

iScience, Volume 25

Supplemental information

**DeepASmRNA: Reference-free prediction
of alternative splicing events with a scalable
and interpretable deep learning model**

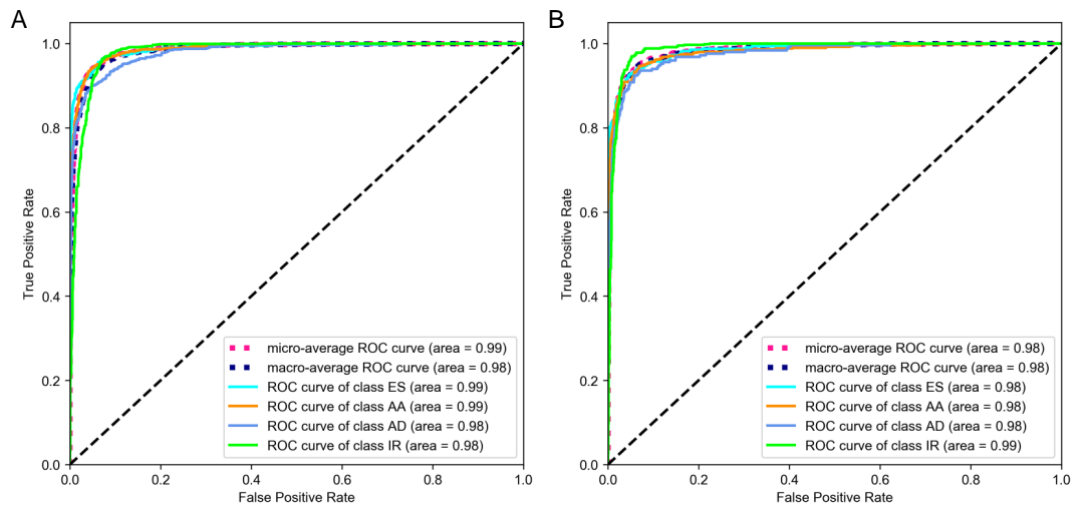
Lei Cao, Quanbao Zhang, Hongtao Song, Kui Lin, and Erli Pang

Supplementary Data Index

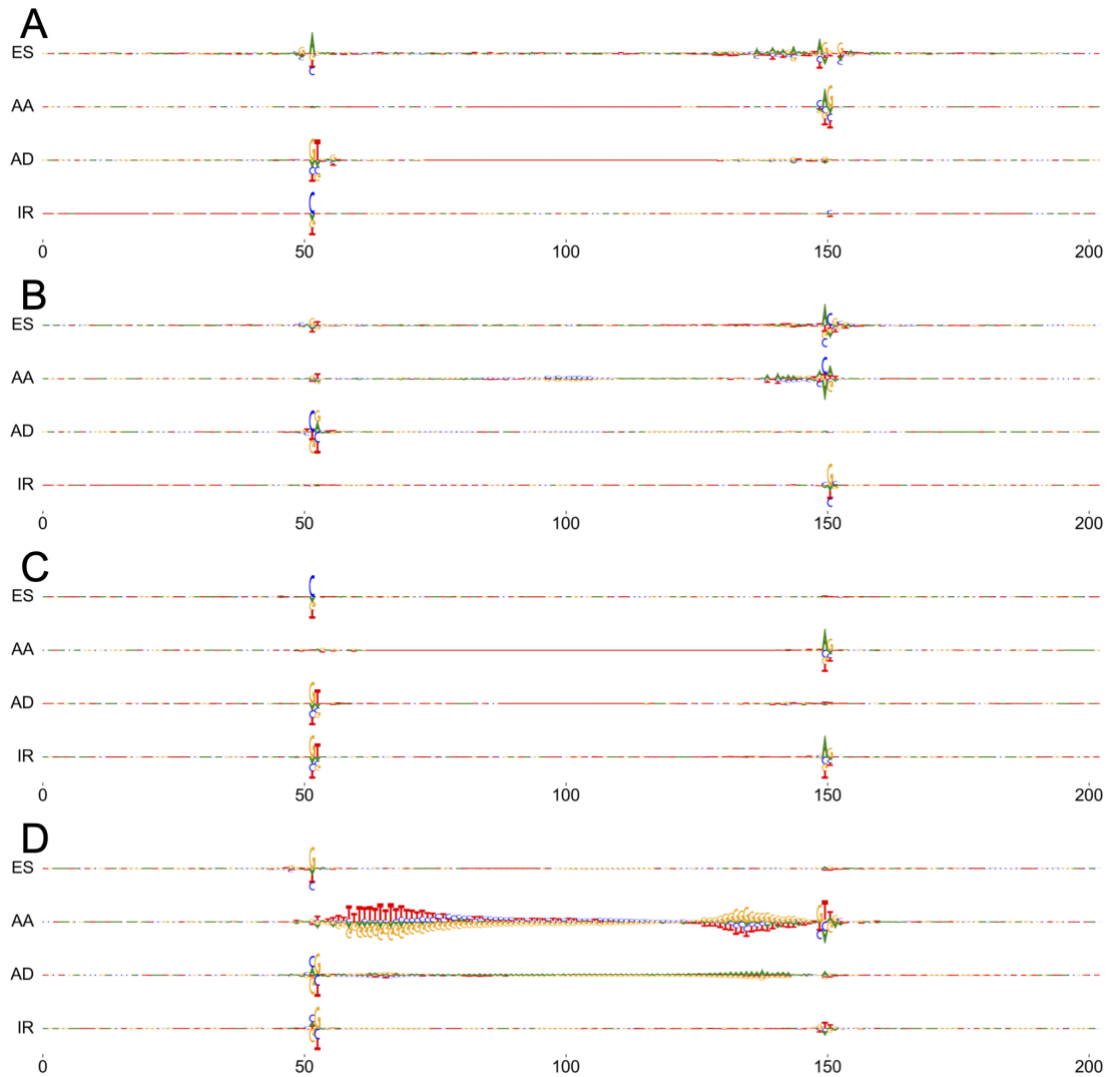
SUPPLEMENTARY FIGURES.....	2
SUPPLEMENTARY TABLES	7

Supplementary Figures

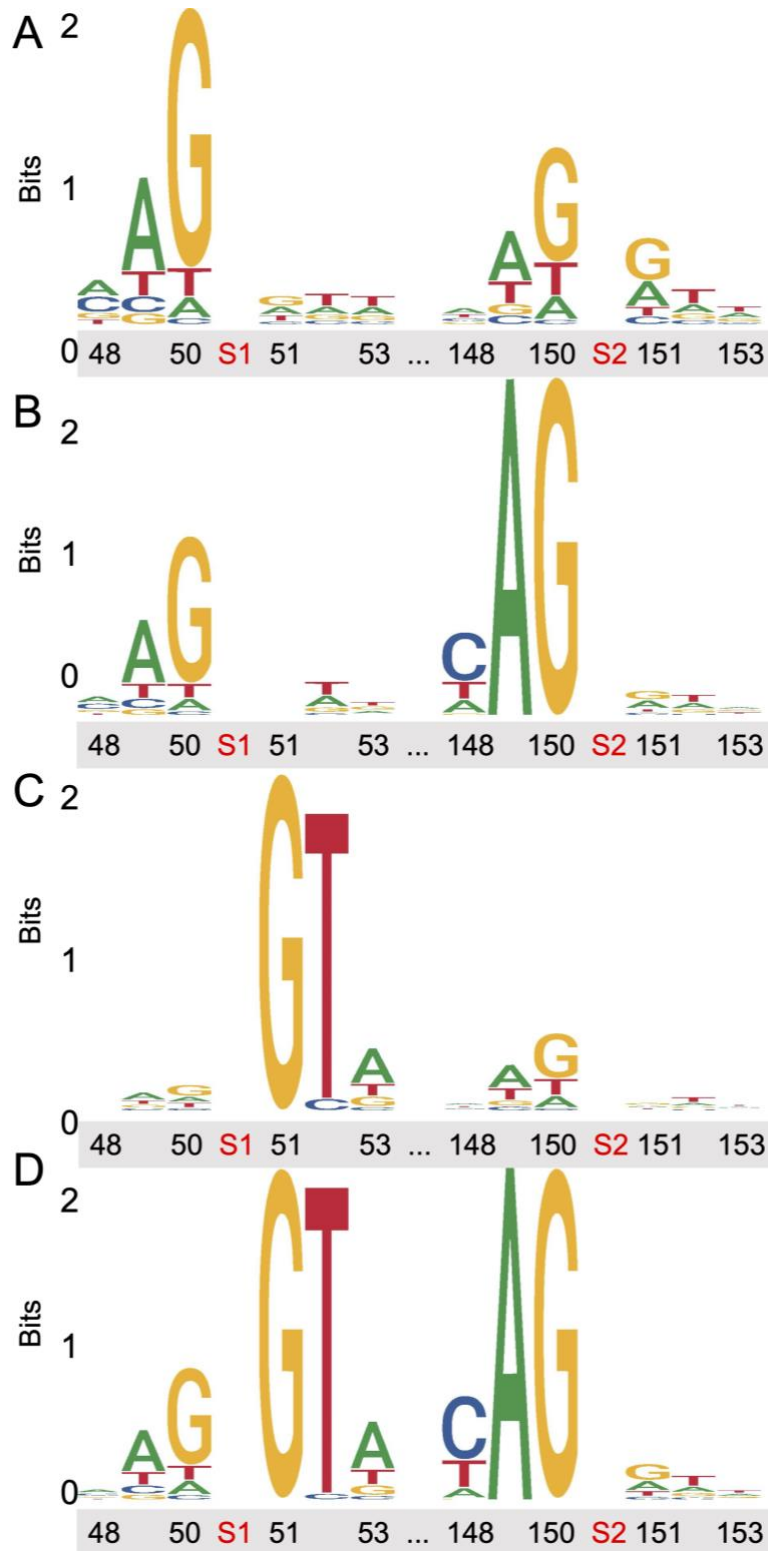
Supplementary Figure S1. The ROC curves (receiver operating characteristic curves) of DeepASmRNA, Related to Figure 4. (A) *Arabidopsis thaliana*. (B) Rice. AUCs (area under the curves) are listed at the bottom right of each subplot. Solid lines with different colors represent four types of AS events, while dotted lines represent micro- and macro-average ROC curves. Micro-average regards each instance equally, while macro-average considers each class equally.



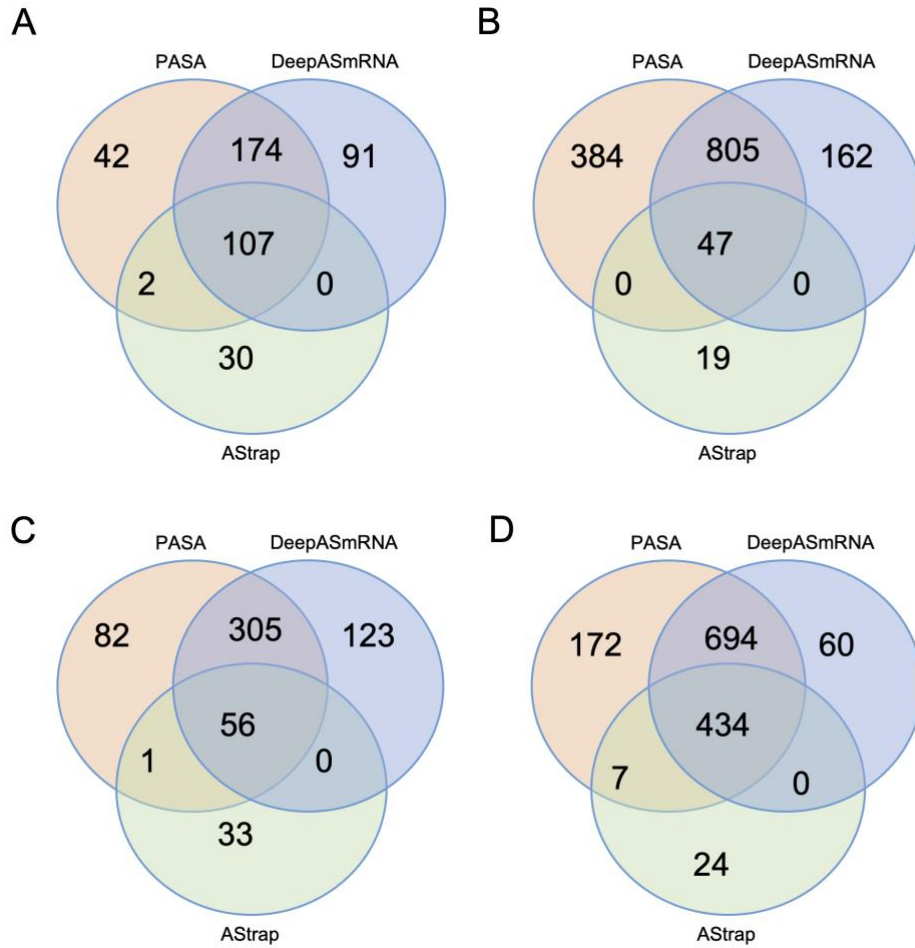
Supplementary Figure S2. The visualization of the precision mask of DeepASmRNA, Related to Figure 5. (A) Cross entropy loss value of human model using 0 mask each base on each position. (B) Cross entropy loss value of human model using 1 mask each base on each position. (C) Cross entropy loss value of *Arabidopsis thaliana* model using 0 mask each base on each position. (D) Cross entropy loss value of *Arabidopsis thaliana* model using 1 mask each base on each position. The x-axis represents the position to be masked, and the y-axis indicates the mean of scaled importance scores of the top ten sequences with the lowest loss in each class of AS event.



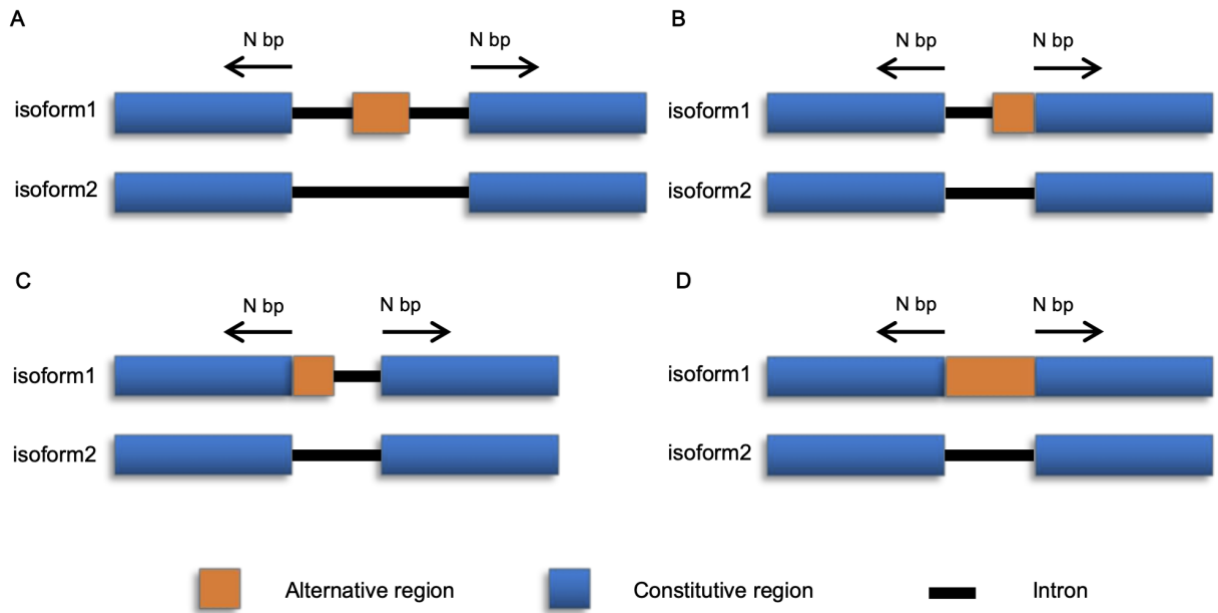
Supplementary Figure S3. Sequence logo of four types of AS events, Related to Figure 5. (A) Exon skipping (ES). (B) Alternative acceptor site (AA). (C) Alternative donor site (AD). (D) Intron retention (IR). The X-axis represents the position of the sequence. Positions S1 and S2 are splicing sites. Three bases are taken upstream and downstream of each of the two splicing sites. Y-axis indicates Shannon entropy.



Supplementary Figure S4. Venn graph of AS events of Iso-Seq data from *Amborella* identified by DeepASmRNA, AStrap, and PASA, of Iso-Seq data from *Amborella*, Related to Figure 6. (A) Exon skipping (ES). (B) Alternative acceptor site (AA). (C) Alternative donor site (AD). (D) Intron retention (IR).



Supplementary Figure S5. The extracted mRNA transcript sequence, Related to STAR Methods. (A) Exon skipping (ES). (B) Alternative acceptor site (AA). (C) Alternative donor site (AD). (D) Intron retention (IR).



Supplementary Tables

Supplementary Table S1. Information of genomes, Related to STAR Methods

Features	Human	<i>Arabidopsis thaliana</i>	Rice
Genome version	GRCh38.p13	Araport11	MSU7
Annotation version	v37	v11	v7.0
Number of genes	19 951	27 654	55 803
Number of transcripts	159 505	48 358	66 338
Number of genes occurring alternative splicing	17 141	10 695	6459

Supplementary Table S2. Performance of identification of AS transcripts in human, *Arabidopsis thaliana*, and rice using the SUPPA2 datasets as the ground truth, Related to Figure 2

Method	Features	Human	<i>Arabidopsis thaliana</i>	Rice
DeepASmRNA	Predicted AS transcript pairs	204 910	21 747	8 873
	True positive	187 083	20 595	8 269
	False positive	17 827	1 152	604
	False negative	45 825	4 369	2 120
	Precision (%)	91.30	94.70	93.19
	Recall (%)	80.32	82.50	79.59
	F1-score	0.85	0.88	0.86
AStrap	Predicted AS transcript pairs	44 085	9 791	5 247
	True positive	22 081	9 373	5 091
	False positive	22 004	418	156
	False negative	210 827	15 591	5 298
	Precision (%)	50.08	95.73	97.03
	<i>P</i> -value ¹	0	0.5363	0.1080
	Recall (%)	9.48	37.55	49.00
	<i>P</i> -value ²	0	0	1.60e-103
	F1-score	0.16	0.54	0.65
IsoSplitter	Predicted AS transcript pairs	259 915	20 376	8 482
	True positive	125 276	16 641	7 246
	False positive	134 639	3 735	1 236
	False negative	107 632	8 323	3 143
	Precision (%)	48.20	81.67	85.43
	<i>P</i> -value ³	0	3.25e-25	8.64e-05
	Recall (%)	53.79	66.66	69.75
	<i>P</i> -value ⁴	0	2.22e-54	5.28e-10
	F1-score	0.51	0.73	0.77

¹ Fisher's exact test of the precision between DeepASmRNA and AStrap.

² Fisher's exact test of the recall between DeepASmRNA and AStrap.

³ Fisher's exact test of the precision between DeepASmRNA and IsoSplitter.

⁴ Fisher's exact test of the recall between DeepASmRNA and IsoSplitter.

Supplementary Table S3. Performance of DeepASmRNA in AS classification for different sequence length, Related to Figure 4

Species	Input Segment Length (bp)	10	20	30	40	50	60
Human	Accuracy	83.59	86.25	87.57	88.00	88.49	88.48
	Micro-average AUC	0.97	0.98	0.98	0.98	0.98	0.98
	Macro-average AUC	0.97	0.98	0.98	0.98	0.98	0.98
	ES accuracy	87.66	90.00	90.86	90.87	91.32	91.46
	ES precision	87.78	90.53	91.64	89.85	90.61	92.02
	ES recall	85.02	87.41	88.20	90.37	90.52	89.18
	ES AUC	0.95	0.97	0.97	0.97	0.97	0.97
	ES F1-score	0.86	0.89	0.90	0.90	0.91	0.91
	ES AUPRC	0.95	0.96	0.97	0.97	0.97	0.97
	AA accuracy	90.71	92.50	93.21	93.29	93.54	93.70
	AA precision	76.96	82.55	82.83	83.61	83.93	83.31
	AA recall	75.20	78.40	82.52	81.87	83.01	84.95
	AA AUC	0.95	0.97	0.97	0.97	0.97	0.97
	AA F1-score	0.76	0.80	0.83	0.83	0.83	0.84
	AA AUPRC	0.84	0.89	0.91	0.91	0.91	0.91
	AD accuracy	93.36	93.77	94.45	94.87	95.19	94.91
	AD precision	78.88	78.20	80.46	84.84	85.45	83.29
	AD recall	84.54	89.08	90.07	85.93	87.29	88.58
	AD AUC	0.98	0.98	0.98	0.98	0.99	0.99
	AD F1-score	0.82	0.83	0.85	0.85	0.86	0.86
	AD AUPRC	0.90	0.92	0.93	0.93	0.94	0.94
	IR accuracy	95.45	96.23	96.61	96.96	96.92	96.89
	IR precision	85.20	88.49	90.66	91.21	91.16	90.98
	IR recall	88.46	89.28	89.14	90.78	90.54	90.57
IR AUC	0.98	0.99	0.99	0.99	0.99	0.99	
IR F1-score	0.87	0.89	0.90	0.91	0.91	0.91	
IR AUPRC	0.95	0.96	0.96	0.97	0.97	0.97	
<i>Arabidopsis thaliana</i>	Accuracy	87.52	89.01	89.66	90.44	90.73	90.60
	Micro-average AUC	0.98	0.98	0.98	0.98	0.99	0.99
	Macro-average AUC	0.97	0.98	0.98	0.98	0.98	0.98
	ES accuracy	96.11	96.60	96.99	97.38	97.38	97.50
	ES precision	90.86	87.16	90.03	94.77	91.99	93.18
	ES recall	78.33	86.95	86.70	84.73	87.68	87.44
	ES AUC	0.98	0.99	0.98	0.99	0.99	0.99

	ES F1-score	0.84	0.87	0.88	0.89	0.90	0.90
	ES AUPRC	0.92	0.94	0.94	0.96	0.96	0.95
	AA accuracy	93.78	94.07	95.01	95.27	95.11	95.11
	AA precision	88.11	88.94	90.87	92.02	90.99	89.60
	AA recall	89.55	89.67	91.08	90.73	91.31	93.08
	AA AUC	0.98	0.98	0.98	0.98	0.99	0.99
	AA F1-score	0.89	0.89	0.91	0.91	0.91	0.91
	AA AUPRC	0.95	0.96	0.96	0.97	0.97	0.97
	AD accuracy	93.62	94.65	94.26	94.88	94.91	94.82
	AD precision	90.96	90.75	91.58	94.15	92.16	94.29
	AD recall	76.91	82.68	79.72	80.34	82.53	79.88
	AD AUC	0.97	0.98	0.97	0.98	0.98	0.98
	AD F1-score	0.83	0.87	0.85	0.87	0.87	0.86
	AD AUPRC	0.93	0.94	0.93	0.94	0.94	0.95
	IR accuracy	91.54	92.71	93.07	93.36	94.07	93.78
	IR precision	84.86	88.85	87.92	86.75	89.53	88.96
	IR recall	94.95	92.67	95.03	97.64	95.79	95.70
	IR AUC	0.97	0.98	0.98	0.98	0.98	0.98
	IR F1-score	0.90	0.91	0.91	0.92	0.93	0.92
	IR AUPRC	0.94	0.95	0.95	0.96	0.96	0.96
Rice	Accuracy	86.80	88.29	89.31	89.78	90.02	89.96
	Micro-average AUC	0.97	0.98	0.98	0.98	0.98	0.98
	Macro-average AUC	0.97	0.98	0.98	0.98	0.98	0.98
	ES accuracy	95.22	94.62	95.16	94.74	95.40	95.40
	ES precision	89.56	84.29	84.11	83.61	88.22	92.31
	ES recall	85.76	89.39	93.03	91.21	88.48	83.64
	ES AUC	0.97	0.98	0.98	0.98	0.98	0.99
	ES F1-score	0.88	0.87	0.88	0.87	0.88	0.88
	ES AUPRC	0.94	0.95	0.96	0.95	0.96	0.96
	AA accuracy	92.35	92.89	93.67	94.15	94.44	93.91
	AA precision	89.64	93.64	91.79	95.71	93.07	91.00
	AA recall	84.29	81.91	86.68	84.29	88.07	88.47
	AA AUC	0.97	0.98	0.98	0.98	0.98	0.98
	AA F1-score	0.87	0.87	0.89	0.90	0.91	0.90
	AA AUPRC	0.95	0.96	0.96	0.96	0.97	0.96
	AD accuracy	94.50	95.28	95.10	95.64	95.52	95.28
	AD precision	92.93	87.34	92.42	87.66	90.70	84.90
	AD recall	68.40	80.00	73.20	82.40	78.00	83.20
	AD AUC	0.97	0.98	0.98	0.98	0.98	0.98

AD F1-score	0.79	0.84	0.82	0.85	0.84	0.84
AD AUPRC	0.90	0.91	0.91	0.92	0.91	0.92
IR accuracy	91.52	93.79	94.68	95.04	94.68	95.34
IR precision	82.03	87.18	89.47	89.94	88.50	90.02
IR recall	97.29	96.62	96.28	96.79	97.63	97.63
IR AUC	0.98	0.98	0.99	0.99	0.99	0.99
IR F1-score	0.89	0.92	0.93	0.93	0.93	0.94
IR AUPRC	0.94	0.96	0.97	0.97	0.97	0.98

Supplementary Table S4. Performance in AS classification of DeepASmRNA and AStrap based on AStrap's Dataset and DeepASmRNA's Dataset, Related to Figure 4

Dataset source	Model	Species	Overall Accuracy (%)	Precision (%)			
				ES	AA	AD	IR
AStrap	AStrap	Human	86.5	85.9	87.5	85.2	87.6
		Rice	88.3	91.7	90.2	86.0	85.9
	DeepASmRNA	Human	87.84	90.96	80.35	78.47	94.31
		Rice	91.17	80.11	92.72	86.14	93.04
DeepASmRNA	AStrap	Human	81.44	94.99	77.00	70.61	66.15
		Rice	90.11	72.41	96.10	82.20	95.43
	DeepASmRNA	Human	88.49	90.61	83.93	85.45	91.16
		Rice	90.02	88.22	93.07	90.7	88.5

Supplementary Table S5. Comparison of DeepASmRNA, LeNet, ResNet, and LSTM in classification, Related to Figure 4

Species	Models	LeNet	ResNet	LSTM	DeepASmRNA
Human	Accuracy	88.07	85.57	88.12	88.49
	Micro-average AUC	0.98	0.97	0.98	0.98
	Macro-average AUC	0.98	0.97	0.98	0.98
	ES accuracy	90.91	89.63	91.32	91.32
	ES precision	89.37	89.62	89.72	90.61
	ES recall	91.09	87.61	91.65	90.52
	ES AUC	0.97	0.96	0.97	0.97
	ES F1-score	0.90	0.89	0.91	0.91
	ES AUPRC	0.97	0.96	0.97	0.97
	AA accuracy	93.42	91.96	93.44	93.54
	AA precision	84.51	77.88	88.81	83.93
	AA recall	81.43	82.46	76.23	83.01
	AA AUC	0.97	0.96	0.97	0.97
	AA F1-score	0.83	0.80	0.82	0.83
	AA AUPRC	0.91	0.88	0.92	0.91
	AD accuracy	94.89	94.17	94.65	95.19
	AD precision	84.68	82.91	82.60	85.45
	AD recall	86.33	83.85	87.82	87.29
	AD AUC	0.98	0.98	0.98	0.99
	AD F1-score	0.85	0.83	0.85	0.86
	AD AUPRC	0.93	0.92	0.93	0.94
	IR accuracy	96.91	95.39	96.83	96.92
	IR precision	92.11	87.05	89.04	91.16
	IR recall	89.34	85.42	92.66	90.54
	IR AUC	0.99	0.98	0.99	0.99
	IR F1-score	0.91	0.86	0.91	0.91
	IR AUPRC	0.97	0.94	0.96	0.97
	<i>Arabidopsis thaliana</i>	Accuracy	89.89	88.72	88.98
Micro-average AUC		0.98	0.98	0.98	0.99
Macro-average AUC		0.98	0.98	0.98	0.98
ES accuracy		97.15	96.08	96.63	97.38
ES precision		91.84	94.67	89.32	91.99
ES recall		85.96	74.38	84.48	87.68
ES AUC		0.99	0.98	0.98	0.99
ES F1-score		0.89	0.83	0.87	0.90
ES AUPRC		0.95	0.93	0.93	0.96
AA accuracy		94.91	94.39	93.88	95.11
AA precision	89.90	87.85	85.92	90.99	

	AA recall	91.90	92.49	93.08	91.31
	AA AUC	0.98	0.98	0.98	0.99
	AA F1-score	0.91	0.9	0.89	0.91
	AA AUPRC	0.97	0.96	0.96	0.97
	AD accuracy	94.20	93.62	93.97	94.91
	AD precision	92.62	83.64	94.52	92.16
	AD recall	78.32	86.12	75.35	82.53
	AD AUC	0.97	0.97	0.97	0.98
	AD F1-score	0.85	0.85	0.84	0.87
	AD AUPRC	0.93	0.93	0.94	0.94
	IR accuracy	93.52	93.36	93.49	94.07
	IR precision	88.17	90.58	88.88	89.53
	IR recall	96.04	92.33	94.95	95.79
	IR AUC	0.98	0.98	0.98	0.98
	IR F1-score	0.92	0.91	0.92	0.93
	IR AUPRC	0.96	0.96	0.96	0.96
Rice	Accuracy	89.01	87.93	88.77	90.02
	Micro-average AUC	0.98	0.98	0.98	0.98
	Macro-average AUC	0.98	0.98	0.97	0.98
	ES accuracy	95.28	94.15	95.34	95.40
	ES precision	84.20	81.69	90.65	88.22
	ES recall	93.64	90.61	85.15	88.48
	ES AUC	0.99	0.98	0.98	0.98
	ES F1-score	0.89	0.86	0.88	0.88
	ES AUPRC	0.96	0.95	0.94	0.96
	AA accuracy	93.55	92.77	93.25	94.44
	AA precision	89.74	92.07	88.39	93.07
	AA recall	88.67	83.1	89.26	88.07
	AA AUC	0.98	0.97	0.97	0.98
	AA F1-score	0.89	0.87	0.89	0.91
	AA AUPRC	0.96	0.95	0.95	0.97
	AD accuracy	94.5	95.1	94.5	95.52
	AD precision	84.05	88.53	85.91	90.70
	AD recall	78.00	77.20	75.60	78.00
	AD AUC	0.96	0.97	0.96	0.98
	AD F1-score	0.81	0.82	0.80	0.84
	AD AUPRC	0.89	0.91	0.85	0.91
	IR accuracy	94.68	93.85	94.44	94.68
	IR precision	93.43	88.36	89.15	88.50
	IR recall	91.37	95.09	95.94	97.63
	IR AUC	0.99	0.98	0.98	0.99

IR F1-score	0.92	0.92	0.92	0.93
IR AUPRC	0.97	0.96	0.96	0.97

Supplementary Table S6. Performance of transfer learning, Related to STAR Methods

Model	Before transfer learning				After fine tune			
	<i>Arabidopsis thaliana</i>	Rice	Human	<i>Arabidopsis thaliana</i>	<i>Arabidopsis thaliana</i>	Rice	Human	<i>Arabidopsis thaliana</i>
Target	Rice	<i>Arabidopsis thaliana</i>	<i>Arabidopsis thaliana</i>	Human	Rice	<i>Arabidopsis thaliana</i>	<i>Arabidopsis thaliana</i>	Human
accuracy	89.67	88.50	72.13	66.50	90.44	88.69	84.96	66.82
micro-average AUC	0.99	0.98	0.91	0.86	0.99	0.98	0.96	0.87
macro-average AUC	0.98	0.98	0.94	0.82	0.98	0.98	0.96	0.83
ES accuracy	94.80	95.85	86.75	73.73	95.28	95.79	94.94	74.26
ES precision	88.33	82.18	49.78	72.77	87.24	82.39	90.85	73.23
ES recall	84.85	87.44	85.47	68.59	89.09	86.45	68.47	69.47
ES AUC	0.99	0.98	0.93	0.80	0.99	0.98	0.95	0.80
ES F1-score	0.87	0.85	0.63	0.71	0.88	0.84	0.78	0.71
ES AUPRC	0.96	0.93	0.83	0.68	0.95	0.92	0.85	0.68
AA accuracy	93.67	94.49	89.44	86.03	94.44	94.62	91.57	86.95
AA precision	88.25	92.10	82.79	61.06	91.84	91.13	79.37	63.99
AA recall	91.05	87.56	77.93	79.75	89.46	89.20	93.90	76.67
AA AUC	0.98	0.98	0.95	0.93	0.98	0.98	0.97	0.93
AA F1-score	0.90	0.90	0.80	0.69	0.91	0.90	0.86	0.70
AA AUPRC	0.97	0.96	0.91	0.80	0.97	0.96	0.94	0.80

AD accuracy	96.00	93.52	86.97	90.92	96.06	93.71	93.26	90.67
AD precision	90.67	92.98	64.14	77.31	90.00	94.43	91.40	74.75
AD recall	81.60	74.41	84.56	67.79	82.80	74.10	74.57	70.18
AD AUC	0.98	0.97	0.95	0.96	0.98	0.97	0.97	0.95
AD F1-score	0.86	0.83	0.73	0.72	0.86	0.83	0.82	0.72
AD AUPRC	0.93	0.91	0.85	0.82	0.93	0.92	0.91	0.80
IR accuracy	94.86	93.13	81.11	82.32	95.10	93.26	90.15	81.78
IR precision	91.19	86.63	90.70	47.50	91.25	87.00	85.35	45.96
IR recall	94.59	97.14	56.70	44.06	95.26	96.97	89.81	44.71
IR AUC	0.99	0.97	0.95	0.61	0.99	0.97	0.95	0.64
IR F1-score	0.93	0.92	0.70	0.46	0.93	0.92	0.88	0.45
IR AUPRC	0.97	0.95	0.87	0.43	0.97	0.95	0.9	0.44

Supplementary Table S7. Transcripts of Iso-Seq data from *Amborella*, Related to STAR Methods

Tissue	Library size (kb)	Read of inserts (ROIs)	Full-length ROIs	ICE-QUIVER isoforms	Polished isoforms	Cd-hit-est	Transcripts produced by annotated genes
leaf	1-2	80 923	45 099	25 274	2523		
leaf	2-3	150 845	61 028	37 084			
leaf	3-6	151 220	45 905	30 083	4248		
flower	1-2	36 098	15 743	11 762	1772		
flower	2-3	61 836	21 983	17 090	1860		
flower	3-6	134 238	33 959	29 103	1969		
all		615 160	223 717	150 396	12 372	10 767	8 918

Supplementary Table S8. Performance of DeepASmRNA of Iso-Seq data from *Amborella* compared with AStrap using the PASA dataset as the ground truth, Related to Figure 6

Type	PASA	DeepASmRNA					AStrap				
		AS events	Overlapping with PASA	Precision (%)	Recall (%)	F1-score	AS events	Overlapping with PASA	Precision (%)	Recall (%)	F1-score
ES	325	372	281	75.54	86.46	0.81	139	109	78.42	33.54	0.47
AA	1236	1014	852	84.02	68.93	0.76	66	47	71.21	3.80	0.07
AD	444	484	361	74.59	81.31	0.78	90	57	63.33	12.84	0.21
IR	1307	1188	1128	94.95	86.30	0.90	465	441	94.84	33.74	0.50
All	3312	3058	2622	85.74	79.17	0.82	760	654	86.05	22.95	0.36
AS transcript pairs	1973	1676	1577	94.09	79.92	0.86	688	675	98.17	34.21	0.51

Supplementary Table S9. Confusion matrix for AS events of Iso-Seq data from *Amborella* identified by DeepASmRNA, AStrap using the PASA dataset as the ground truth, Related to STAR Methods

Type	DeepASmRNA				AStrap			
	ES	AA	AD	IR	ES	AA	AD	IR
ES	281	22	25	13	109	5	13	4
AA	27	852	11	30	8	47	6	9
AD	33	47	361	17	17	3	57	11
IR	31	93	87	1128	5	11	14	441

Supplementary Table S10. Performance of DeepASmRNA before data augmentation in classification, Related to STAR Methods

Model		Human	<i>Arabidopsis thaliana</i>	Rice
Overall accuracy (%)		87.68	90.29	89.12
ES	Precision (%)	91.62	83.33	79.31
	Recall (%)	90.45	71.82	70.99
AA	Precision (%)	83.16	93.65	91.36
	Recall (%)	83.24	90.02	91.74
AD	Precision (%)	85.56	87.58	87.98
	Recall (%)	86.39	85.76	84.70
IR	Precision (%)	81.55	90.28	90.15
	Recall (%)	85.42	95.51	93.94

Supplementary Table S11. The results of the grid search for the criteria of filtering the transcripts in identification, Related to STAR Methods

Species	Identity (%)	Metrics	Coverage (%)				
			50	60	70	80	90
Human	96	Recall	81.72	80.89	80.85	80.55	74.46
		Precision	91.98	92.32	92.43	92.63	93.47
		F1-score	0.87	0.86	0.86	0.86	0.83
	97	Recall	81.31	81.17	80.96	78.57	74.29
		Precision	91.43	91.46	91.94	92.39	93.01
		F1-score	0.86	0.86	0.86	0.85	0.83
	98	Recall	80.98	80.7	80.66	78.78	74.18
		Precision	91.76	91.89	92.17	92.68	93.4
		F1-score	0.86	0.86	0.86	0.85	0.83
	99	Recall	80.32	80.23	80.02	78.65	73.37
		Precision	91.79	92.12	92.31	92.76	93.37
		F1-score	0.86	0.86	0.86	0.85	0.82
	100	Recall	80.09	80.01	79.7	78.3	73.94
		Precision	91.86	92.36	92.58	92.47	93.32
		F1-score	0.86	0.86	0.86	0.85	0.83
<i>Arabidopsis thaliana</i>	96	Recall	83.78	83.67	83.22	83.36	77.13
		Precision	94.29	94.33	94.43	94.64	95.53
		F1-score	0.89	0.89	0.88	0.89	0.85
	97	Recall	83.57	83.43	83.01	81.17	76.88
		Precision	94.32	94.35	94.44	94.66	95.56
		F1-score	0.89	0.89	0.88	0.87	0.85
	98	Recall	83.36	83.21	82.79	80.91	76.57
		Precision	94.36	94.38	94.49	94.69	95.61
		F1-score	0.89	0.88	0.88	0.87	0.85
	99	Recall	83.17	83.01	82.5	80.77	76.28
		Precision	94.62	94.64	94.7	94.91	95.88
		F1-score	0.89	0.88	0.88	0.87	0.85
	100	Recall	82.98	82.86	82.33	80.59	76.07
		Precision	94.64	94.65	94.71	94.93	95.89
		F1-score	0.88	0.88	0.88	0.87	0.85
Rice	96	Recall	81.07	80.8	80.6	80.36	75.03
		Precision	92.66	92.71	92.81	93.44	94.49
		F1-score	0.86	0.86	0.86	0.86	0.84
	97	Recall	81.4	81.22	80.26	78.96	74.17
		Precision	92.74	92.89	92.94	93.5	94.74

	F1-score	0.87	0.87	0.86	0.86	0.83
98	Recall	81.21	80.5	79.99	78.13	74.51
	Precision	92.88	92.95	93.02	94.57	94.61
	F1-score	0.87	0.86	0.86	0.86	0.83
99	Recall	80.97	80.88	79.59	78.13	73.58
	Precision	92.98	93.07	93.19	94.82	95.29
	F1-score	0.87	0.87	0.86	0.86	0.83
100	Recall	80.67	80.04	79.47	78.3	73.81
	Precision	93.01	93.09	93.23	94.84	95.31
	F1-score	0.86	0.86	0.86	0.86	0.83

Supplementary Table S12. Hyperparameters of DeepASmRNA in classification, Related to STAR Methods

Parameters	Value	Explanation or citation
Kernel size of each CNN layer except the last one	12	According to JASPAR (https://jaspar.genereg.net/#), most motifs are less than 12bp in length; we also tried 6 bp and found 12 bp had a better performance.
Stride of each CNN layer	1	The default setting of tensorflow.
Activation of each CNN layer	relu	According to the tutorial of tensorflow (https://tensorflow.google.cn/tutorials/images/cnn), we also tried sigmoid and relu has faster convergence and better performance.
Filters of the first CNN layer	30	Similar to a tutorial of tensorflow (https://tensorflow.google.cn/tutorials/images/cnn).
Filters of the second CNN layer	100	Similar to the tutorial of tensorflow (https://tensorflow.google.cn/tutorials/images/cnn). We tried 64 as they did, but sometimes the model will be trained into a local minimum and get only about 44% accuracy. So, we enlarge this hyper-parameter to make the model have more parameters to train. This value is also chosen based on the API of tensorflow (https://tensorflow.google.cn/versions/r2.1/api_docs/python/tf/keras/layers/Attention)
Pool size of the maxpool layer	2	Same as https://tensorflow.google.cn/tutorials/images/cnn
Stride of the max pool layer	2	Same as https://tensorflow.google.cn/tutorials/images/cnn
Filters of the last CNN layer	200	We kept the filters getting more and more when the layer went to the latter, similar to (https://tensorflow.google.cn/tutorials/images/classification). we also tried 100 as the second layer, and the local minimal happens as the description of filters of the second CNN layer, so we also expand the filters of the last CNN layer to eliminate local minimal.
Kernel size of the last CNN layer	42	This is a global CNN layer, 42 is the input length of this layer.
Units of the first dense layer	32	Same as https://tensorflow.google.cn/tutorials/images/cnn .
Activation of the first dense layer	relu	Same as https://tensorflow.google.cn/tutorials/images/cnn .

A dropout rate of the first dense layer	0.6	We tried 0.5, 0.6, 0.7, 0.8, and 0.9 which have the best performance.
Units of the last dense layer	4	The classes of our model to predict
Activation of the last dense layer	softmax	The most used activation of multi-classifier.
