

Supplementary Material

WITMSG: Large-scale Prediction of Human Intronic m⁶A RNA Methylation Sites from Sequence and Genomic FeaturesLian Liu¹, Xiujuan Lei^{1,*}, Jia Meng² and Zhen Wei^{2,*}¹School of Computer Sciences, Shannxi Normal University, Xi'an, Shaanxi, 710119, China; ²Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, 215123, China

Table S1 Genomic features considered in WITMSG

ID	Name	Description	Note
1	Stop_codons	stop codons 203bp	whether the site is overlapped to the
2	Start_codons	start codons 203bp	topological region on
3	TSS	downstream 100bp of transcription start site	the major RNA transcript.
4	exon_stop	exon with stop codon	
5	length_gene_ex	mature transcript length	The region length in bp.
6	length_gene_full	full transcript length	
7	AAACA	motif --- AAACA	
8	AGACA	motif --- AGACA	
9	AAACT	motif --- AAACT	
10	AGACT	motif --- AGACT	
11	AAACC	motif --- AAACC	
12	AGACC	motif --- AGACC	
13	GAACA	motif --- GAACA	
14	GGACA	motif --- GGACA	
15	GAACT	motif --- GAACT	
16	GGAAT	motif --- GGAAT	
17	GAACC	motif --- GAACC	
18	GGACC	motif --- GGACC	
19	TAACA	motif --- TAACA	
20	TGACA	motif --- TGACA	
21	TAACT	motif --- TAACT	
22	TGACT	motif --- TGACT	
23	TAACC	motif --- TAACC	
24	TGACC	motif --- TGACC	
25	dist_sj_5_p2000	distance to the upstream (5' end) splicing junction	
26	dist_sj_3_p2000	distance to the downstream (3' end) splicing junction	
27	dist_nearest_p2000	distance to the nearest neighbors (peaked at 2000bp)	
28	dist_nearest_p200	distance to the nearest neighbors (peaked at 200bp)	
29	clust_f1000	number of neighbors within 1000bp flanking regions	
30	clust_f100	number of neighbors within 100bp flanking regions	
31	clust_DRACH_f1000	number of DRACH motif neighbors within 1000bp flanking regions is generated.	clustering indicators and motif clustering
32	clust_DRACH_f100	number of DRACH motif neighbors within 100bp flanking regions is generated.	
33	dist_DRACH_p2000	distance to the nearest DRACH motif (peaked at 2000bp)	
34	dist_DRACH_p200	distance to the nearest DRACH motif (peaked at 200bp)	
35	PC_1bp	phast cons scores 1bp	
36	PC_101bp	phast cons scores 101bp	Scores related to evolutionary conservation
37	FC_1bp	fitness consequences scores 1bp z score	
38	FC_101bp	fitness consequences scores 101bp z score	
39	struct_hybridize	predicted hybridized region	RNA secondary structures
40	struct_loop	inferred loop structures between hybridized region	
41	HNRNPC_eCLIP	eCLIP data of HNRNPC RNA binding sites	
42	YTHDC1_TREW	TREW data of YTHDC1 RNA binding sites	
43	YTHDF1_TREW	TREW data of YTHDF1 RNA binding sites	
44	YTHDF2_TREW	TREW data of YTHDF2 RNA binding sites	
45	METTL3_TREW	TREW data of METTL3 RNA binding sites	
46	METTL14_TREW	TREW data of METTL14 RNA binding sites	
47	WTAP_TREW	TREW data of WTAP RNA binding sites	
48	METTL16_CLIP	CLIP data of METTL16 RNA binding sites	
49	FTO_CLIP	CLIP data of FTO RNA binding sites	
50	FTO_eCLIP	eCLIP data of FTO RNA binding sites	
51	TargetScan	Predicted miRNA targeted sites by TargetScan	
52	Verified_miRtargets	miRNA targeted sites verified by experiment	
53	miR_targeted_genes	miRNA targeted genes	
54	HK_genes	house keeping genes	Attributes of the genes or transcripts.
55	isoform_num	isoform number z score	Isoform number z score
56	GC_cont_101bp	101bp GC content z score	
57	GC_cont_genes	gene level GC content z score is generated.	GC content z score