Shen et al., Supplemental Data S3

Relationship of APA and the expression levels of *trans*-acting factors

It is well known that some polyadenylation-related protein factors play crucial roles in 3'-end formation (Simpson et al. 2003; Xing et al. 2008; Zhang et al. 2008), and mRNA splicing is also known to play a role in polyadenylation (Millevoi et al. 2006; Tian et al. 2007). To further determine the basis for the different usage of APA sites in different tissue types, we examined the expression level of known polyadenylation factors and splicing factors in each MPSS-DGE library. The hypothesis was that different expression levels of these *trans*-acting factors might correspond to different APA levels among different libraries. Since Arabidopsis has more libraries showing significant differences in APA site usages (Supplemental Data S2) and the polyadenylation factors are better understood than those in rice, we only used the polyadenylation factors of Arabidopsis in this study. We did not use signatures in the five oldest libraries sequenced by the "classic" MPSS-DGE method to make sure that comparisons between libraries are not affected by sequencing methods. In 12 libraries, the TPM values were used as the expression levels of genes involved. The 24 known Arabidopsis polyadenylation factors (Hunt et al. 2008a) and expression value of 19 splicing factors (Campbell et al. 2006) were examined in the same 12 libraries obtained from Arabidopsis MPSS-DGE databases.

To find genes that are critical to the library-specific behavior of APA, we used a Spearman correlation to calculate the correlation of each gene's expression value with different usage of APA. We further categorized APA into two conditions, APA found in exons (classes 4, 5 and 6), and APA associated with alternative splicing (classes 6 and 7). In the first case, several polyadenylation genes *PCFS1, U2AF65, PAPS4* and *PABN3,* were found to have a positive correlation with the increased usage of library-specific APA (Supplemental Table S1; detail expression value of each genes are in Supplemental Table S2). Meanwhile *CFIS25*, *SYM2* and *PABN2* have positive APA effect in introns. This result correlates with some previous findings using traditional experimental methods. For example, *PCFS4*, a homologue of yeast polyadenylation factor Protein 1 of Cleavage Factor 1 (*Pcf11p*), is the only gene on the list significantly correlated with APA (Pearson correlation test, $p<0.05$) in exons and introns and is also known to regulate flowering time through regulation of APA (Xing et al. 2008). The splicing factor U2AF65 was recently associated with the splicing of the last exon and polyadenylation, through its interaction with polyadenylation factor CFI-m in mammals (Millevoi et al. 2006). Moreover, the homologue of CFI-m in Arabidopsis, *CFIS25*, may also be implicated here as it demonstrates a positive correlation with APA in introns.

It was surprising to see there were no genes on the list for which their expression level is significantly correlated with APA (Pearson correlation test, $p<0.05$) in both exons and introns (Supplemental Table S2). This result suggests a possibility that these events could be mutually exclusive, and the expression of, rather than lack thereof, certain genes may promote APA in some tissues. While the result may also validate our analysis protocols, the correlations warrant further studies to reveal more specific relationships between polyadenylation and splicing factors and particular library preference towards APA.

Method

To study which *trans*-acting factors play a role in determination of usage of APA in Arabidopsis, we identified known polyadenylation and splicing factors (Campbell et al. 2006; Hunt et al. 2008b). For each library, expression levels of each factor were determined by the total number of TPMs (without considering the number of signatures each gene has) and expression values from all libraries were compared with relative distance calculated in the previous section. Spearman correlation values for each gene were calculated by comparing the expression value of the gene with the preference of APA discussed in previous paragraph. We classified the APA preference into two groups: an exon-related group that includes signatures from APA classes 4, 5 and 6 and an intron-related group from APA classes 6 and 7. Genes with p-values < 0.1 were considered significant.

Table S1. List of polyadenylation or splicing related proteins found to be correlated with the usage of APA in exons or introns. Library specific expression value of all known polyadenylation or splicing factors are compared with APA-class usages in same library (see Figure 3) using Spearman correlation method.

Gene ID	Name	*p*-value-exon[a]	*rho*-value-exon[b]	*p*-value-Intron[a]	*rho*-value-Intron[b]
AT4G32850	*PAPS4*	0.068	0.545	0.044	0.594
AT4G36690	*U2AF65*	0.017	0.670	-	-
AT1G66500	*PCFS1*	0.075	0.531	-	-
AT5G10350	*PABN3*	0.077	0.531	-	-
AT5G51120	*PABN2*	-	-	0.009	0.708
AT1G27595	*SYM2*	-	-	0.056	0.562
AT4G25550	*CFIS2*	-	-	0.072	0.535

[a]*p*-value and [b]*rho*-value (Spearman ranking) were calculated based on expression values of the protein coding genes and APA usages in the 12 Arabidopsis libraries using Spearman correlation methods.

Supplemental Table S2. The expression values (tag numbers) of polyadneylation and splicing related genes in Arabidopsis associated with each library.

Gene ID	Name	AP1	AP3	AGM	INS	ROS	SAP	S04	S52	LES	GSE	CAS	SIS	p-value	rho-exon	pvalue-intro	rho-intron
AT5G51660	CPSF160	26	22	11	25	15	1	51	9	48	17	25	16	0.672054	-0.1366027	0.556033	-0.189142
AT5G23880	CPSF100,ESP5	61	33	28	41	63	36	24	2	10	5	26	9	0.450277	0.2377622	0.477883	0.223776
AT1G61010	CPSF73-I	16	24	19	54	28	20	3	7	7	1	18	15	0.541214	0.1961474	0.533867	0.19965
AT1G30460	CPSF30,OXT6	5	16	18	16	1	31	0	0	4	0	24	0	0.587049	0.1747216	0.211804	0.388667
AT1G17760	CSTF77	17	12	27	21	17	48	4	20	38	12	39	29	0.786327	-0.0877198	0.695662	0.126317
AT1G71800	CSTF64	0	0	0	0	0	0	0	0	0	0	22	0	0.684849	-0.1310139	0.206256	0.393042
AT5G60940	CSTF50	40	56	22	27	24	13	14	0	48	2	64	56	0.931114	-0.0280211	0.504913	0.213661
AT5G13480	FY	0	4	0	2	5	0	0	0	0	0	2	5	0.799433	0.0822373	0.485576	0.223216
AT1G17980	PAPS1	6	6	10	6	38	8	3	5	6	0	17	0	0.732356	0.1105382	0.526366	0.203248
AT2G25850	PAPS2	7	37	17	29	24	23	0	18	61	0	35	12	0.347292	0.2977237	0.854155	0.059545
AT4G32850	PAPS4	44	75	46	74	139	107	41	17	64	49	174	29	0.068633	0.5454545	0.044262	0.594406
AT3G66652	FIPS3	12	19	9	6	9	0	0	0	0	0	0	0	0.518389	0.2070976	0.74418	0.105503
AT4G25550	CFIS2	33	36	6	28	60	59	0	0	3	0	78	45	0.539027	0.197188	0.072939	0.535225
AT4G29820	CFIS1	0	0	0	0	0	0	0	0	0	0	6	0	0.684849	-0.1310139	0.206256	0.393042
AT3G04680	CLPS3	1	4	0	16	15	5	0	0	16	0	26	3	0.387789	0.2745626	0.387789	0.274563
AT1G66500	PCFS1	0	8	0	5	0	0	0	5	10	14	0	0	0.075394	0.5314203	0.856417	0.058613
AT4G04885	PCFS4	14	6	2	9	10	0	0	0	17	1	5	22	0.810877	0.0774667	0.930749	0.02817
AT5G43620	PCFS5	0	0	0	0	0	0	0	0	0	0	0	4	0.333893	-0.3056992	0.8928	-0.043671
AT5G10350	PABN3	48	113	53	125	132	14	6	63	54	80	64	8	0.07708	0.5314685	0.313863	0.314685
AT5G51120	PABN2	32	47	57	41	29	55	13	19	13	52	74	32	0.518551	0.2070188	0.009865	0.708776
AT5G65260	PABN1	5	4	15	15	59	4	0	37	8	13	11	26	0.63974	-0.1508781	0.679528	-0.133334
AT5G01400	ESP4,SYM5	27	45	40	64	61	86	58	14	19	6	160	22	0.86898	-0.0559441	0.477883	0.223776
AT1G27590	SYM1	0	0	0	5	0	0	0	0	0	0	0	0	0.8928	-0.0436713	0.684849	-0.131014
AT1G27595	SYM2	38	35	8	35	11	45	0	0	38	20	170	66	0.461214	0.2355042	0.056984	0.562398
AT5G09880	PUF60I	18	16	13	3	1	15	0	0	2	0	0	1	0.356407	0.2923913	0.634252	0.153327
AT5G51660	PUF60II	26	22	11	25	15	1	51	9	48	17	25	16	0.672054	-0.1366027	0.556033	-0.189142
AT5G11170	UAP56I	5	6	9	11	1	2	0	43	3	22	0	0	0.375065	0.2816971	0.913488	-0.035212
AT5G11200	UAP56II	2	6	4	9	0	0	0	0	3	9	6	0	0.276412	0.342092	0.175728	0.418517
AT5G51300	SFI	54	92	73	90	18	32	2	9	21	31	68	26	0.410431	0.2587413	0.112478	0.482518
AT1G14650	SF3AI	20	81	40	50	46	14	3	21	27	12	33	50	0.956919	0.0175132	0.93971	0.024518
AT1G14640	SF3AIII	0	0	0	0	0	0	0	24	0	0	0	0	0.495367	-0.2183566	0.206256	-0.393042
AT2G32600	SF3A, subunit 2	19	30	49	46	20	27	23	0	76	5	86	23	0.664076	-0.1401053	0.974145	0.010508
AT5G06160	SF3A60	15	42	26	51	19	56	6	25	11	3	23	14	0.774731	0.0909091	0.724465	0.111888
AT5G64270	SF3B, subunit 1	216	103	95	140	87	53	52	12	200	99	318	59	0.360342	0.2867133	0.165145	0.426573
AT4G21660	SF3B, subunit 2	92	131	100	73	89	31	14	21	36	37	73	145	0.871178	0.0525395	0.341372	0.301226
AT5G40340	SRm300	81	140	100	148	33	123	7	4	46	0	86	24	0.643088	0.1468531	0.535464	0.195804
AT3G50670	U1 70K	159	165	216	169	155	120	75	70	135	66	108	99	0.92992	0.027972	0.886335	-0.048951
AT2G43370	U1 70K	0	1	0	1	2	1	0	0	0	0	0	0	0.094565	0.5042729	0.453361	0.23953
AT1G60900	U2AF65	45	101	107	114	43	21	0	5	43	0	109	85	0.616271	-0.1614045	0.671506	0.136843
AT4G36690	U2AF65	35	40	4	38	0	25	0	15	37	50	25	2	0.01709	0.6701796	0.10773	0.487722
AT1G27650	U2AF35	139	80	83	109	167	29	18	113	70	18	223	6	0.974145	0.0105079	0.736944	0.108582
AT5G42820	U2AF35	134	156	166	107	251	48	28	57	89	89	136	85	0.497785	0.2171632	0.347292	0.297724
AT1G10320	U2AF35	4	4	1	9	5	0	0	0	2	0	0	0	0.404218	0.2655249	0.826308	-0.071056