Gene	3'UTR binding sites
ZEB1	357…GTTTTATCTTATCAGTATTA…376
	451…CTAAATCCGCTTCAGTATTT…470
	880…AGTGCCATTTCTCAGTATTT…899
	1231…ATTTTTACCTATCAGTATTA…1250
	1301…CTTCAAACCTGGCAGTATTA…1320
	1952…TTTCATCATTATCAGTATTT…1971
ZEB2	380…ACTACCATACATCAGTATTT…399
	443…ACTACAATGCATCAGTATTA…462
	801…AAGCACCCATGTCAGTATTA…820
	886…CATTAATTTTGTCAGTATTA…905
	1017…TACTGTAGTGTACAGTATTA…1036
LIFR	500…CTCCTCTATCCACAGCATTC…519
	3255…CATTTTCCAAACCAGCATTA…3274