

## Shen et al. Supplemental Materials












Table S1 Parameters of forward-backward algorithm

Symbol	Description
N	Number of states
$S = \{S_1, S_2, \dots, S_N\}$	Set of states
M	Number of symbols can be observed at each state
$V = \{v_1, v_2, \dots, v_M\}$	Set of observable discrete symbols
$A = \{a_{ij}\}$	The probability from state $S_i$ transferring to $S_j$
$B = \{b_j(k)\}$	Probability distribution of each observable symbols at state $S_j$
$\pi = \{\pi_i\}$	Initial distribution of state
T	Length of observed sequence
$D = \{D_i\}$	The largest sustained time in the state $S_i$
$P = \{p_i(d)\}$	The probability of the sustained time to be d in the state $S_i$

Table S2 The signal state parameters

Signal state	State length	Signal region	A	T	C	G
NUE	6	269-288	Obtained from matrix of transition probabilities			
CS	2	299-300	Obtained from matrix of transition probabilities			
FUE	8	169-268	0.20806	0.43528	0.16643	0.19023
CEL	6	289-298	0.20910	0.42750	0.17137	0.19203
CER	7	301-310	0.26262	0.37313	0.17679	0.18746

Table S4 The *cis*-element logos for Arabidopsis

Region	Signal element	Name	# of hexamers	Top hexamer	% of Hits
		FUE.1	24	AAUGUA	99
		FUE.2	10	UAGUAG	85
		FUE.3	8	UGAUUC	91
-150/-36		FUE.4	3	CAAGUU	33
		FUE.5	17	UAAUAA	89
		FUE.6	24	UUCUUC	99
		FUE.7	4	UUAAGU	25
		FUE.8	3	CACUUG	14
-35/-10		NUE.1	25	AAUAAA	53
-10/+10		CE.1	2	CUUUUG	4
		CE.2	2	AAAAAG	5

**Table S5. Distance of poly(A) sites in 55K to the annotated poly(A) sites\*.**

Distance (nt)	Number of poly(A) sites	%
0-10	12794	27.56
11-20	5189	11.18
21-30	4103	8.84
31-50	5497	11.84
51-100	8888	19.15
101-500	7761	16.72
501-1000	720	1.55
1000-10000	1392	3.00
>10000	80	0.17
Total	46424	100

\*The Build 3 of the rice genome was used.

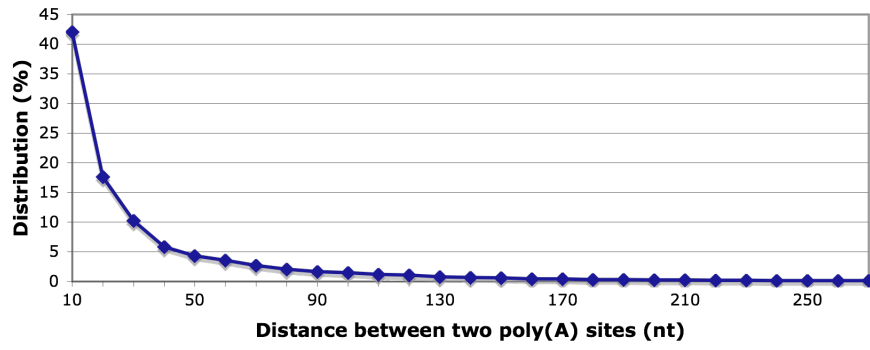


Figure S1. Distribution of the distances between poly(A) sites of the transcripts that undergo APA.

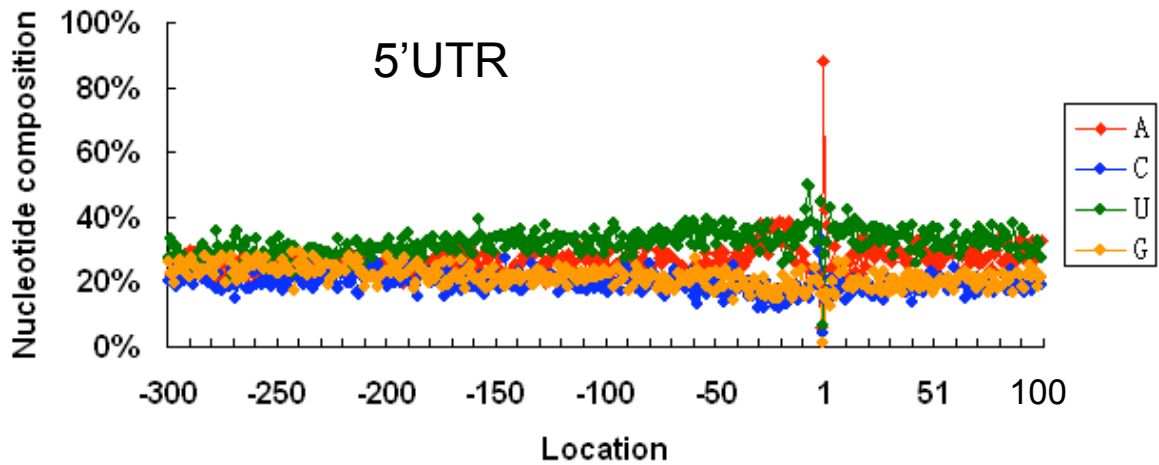
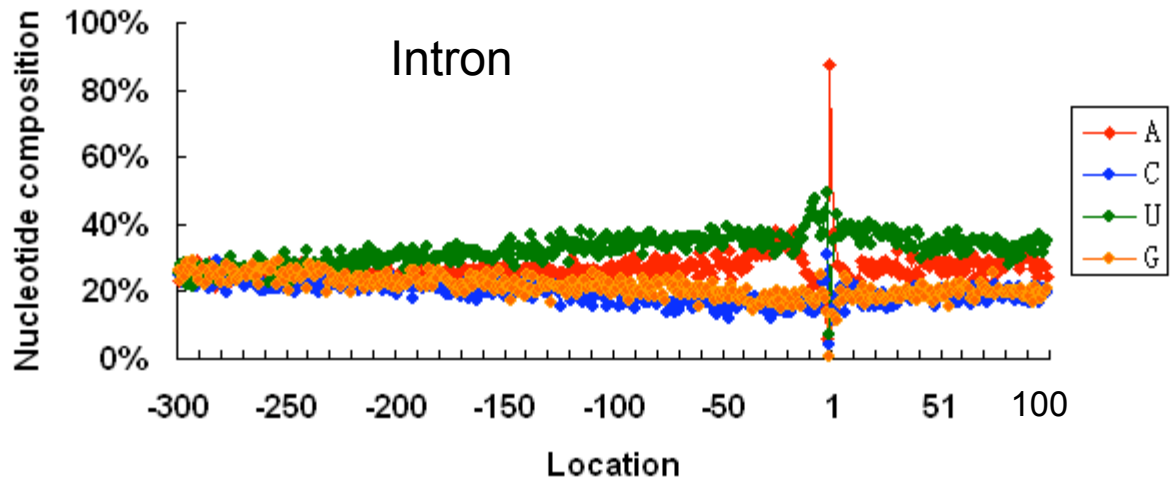


Figure S2. Single nucleotide profiles around the APA sites located in the 5'-UTR and intron regions, respectively. The location designation is the same as Figure 1. Each of the profiles was drawn from about 250 sequences.

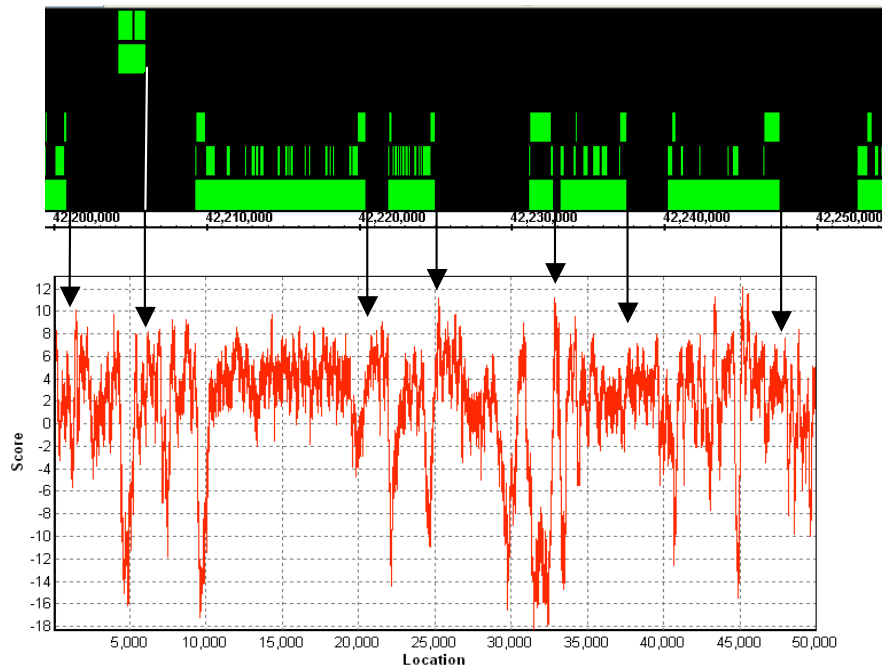


Figure S3. PASS-Rice predictions match the ends of the annotated genes in rice genome. The genome segment was randomly chosen from the Build 3 genome annotation and visualized by using Integrated Genome Browser (<http://www.Affymetrix.com>). The arrows are connecting the ends of the annotated genes and the peaks of PASS-Rice outputs. Genes shown here are (from left to right) Os01g0957600, Os01g0957800, Os01g0957900, Os01g0958000, Os01g0958200, Os01g0958400, Os01g0958500.