SUPPLEMENTARY TABLES，FIGURES AND METHODS

Tables

Table S7. Prognosis-associated lincRNAs obtained in UCEC discovery (FDR < 0.01) and validation set

LincRNA	Discovery (FDR)	Validation (Log-rank p < 0.05 / reference / neither : -)
XLOC_013781	4.00E-04	3.00E-02
XLOC_006505	1.40E-03	3.00E-02
nr_XLOC_004488_GAS5*	1.60E-03	4.10E-02/(1)
XLOC_007617*	1.60E-03	3.10E-04
XLOC_000819	1.60E-03	1.50E-04
XLOC_011604	2.00E-03	2.30E-02
XLOC_003068	2.00E-03	1.30E-02
XLOC_001907	2.60E-03	-
XLOC_001436	3.20E-03	-
XLOC_002624	3.40E-03	4.70E-02
XLOC_009017	3.80E-03	1.30E-04
XLOC_014121	4.00E-03	4.90E-02
XLOC_003350	4.20E-03	4.70E-03
XLOC_000854	4.20E-03	-
XLOC_013424	4.40E-03	1.00E-03
XLOC_006333	4.60E-03	2.10E-02
XLOC_008698	5.20E-03	2.10E-02
XLOC_008342	5.80E-03	-
XLOC_009707	6.00E-03	3.30E-02
XLOC_002323	6.00E-03	-
XLOC_011704	6.20E-03	1.60E-02
XLOC_011563	6.20E-03	3.10E-02
XLOC_003221	6.60E-03	2.90E-03
XLOC_006588	8.00E-03	1.20E-02
XLOC_005176	8.20E-03	2.12e-05/smoking cessation(2)
XLOC_011131	8.40E-03	-
XLOC_006456	8.40E-03	3.60E-02
XLOC_003734	8.80E-03	-
XLOC_010706	9.40E-03	3.30E-02
XLOC_007478	9.40E-03	1.30E-03

Table S8. Top 10 prognosis-associated lincRNAs in RBCA discovery (FDR < 0.01) and validation set

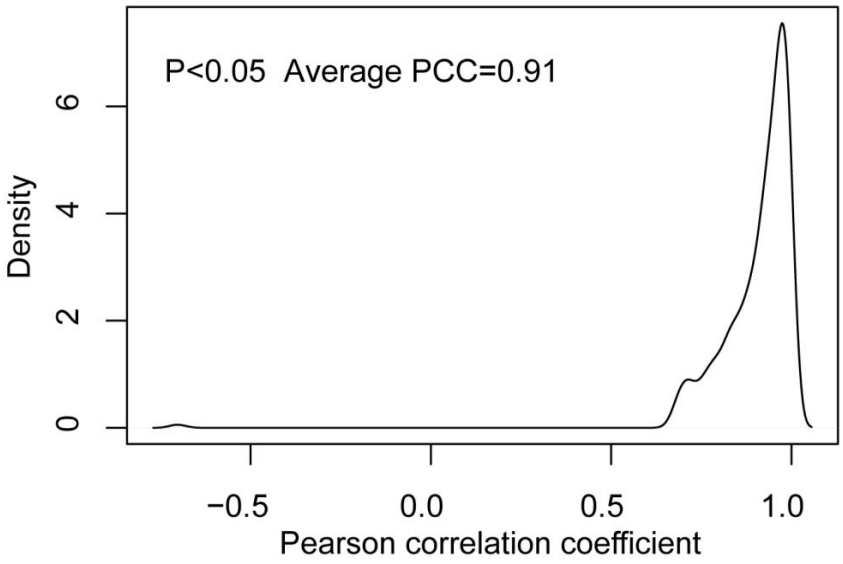
LincRNA	Discovery (FDR)	Validation (Log-rank p < 0.05 / reference /neither : -)
XLOC_000643	2.00E-04	-
XLOC_006991	8.00E-04	8.70E-03
XLOC_000633	8.00E-04	-
XLOC_010595	1.60E-03	1.70E-02
XLOC_013029	2.20E-03	-
XLOC_009284*	2.40E-03	3.20E-02
XLOC_013558	2.80E-03	-
XLOC_003378	3.00E-03	2.50E-02
XLOC_003069	4.00E-03	-
XLOC_010422	6.60E-03	2.70E-02

Table S9. Prognosis-associated lincRNAs in LUSC discovery (FDR < 0.01) and validation set

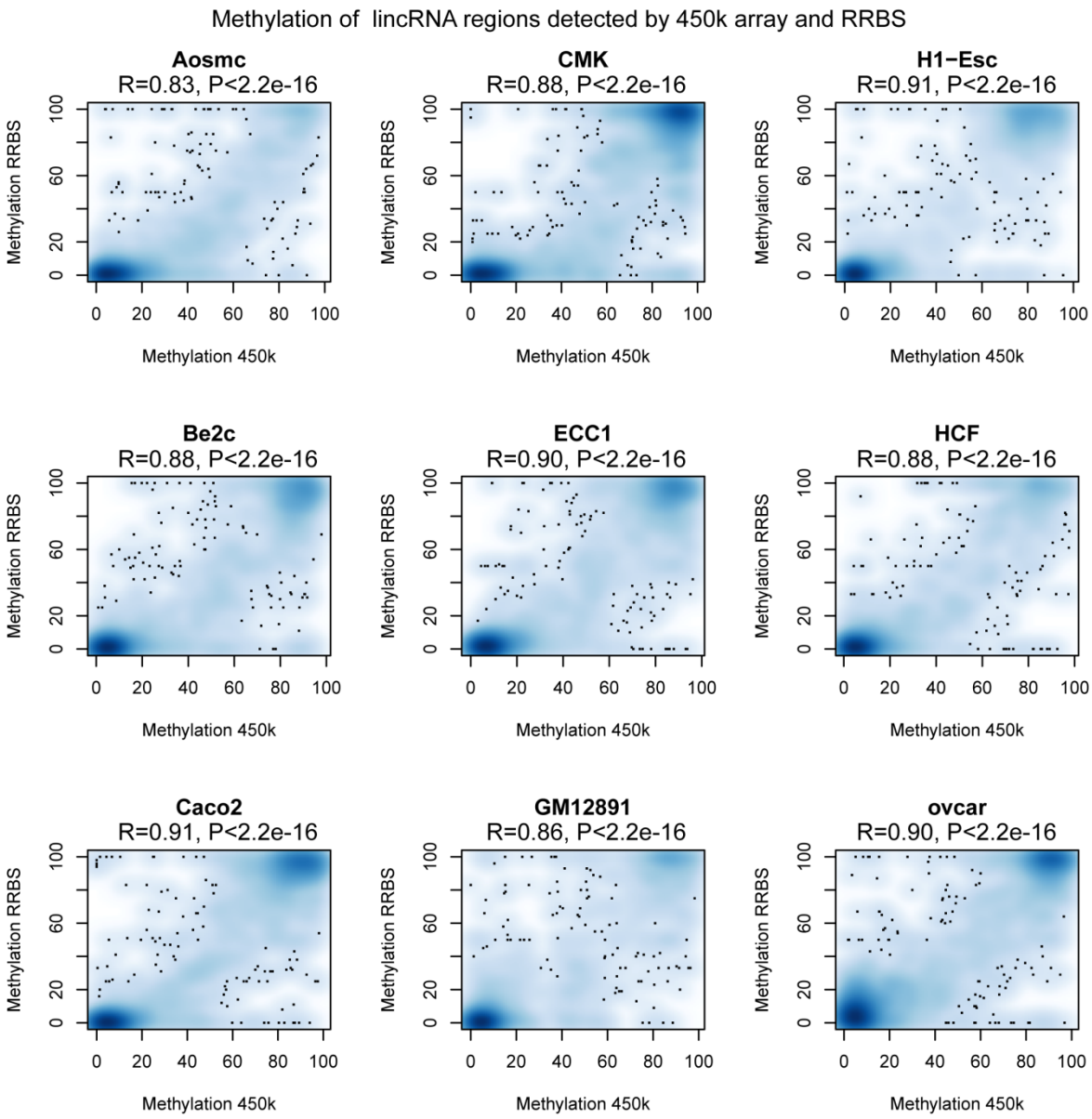
lincRNA	Discovery (FDR)	Validation (Log-rank p < 0.05 / reference / neither : -)
XLOC_000377	2.00E-04	-
XLOC_004549	2.00E-04	-
XLOC_009607	2.00E-04	-
nr_XLOC_088170_MEG3	8.00E-04	(3)
XLOC_005028	8.00E-04	-
XLOC_010922	1.20E-03	-
XLOC_012488	1.20E-03	-
XLOC_011171	1.40E-03	-
XLOC_004706	2.20E-03	4.30E-02
XLOC_005912	2.80E-03	-
XLOC_010544	3.00E-03	-
XLOC_003038	3.20E-03	-
XLOC_008503	4.20E-03	-
XLOC_009367*	5.60E-03	9.50E-04
XLOC_004905	6.00E-03	-
XLOC_005883	6.00E-03	4.00E-02
XLOC_005815	7.60E-03	9.92E-05
XLOC_005560	7.80E-03	-
XLOC_005087	8.00E-03	1.70E-02
XLOC_000895	8.60E-03	4.50E-02
XLOC_000360	8.80E-03	2.70E-02

For each cancer, we randomly divided the tumors into a discovery set and a validation set. Shown are lincRNAs with significant OS-associated methylation patterns for BRCA, LUSC and UCEC (log-rank test, FDR < 0.01 in the discovery sets). * LincRNAs mentioned in the manuscript.

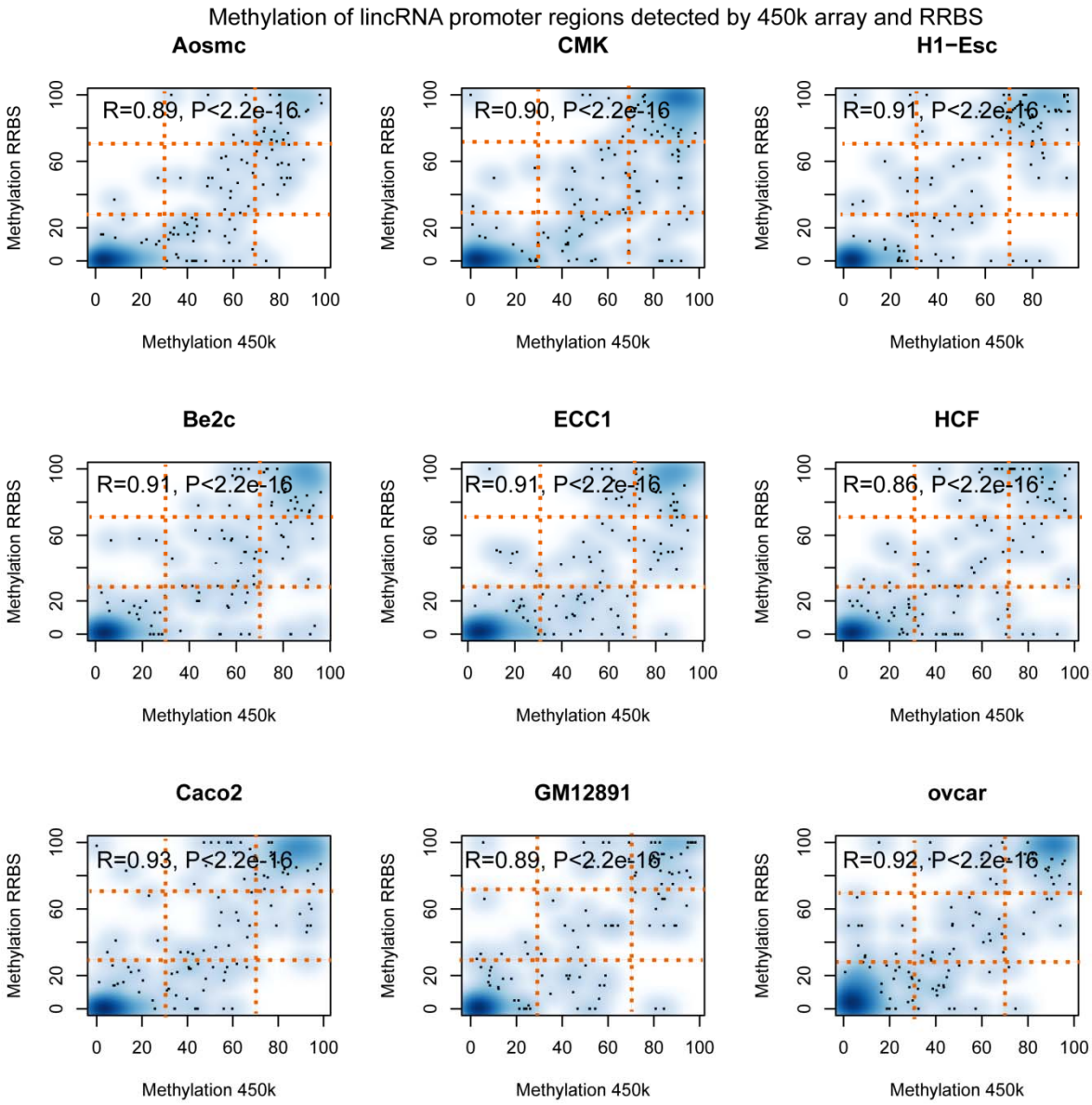
Figures



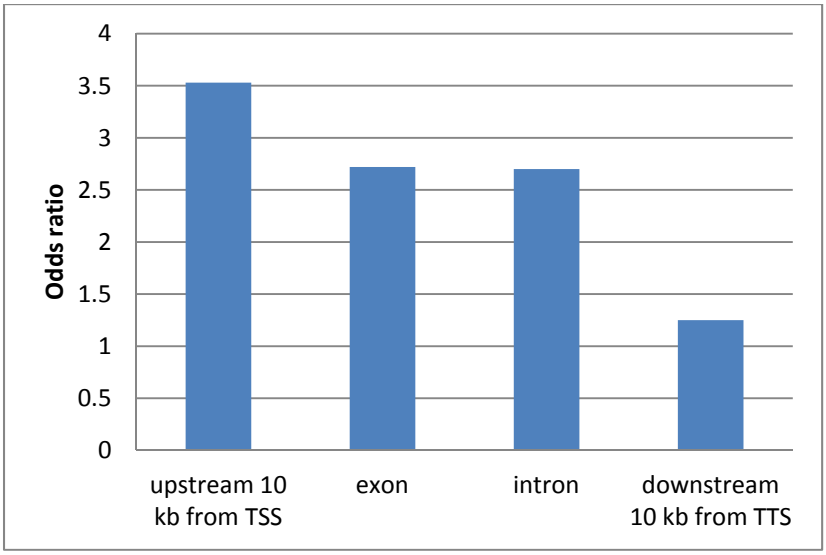
Supplementary Figure S1. Methylation levels of probes mapped to regions associated with lincRNAs were reliable. Density plot with Pearson correlation coefficient (PCC) ($p < 0.05$) for distribution of methylation levels between Infinium 450k array and RRBS in nine cell lines from the ENCODE project. Only probes mapped to regions associated with lincRNAs were considered.



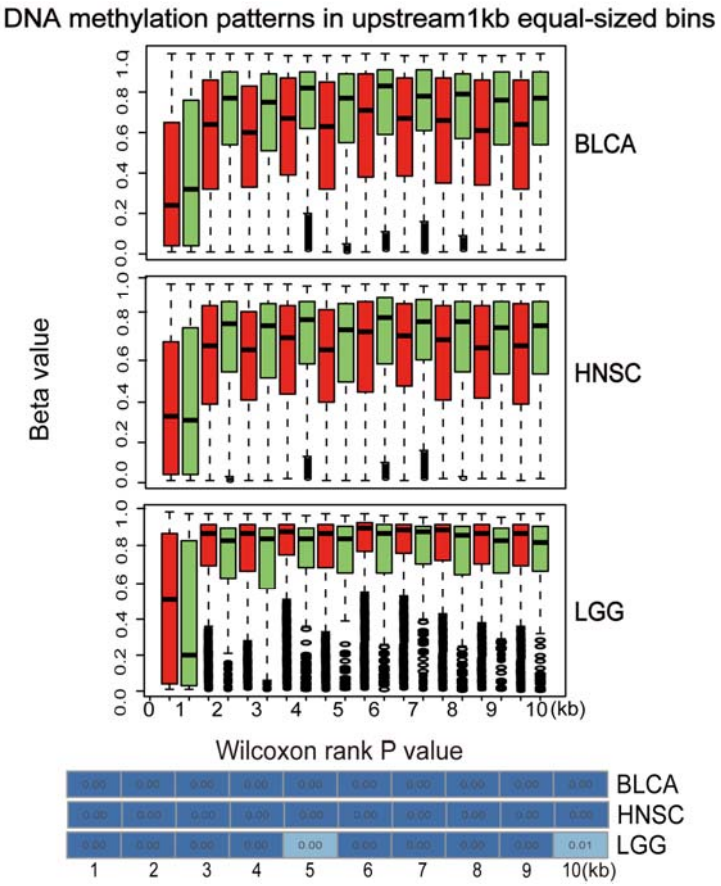
Supplementary Figure S2. Methylation of probes annotated to regions associated with lincRNAs from Infinium 450k data vs. RRBS data of nine cell lines. Scatter plot of percent methylation values of probes mapped to lincRNA functional regions from Infinium 450k arrays (x-axis) and percent methylation values of corresponding sites from RRBS (y-axis). High density scatter plots were generated in R using a smoothed density method.



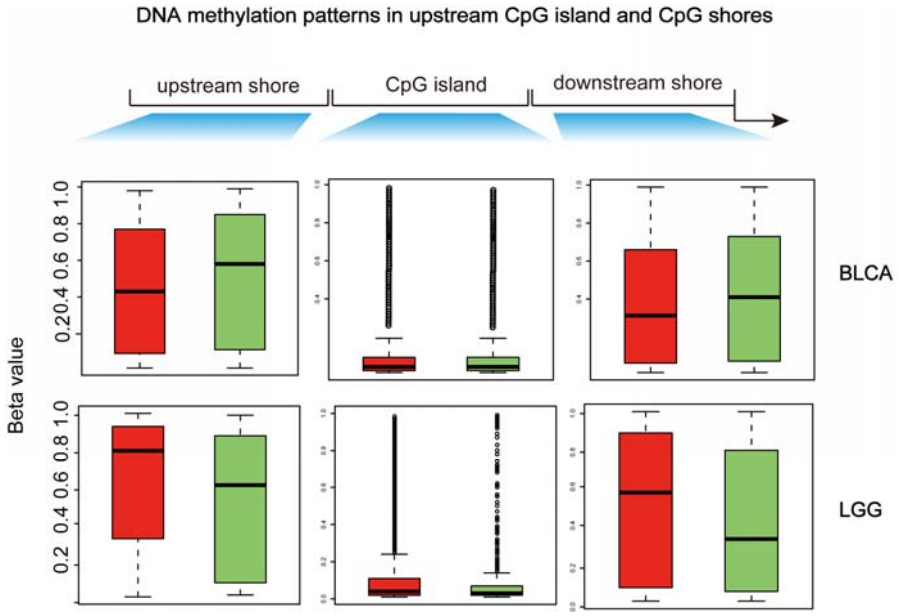
Supplementary Figure S3. Methylation of probes annotated to lincRNAs promoters from Infinium 450k data vs. RRBS data of nine cell lines. Scatter plot of percent methylation values of probes mapped to lincRNA promoters from Infinium 450k arrays (x-axis) and perent methylation values of corresponding sites from RRBS (y-axis). High density scatter plots were generated in R using a smoothed density method.



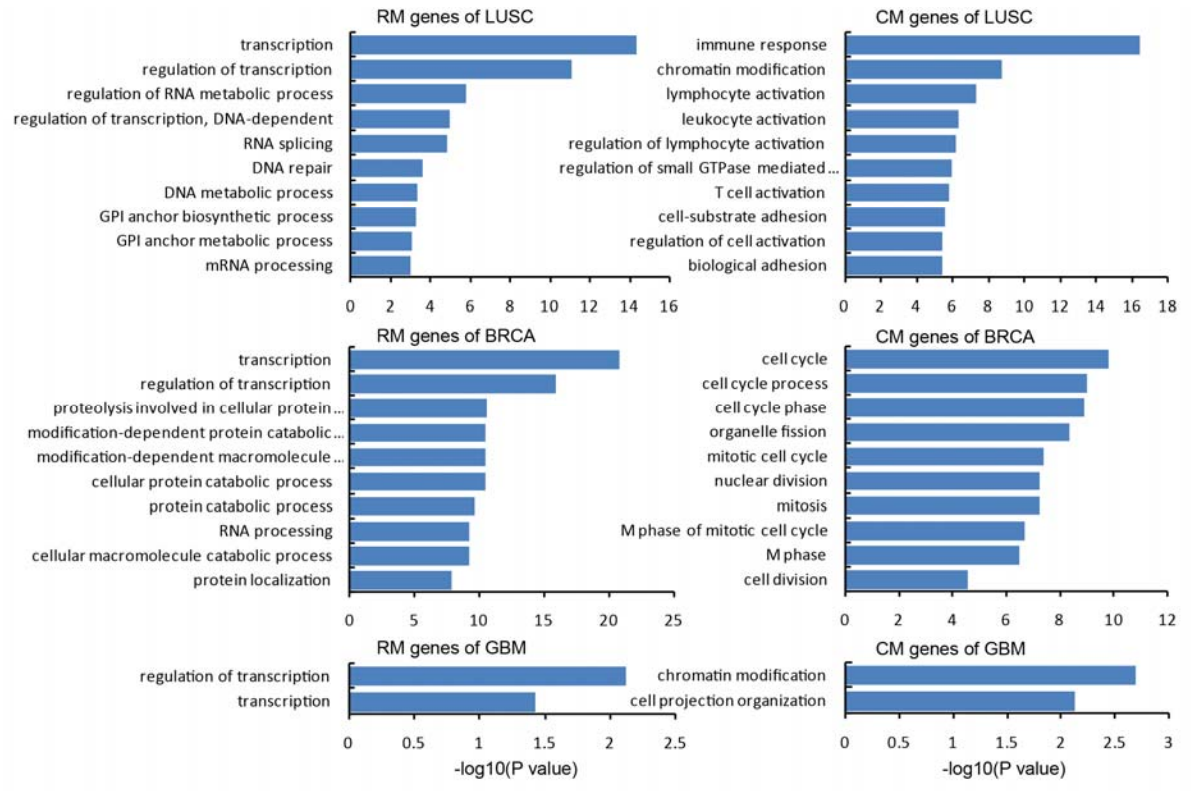
Supplementary Figure S4. The odds ratio indicates that promoter hypermethylation is tightly linked to transcriptional silencing of lincRNAs. To quantify the strength of the correlations between expression and lincRNA component methylation, we calculated the odds ratio to quantify the likelihood of transcriptional silencing of a lincRNA (expression level lower than the highest 10% expression quantile) if it is methylated (beta value > 0.3) within each lincRNA-associated region.



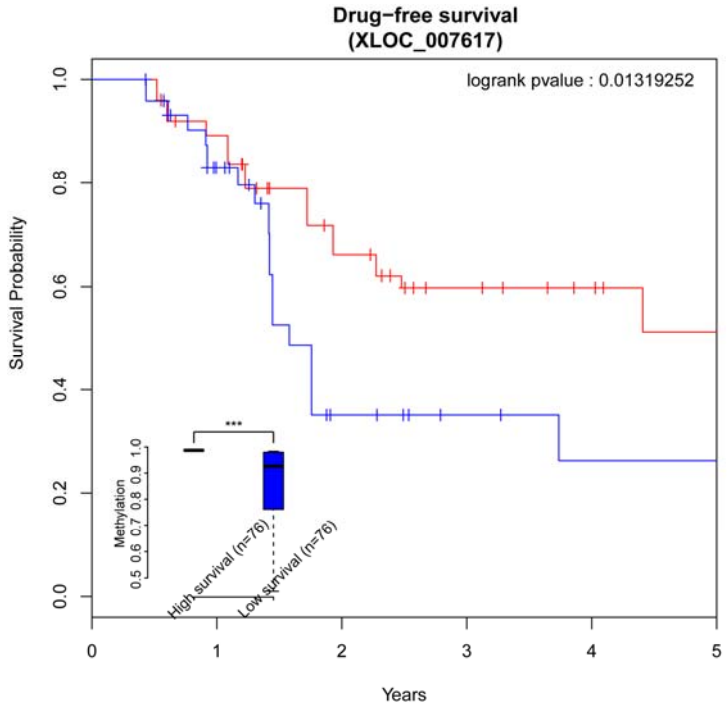
Supplementary Figure S5. Three representative DNA methylation patterns of 1kb equal-sized bins for lincRNA promoter in BLCA, HNSC and LGG.



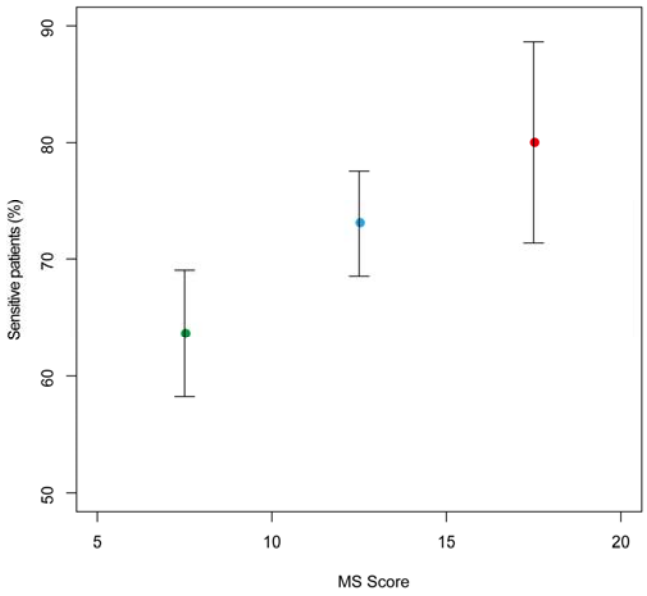
Supplementary Figure S6. Two representative methylation patterns of CGIs and CpG shores covered probes in BLCA and LGG. All tumors boxes were significantly different with the corresponding boxes of normal samples (Wilcoxon rank sum tests).



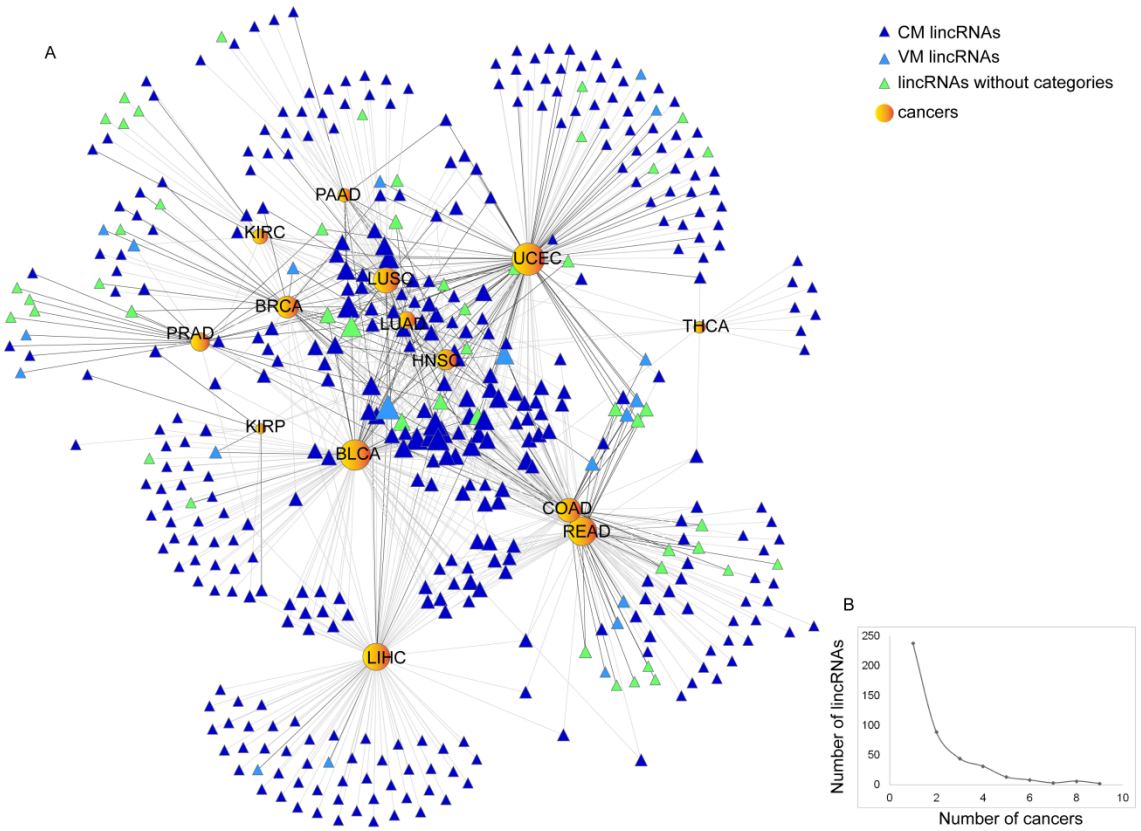
Supplementary Figure S7. Analysis of gene ontology (GO) term enrichment in the 'biological process' category for RM and CM genes. Plotted are the top 10 GO terms ranked according to the FDR adjusted p-values ($p < 0.05$) obtained using the David gene functional enrichment analysis tool. Full lists of enrichment GO terms (Biological Processes) of RM, VM and CM genes are shown in Supplementary Table S12.



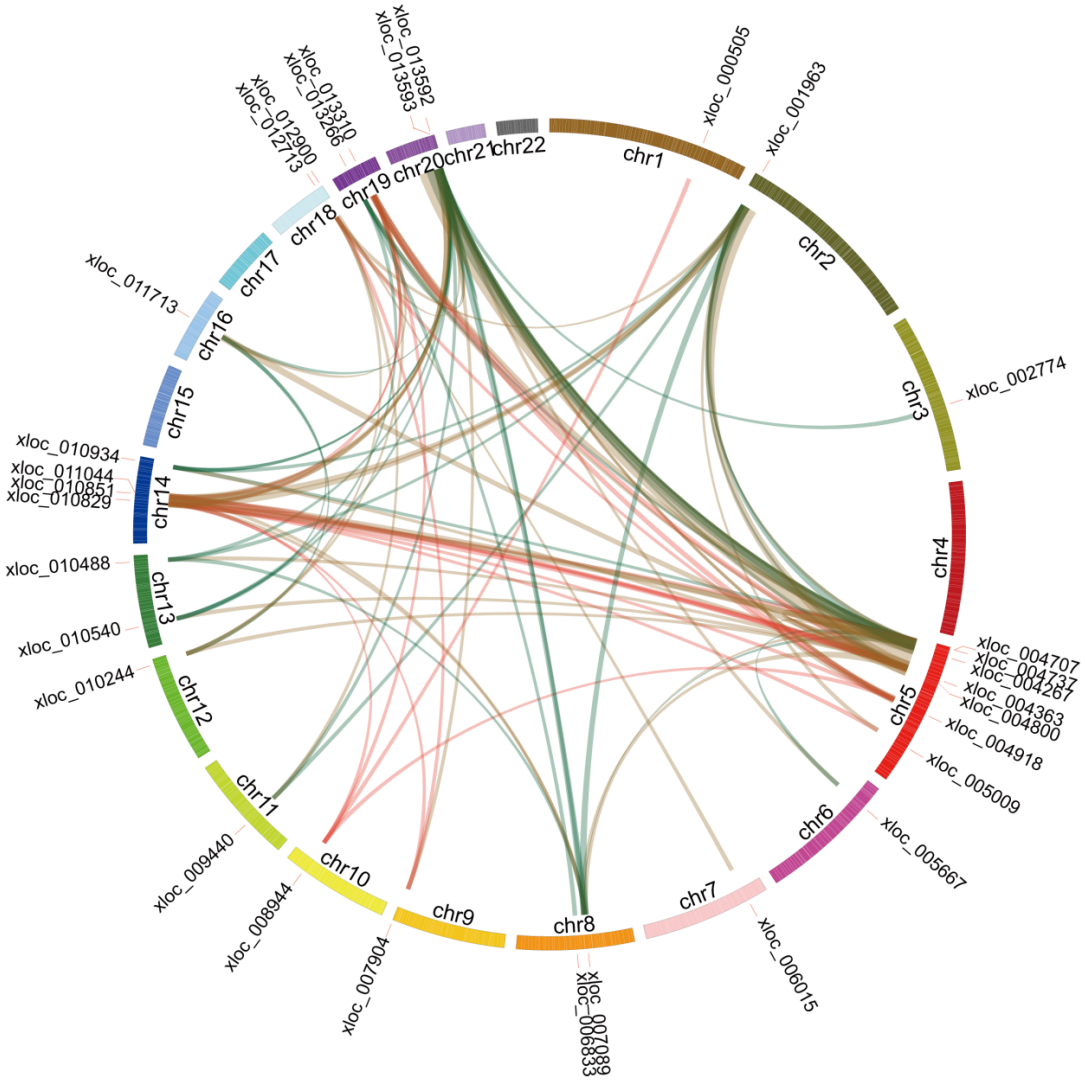
Supplementary Figure S8. Drug-free survival of XLOC_007617 in UCEC. The percent probability of survival is plotted vs time since the end of chemotherapy.



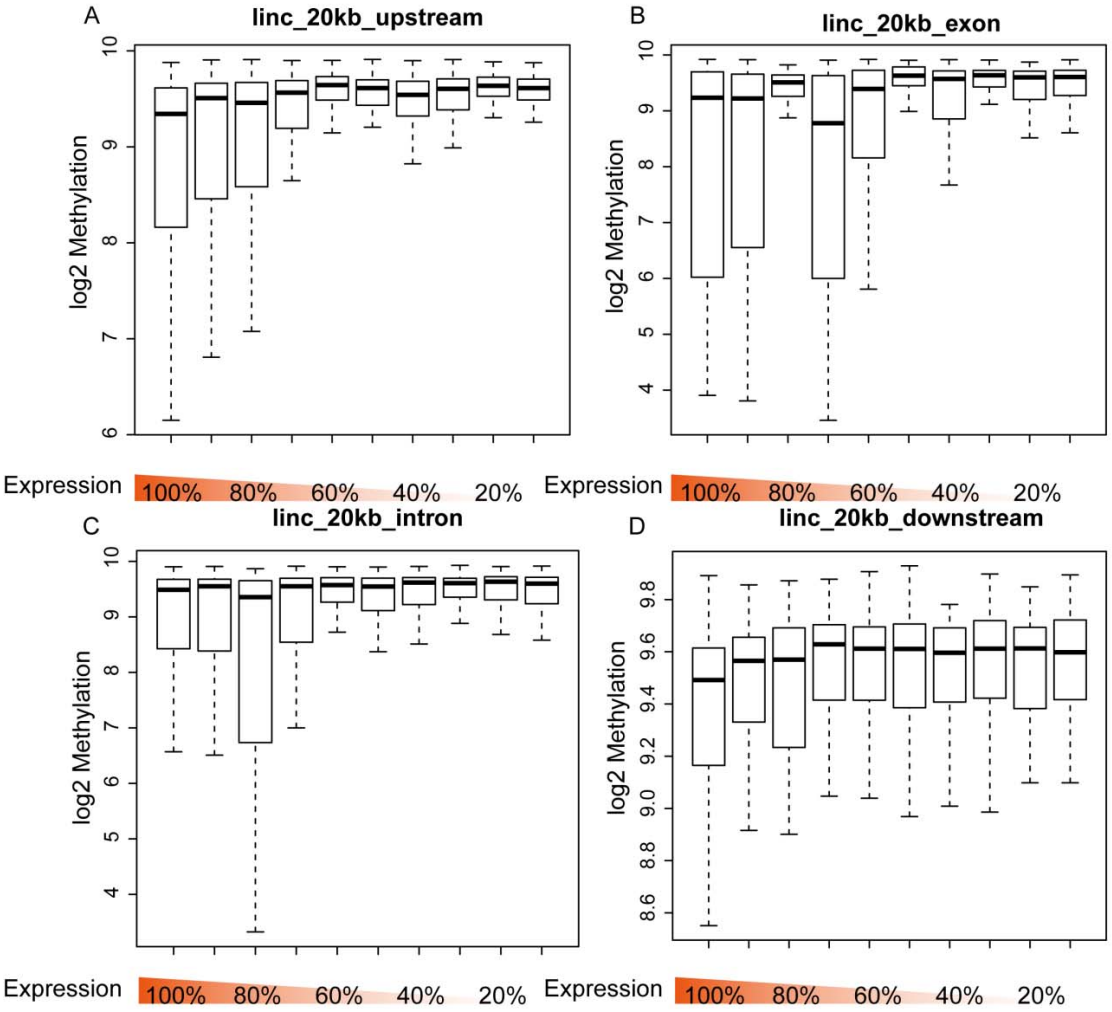
Supplementary Figure S9. Patients with higher MS scores were more sensitive to drugs. Points are the percent sensitive patients out of all patients with the MS scores of 5 to 10, 10 to 15, or 15 to 20. Error bars, standard deviation of 1000 bootstrap samples.



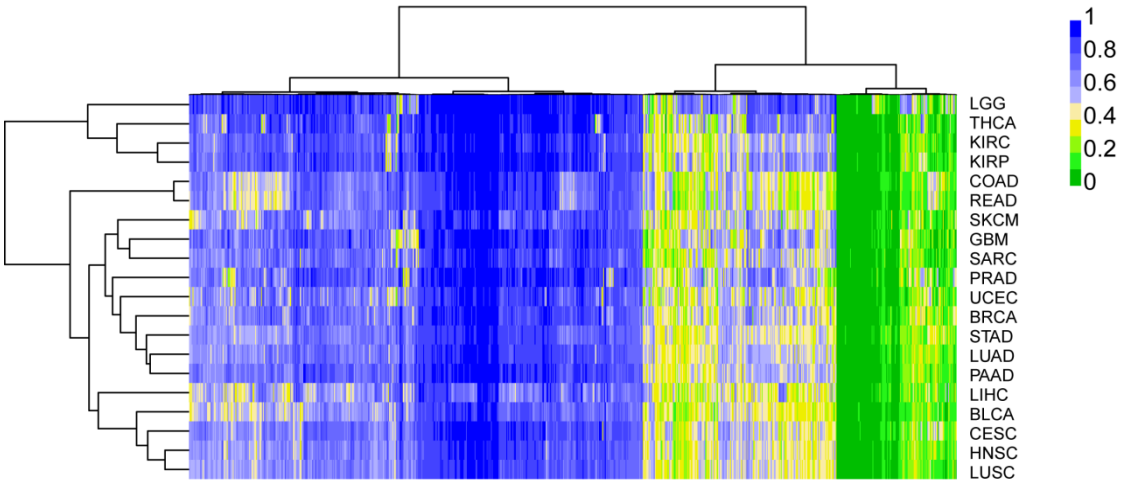
Supplementary Figure S10. Aberrantly methylated lincRNA-cancer network. **(A)** AMCN of 14 cancers and 434 AM lincRNAs. **(B)** Distribution of AM lincRNAs in cancers. Most lincRNAs were aberrantly methylated in one cancer; a few were aberrantly methylated in multiple cancers.



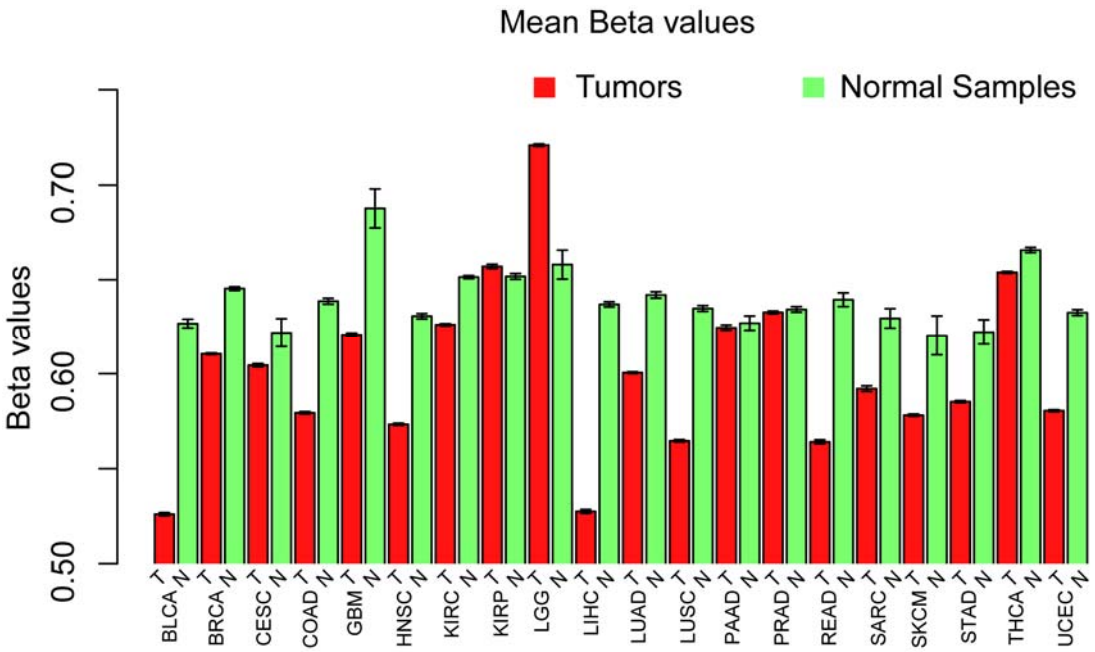
Supplementary Figure S11. Circos plot of AM lincRNA pairs that co-occurred in more than three cancers. Green, hypomethylated in tumors; red, hypermethylated in tumors; brown either hypomethylated or hypermethylated. Line width indicates number of cancer types containing the pair.



Supplementary Figure S12. Methylation patterns of a stringent subset of lincRNAs (>20 kb from the nearest PCGs) in 10 expression quantiles (lowest 10% to 100%). Box plots with methylation in **(A)** promoters, **(B)** exons, **(C)** introns, **(D)**10 kb downstream of TTS.



Supplementary Figure S13. Unsupervised hierarchical clustering of average methylation profiles of 883 lincRNAs (>20 kb from the nearest PCGs) in 20 cancer types.



Supplementary Figure S14. Bar plots with average methylation levels of promoters in a stringent subset of lincRNAs (>20 kb from nearest PCGs) in tumors and corresponding normal samples. Error bars, mean ± SEM.

Methods

Calculating the tissue-specificity of lincRNA expression

We used a method based on information theory proposed by Octavio et al.(4) to calculate the tissue-specificity of lincRNA expression. Similar to the previous method, we calculated a set of relative frequencies, p_{ij}, for the *i*th lincRNA, (*i*=1,2,…,*l*) in the *j*th tissue (*j*=1,2,…,*t*), to describe the lincRNA transcriptome of each tissue.

$$p_{ij} = \frac{\text{the expression of the } i\text{th lincRNA in the } j\text{th tissue}}{\sum_{i=1}^{l} \text{the expression of the } i\text{th lincRNA in the } j\text{th tissue}}$$

Tissue-specificity of each lincRNA was defined as:

$$S_i = \frac{1}{t}\left(\sum_{j=1}^{t} \frac{p_{ij}}{p_i} \log_2 \frac{p_{ij}}{p_i}\right)$$

Where p_i was the average frequency of the *i*th lincRNA among tissues,

$$p_i = \frac{1}{t}\sum_{j=1}^{t} p_{ij}$$

S_i was zero if the lincRNA was transcribed at the same frequency in all tissues and a maximum value of log2(*t*) if the lincRNA was exclusively expressed in a single tissue. We calculated the S_i value for each lincRNA expressed in all 22 tissues and cell lines and used it to measure the tissue-specificity of lincRNAs belonging to each category (RM, CM and VM). Figure 3B showed the cumulative distribution curves of the S_i values of RM, VM and CM lincRNAs (differences between lincRNA sets were tested using Wilcoxon rank sum tests).

Reference

1. Mourtada-Maarabouni, M., Pickard, M.R., Hedge, V.L., Farzaneh, F. and Williams, G.T. (2009) GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene*, 28, 195-208.
2. Volders, P.J., Helsens, K., Wang, X., Menten, B., Martens, L., Gevaert, K., Vandesompele, J. and Mestdagh, P. (2013) LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res*, 41, D246-251.
3. Lu, K.H., Li, W., Liu, X.H., Sun, M., Zhang, M.L., Wu, W.Q., Xie, W.P. and Hou, Y.Y. (2013) Long non-coding RNA MEG3 inhibits NSCLC cells proliferation and induces apoptosis by affecting p53 expression. *BMC Cancer*, 13, 461.
4. Martinez, O. and Reyes-Valdes, M.H. (2008) Defining diversity, specialization, and gene specificity in transcriptomes through information theory. *Proc Natl Acad Sci U S A*, 105, 9709-9714.