Supplementary Materials

3'UTR shortening identifies high-risk cancers with targeted dysregulation of the ceRNA network

Li Li[1*], Duolin Wang[2*], Mengzhu Xue[1*], Xianqiang Mi[1], Yanchun Liang[2], Peng Wang[1,3‡]

1. Key Laboratory of Systems Biology, Shanghai Advanced Research Institute, Chinese Academy of Sciences
2. College of Computer Science and Technology, Jilin University
3. School of Life Science and Technology, ShanghaiTech University

[‡]Correspondence should be addressed to P.W. (email: wangpeng@sari.ac.cn)

[*]These authors contributed equally to this work

Supplementary Tables

Supplementary Table 1. Motifs enriched in the extended 3'UTR region of genes with negative correlation between gene expression and 3'UTR shortening

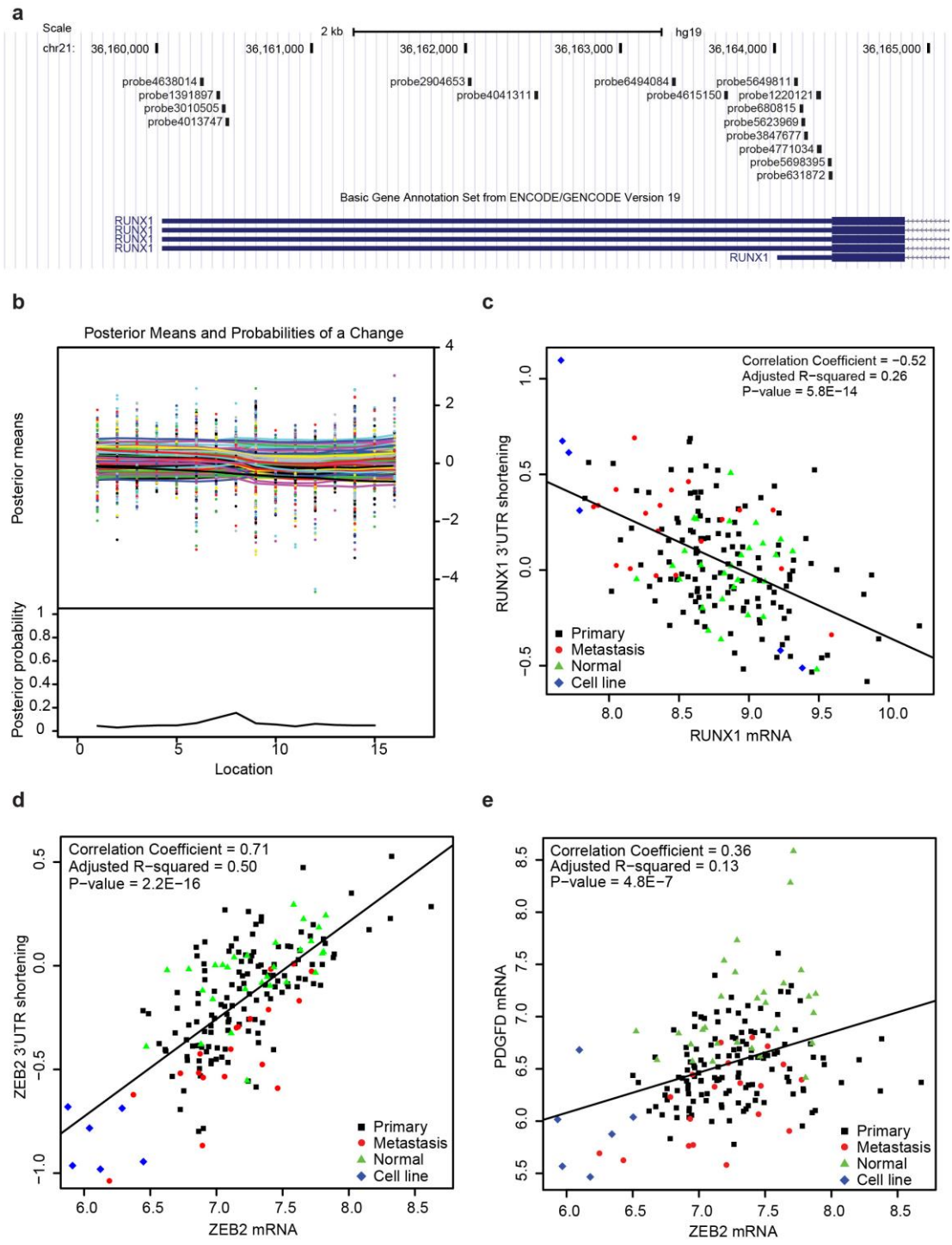
Motif	p-value	q-value	Motif annotation
GAGCCC	1.08E-08	4.43E-05	
CCCCCA	1.71E-05	0.01	poly(C)-binding protein
AGAGCC	1.98E-05	0.01	
CCCCCC	2.21E-05	0.01	poly(C)-binding protein
GCCCCC	2.34E-05	0.01	poly(C)-binding protein
CCCCCT	6.39E-05	0.04	poly(C)-binding protein
GCGGCC	7.11E-05	0.04	

Supplementary Table 2. DAVID analysis result of significantly dysregulated genes in high-risk prostate cancers

Only the top 10 pathways are shown.

Pathway	P-value	Fold enrichment	Benjamini corrected p-value	Genes
Jak-STAT signaling pathway	1.97E-04	5.37	0.007	ZFP91, CBLB, IL6ST, SOS1, SOS2, JAK1, SOCS5, AKT3, PIK3R1
Chronic myeloid leukemia	1.24E-04	8.63	0.009	E2F1, CBLB, SOS1, SOS2, SMAD4, AKT3, PIK3R1
Pancreatic cancer	9.12E-04	7.70	0.012	E2F1, CDC42, SMAD4, JAK1, AKT3, PIK3R1
Focal adhesion	1.13E-03	4.14	0.012	CDC42, ARHGAP5, ITGA6, SOS1, SOS2, RAP1B, PTEN, AKT3, PIK3R1
Glioma	4.93E-04	8.81	0.013	E2F1, SOS1, SOS2, PTEN, AKT3, PIK3R1
T cell receptor signaling pathway	9.01E-04	5.99	0.014	CDC42, CBLB, SOS1, SOS2, NFATC3, AKT3, PIK3R1
Renal cell carcinoma	8.02E-04	7.92	0.015	CDC42, SOS1, SOS2, RAP1B, AKT3, PIK3R1
Non-small cell lung cancer	2.43E-03	8.56	0.017	E2F1, SOS1, SOS2, AKT3, PIK3R1
Endometrial cancer	2.11E-03	8.89	0.018	SOS1, SOS2, PTEN, AKT3, PIK3R1
Prostate cancer	2.36E-03	6.23	0.018	E2F1, SOS1, SOS2, PTEN, AKT3, PIK3R1

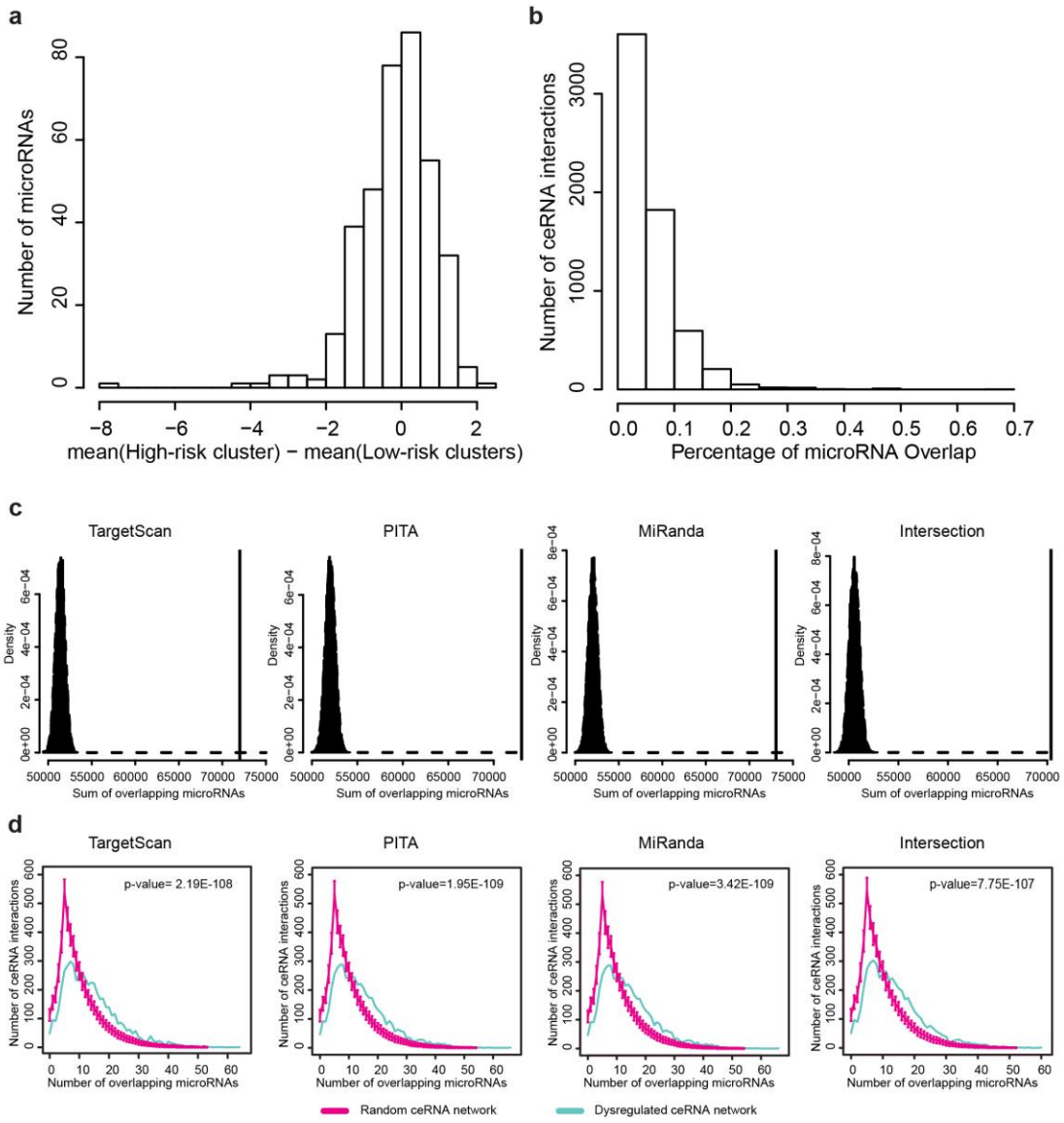
Supplementary Figures



Supplementary Figure 1. Example genes whose mRNA expression displays positive or negative correlation with 3'UTR shortening

(**a**) Plot showing the mapping of exon array probes to the RUNX1 3'UTR region. (**b**) Plot showing the BCP analysis results. In the upper panel, each dot represents a normalized probe value for a particular sample, and the lines represent posterior estimated probe mean values. Probes are ordered according to their genomic locations. The lower panel shows the estimated posterior probabilities of changes. Position 8 is of the highest posterior change probability which matches

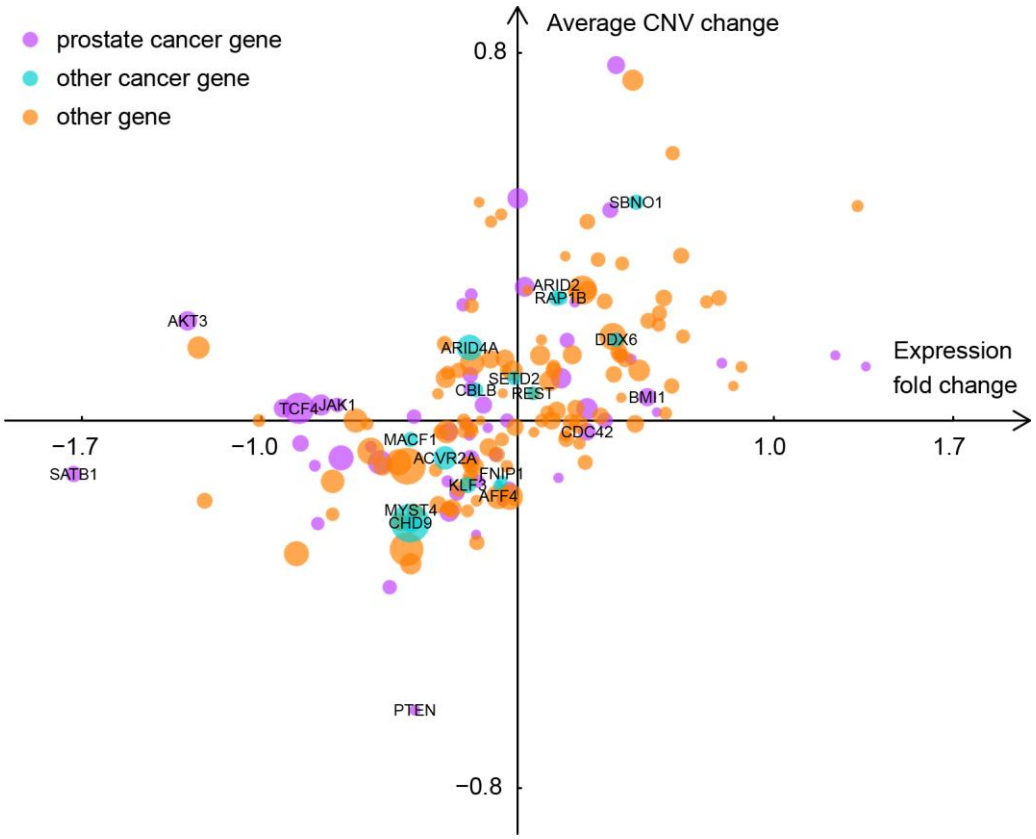
the position inferred from the probe mapping result in (**a**). (**c**) Scatter plot showing the negative correlation between RUNX1 mRNA expression and RUNX1 3'UTR shortening. (**d**) Scatter plot showing the positive correlation between ZEB2 mRNA expression and ZEB2 3'UTR shortening. (**e**) Scatter plot showing the positive correlation between PDGFD mRNA and ZEB2 mRNA. Samples are colored according to their phenotypes.



Supplementary Figure 2. Dysregulated ceRNA interactions share unusually high number of microRNAs with 3'UTRs displaying large APA dynamics

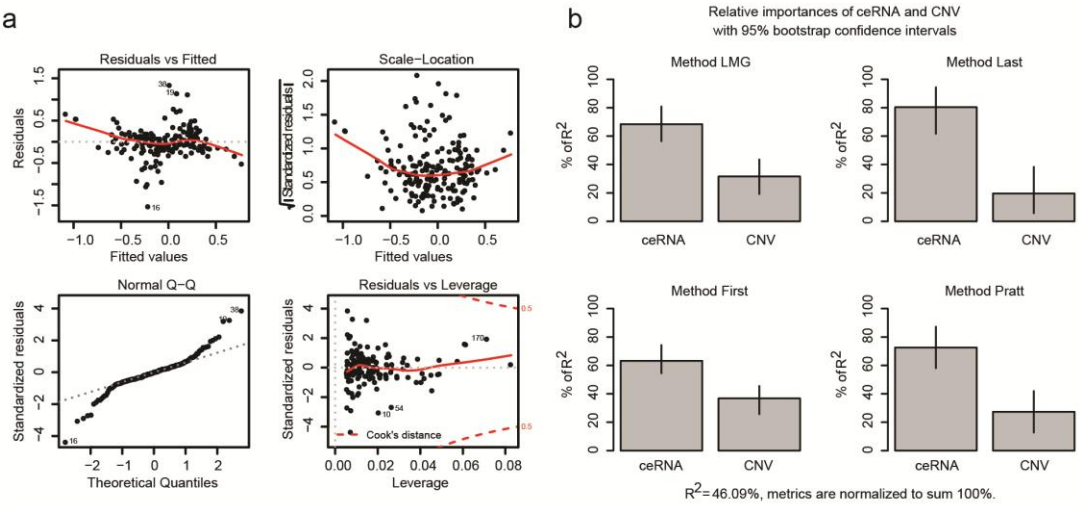
(**a**) MicroRNAs are predominately downregulated in high-risk prostate cancers. Only six of the differentially expressed microRNAs are upregulated, all with limited fold of changes (< 2). (**b**) Histogram shows that for vast majority of the dysregulated ceRNA interactions (86%), there is less than 10% overlap between the differentially expressed microRNAs and those microRNAs mediating that particular dysregulated ceRNA interaction. (**c**) Vertical lines represent the numbers of overlapping microRNAs between the dysregulated ceRNA interactions and the microRNAs in the 3'UTRs of genes displaying large APA dynamics, which are significantly higher that of random ceRNA interactions (black histograms on left part of the figures, p-values < 0.001). (**d**) The overlapping distributions between microRNAs in the 3'UTRs of genes displaying large APA dynamics and microRNAs mediating dysregulated ceRNA interactions (turquoise) are strongly bias towards larger overlaps (p-values < 0.001, chi-squared test) comparing with average overlaps from 10,000 random ceRNA networks (magenta). For (**c**) and (**d**), four different comparisons were carried out using results from TargetScan, PITA, miRanda and their intersection. All four

comparisons gave very similar results, suggesting that variations in microRNA target prediction algorithms have limited impacts on the high microRNA overlap between Dysregulated ceRNA interactions and 3'UTRs with large APA dynamics.



Supplementary Figure 3. Driver CNV vs. expression change for genes with significant ceRNA dysregulation

The x-axis represents the up/downregulation of genes (high-risk vs. low-risk). Y-axis represents the average driver CNV difference of genes (high-risk vs. low-risk). The gene symbols are sized according to the FDR of enrichment of up/downregulated genes in that gene's dysregulated ceRNA interactions (same as in Fig. 3e). There is a significant positive correlation between gene expression fold changes and CNV status (cor = 0.49, p-value = 3.58E-12). The value represents a smaller and less significant correlation than that between expression changes and FDRs of enrichment of up/downregulated genes in gene's dysregulated ceRNA interactions (cor = -0.64, p-value = 4.04E-22). This observation suggests that the observed expression changes for genes with significant ceRNA dysregulation mostly result from enrichment of genes with opposite expression changes in their dysregulated ceRNA interactions, with CNVs contributing to a lesser extent.



Supplementary Figure 4. Preferred dysregulation of ceRNA interactions involving genes displaying opposite expression changes is the main factor contributing to the up/downregulation of genes with significant ceRNA dysregulation

(**a**) Results from multivariate linear regression analysis. A simple model ($Y = \varepsilon + \beta_1 X1 + \beta_2 X2$) was utilized to evaluate the effects of enrichment of up/downregulated genes in a gene's dysregulated ceRNA network (X1) and that gene's CNV status (X2) on that gene's expression change (Y). Except a few outliers, majority of the data points fit well to a linear model. An adjusted R-squared value of 0.46 suggests that the ceRNA enrichment and CNV can largely explain observed expression changes. (**b**) Bootstrapping demonstrated that ceRNA enrichment is the main factor contributing to the observed gene expression changes. On average over 60% of the explained variance is due to the enrichment of genes with opposite expression changes in the dysregulated ceRNA network.