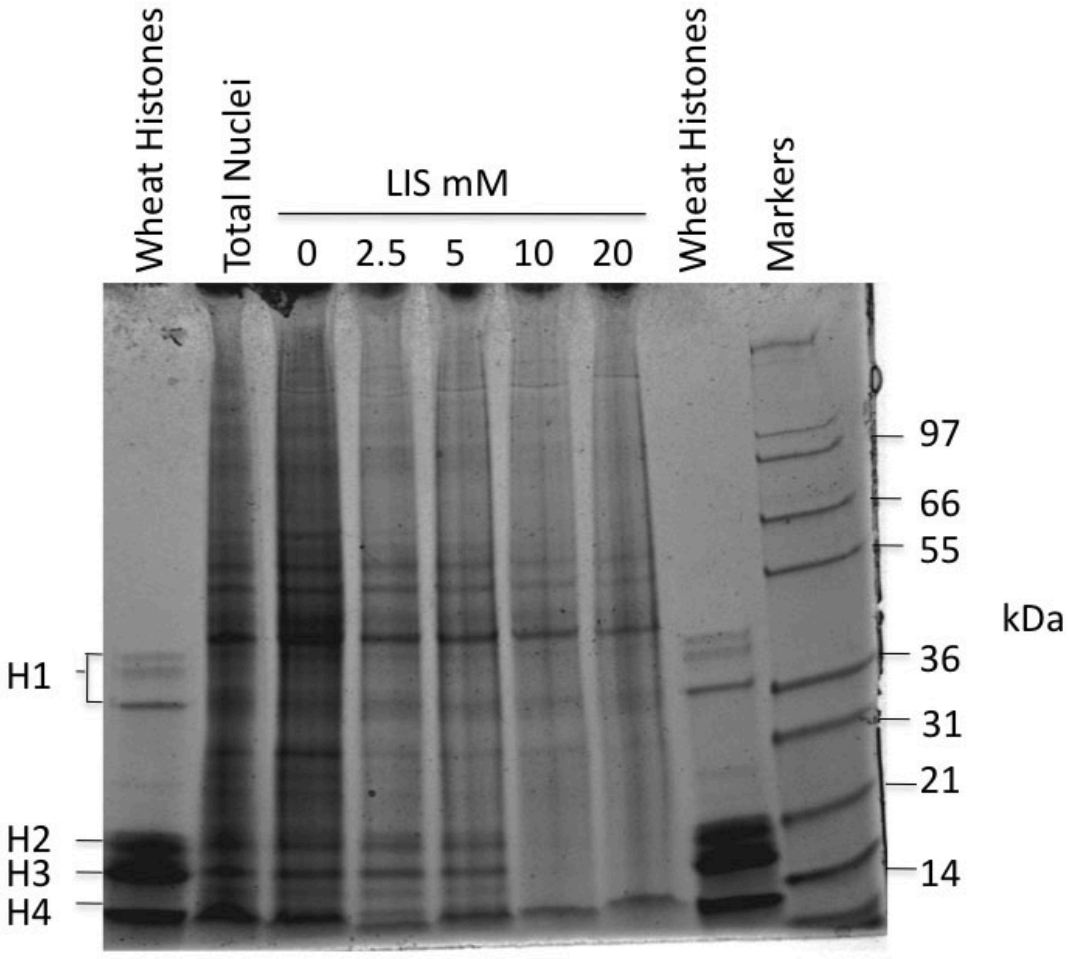
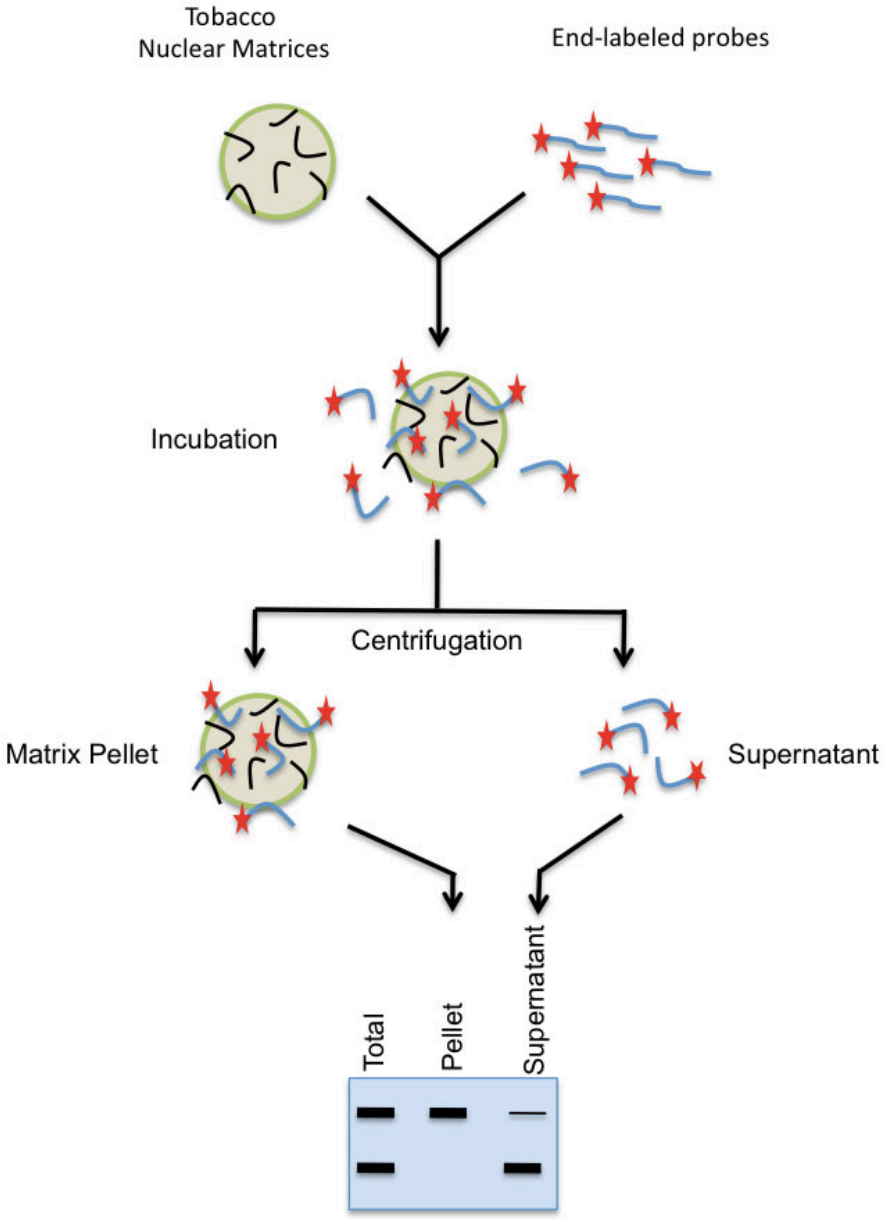
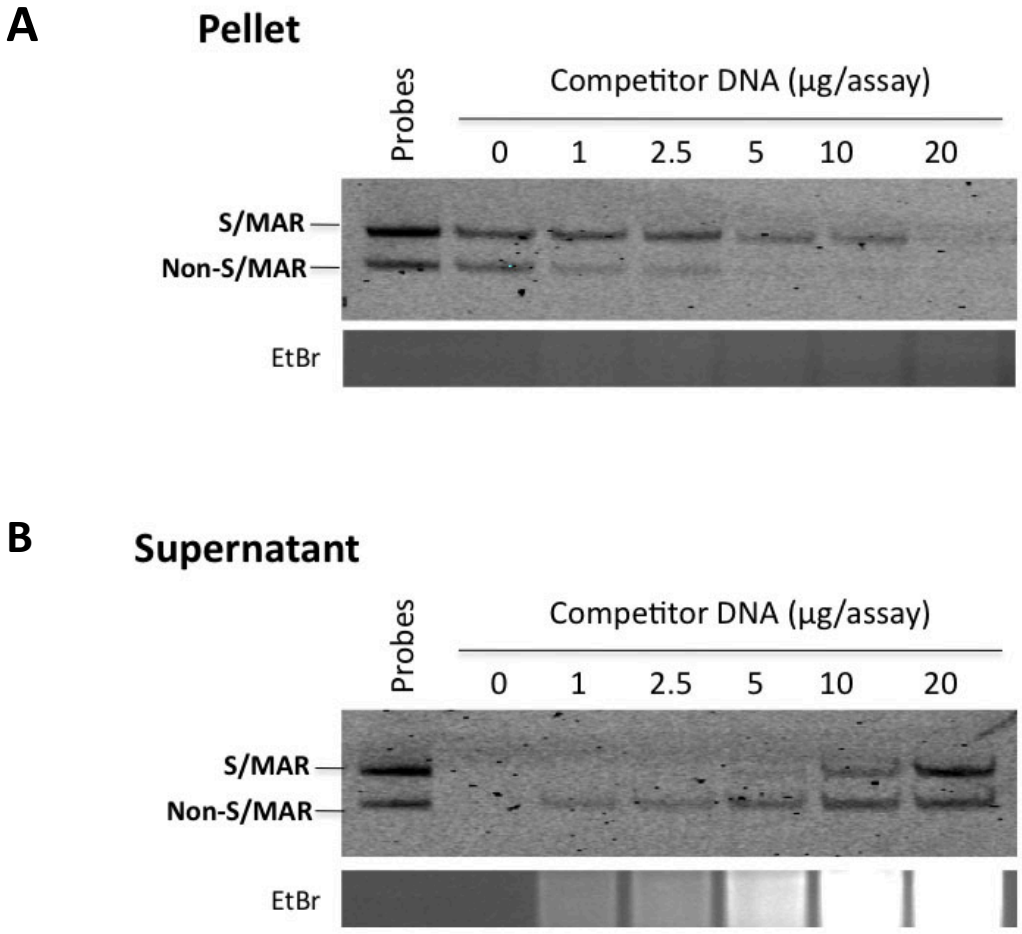


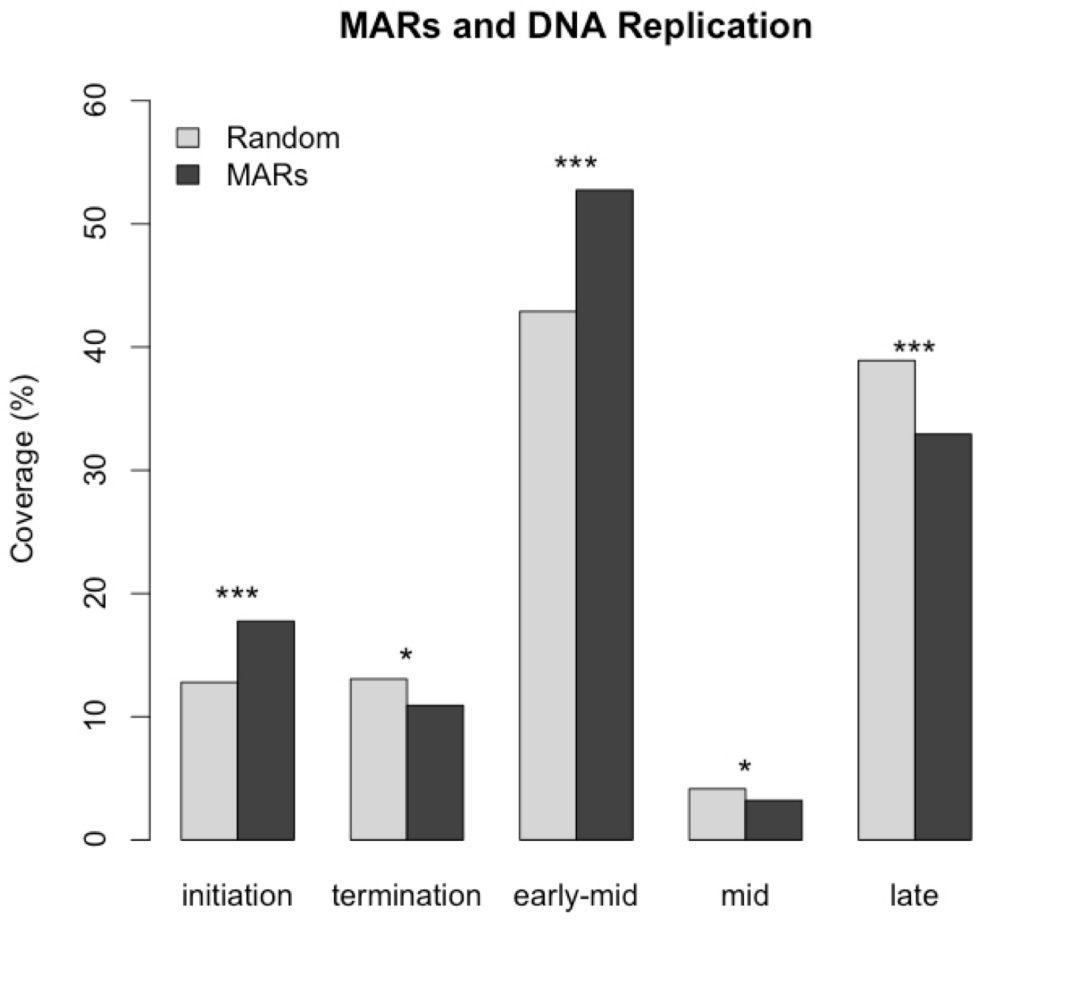
Supplemental Figure 1: Histone removal from Arabidopsis nuclei. A Coomassie-stained SDS/polyacrylamide gel of proteins remaining after extracting nuclei with the LIS- containing buffer HIB2. LIS concentration (mM) is indicated above the each lane. From left to right: Lane 1, purified wheat histones; Lane 2, total nuclear proteins; Lanes 3 to 7, nuclear proteins after treatment with increasing LIS concentrations; Lane 8; purified wheat histones; Lane 9, molecular mass markers. Locations of wheat and Arabidopsis histones are indicated on the left, and masses (kDa) of the markers are shown on the right.



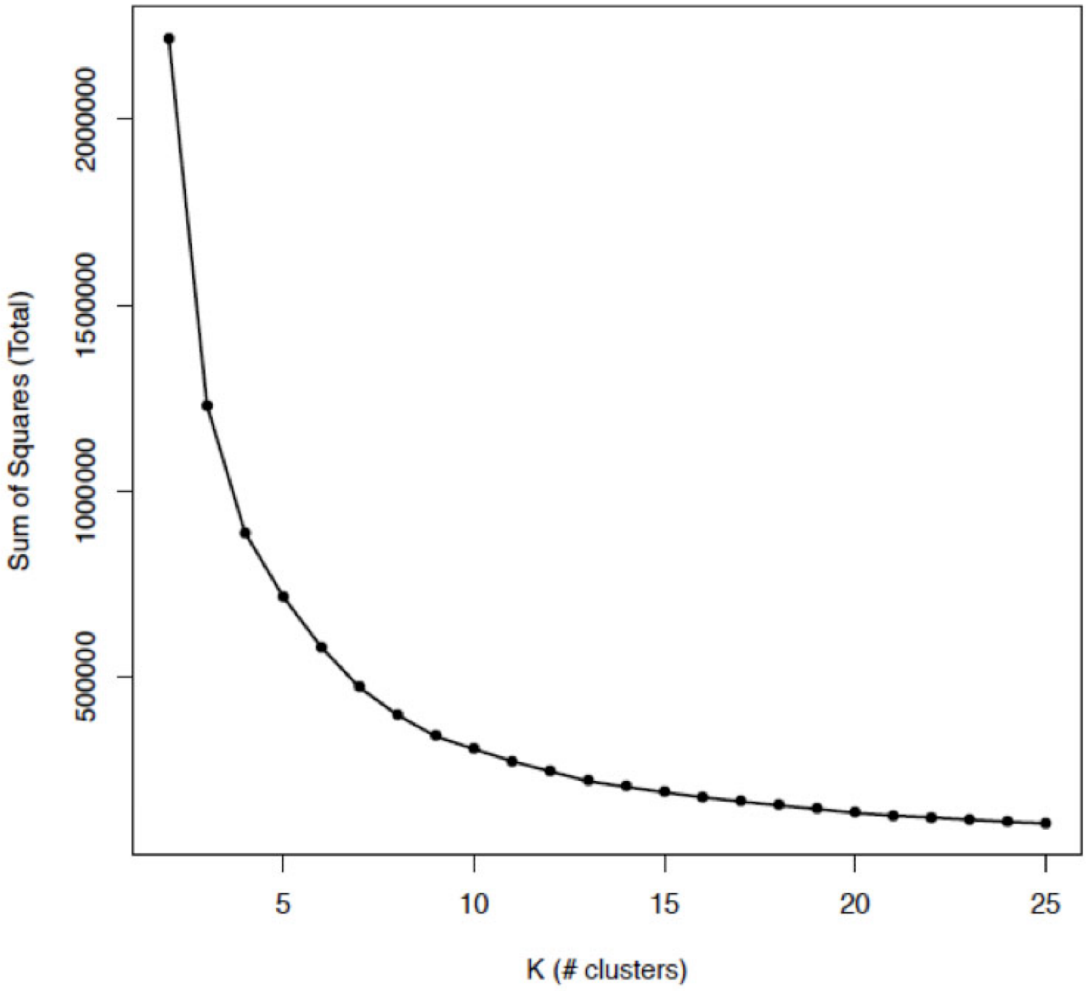
Supplemental Figure 2: In vitro binding assay steps. The binding of a particular DNA fragment was tested by incubation of end-labeled exogenous fragment with nuclear matrices. Details are described within the Methods.



Supplemental Figure 3: Matrix binding competition between S/MARs and *E.coli* DNA. The total assay volume was 50 µL and contained ~6 x 10^4 NT-1 nuclear matrices. The non-S/MAR DNA served as an internal negative control. The total DNA stained with ethidium bromide (EtBr) is shown below each binding assay for both the pellet and the supernatant fractions. (**A)** Pellet lanes show the labeled S/MAR or non-S/MAR DNA bound to the nuclear matrix in the presence of 0, 1, 2.5, 5, 10, or 20 µg of sonicated *E. coli* competitor DNA. Binding of S/MAR and non-S/MAR DNA are similar in the absence of competitor DNA, but only the S/MAR binds to the matrix in the presence of high concentrations of competitor DNA. **(B**) Supernatant lanes show the labeled S/MAR or non-S/MAR DNA that did not bind to the nuclear matrix in the presence of 0, 1, 2.5, 5, 10, or 20 µg of *E. coli* competitor DNA. The S/MAR DNA does not appear in the supernatant until high concentrations of competitor DNA are added.



Supplemental Figure 4: Modest S/MARs enrichment in DNA replication initiation zones and early replicating regions. Determination of the intersection of S/MARs with putative DNA replication initiation and termination zones and segments of DNA replicating at three times in S phase, early-mid, mid only or late only. For comparison, 100,000 random samples of the same size distribution of the S/MARs were generated. S/MARs are modestly enriched in initiation zones and segments of early-mid replication, and depleted in termination zones, and segments of mid-only and late-only replication. P value significance codes *** p <= 0.001, * p <= 0.05 based on permutation tests.

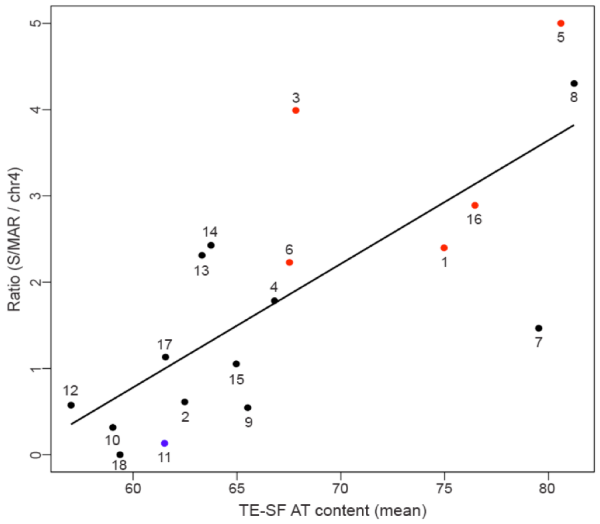


Supplemental Figure 5: Selection of K for K-means clustering of S/MAR data. K-means clustering was used on the gene, exon and TE content to cluster our S/MARs into putative functional groups for further analysis. To determine a reasonable number of clusters, k = 2 to k = 25 was scanned. K = 5 was selected because the reduction in the sum of squares of the residuals began to plateau at this point, and the resulting clusters at k = 5 seemed biologically relevant, i.e. unannotated, TE-associated, flanking genes, exonic, and intronic.

A

ID number	TE superfamily	TE count (chr4)	TE count (S/MARs)	S/MAR count
1	DNA	121	31	29
2	DNA/En-Spm	46	7	7
3	DNA/Harbinger	39	15	15
4	DNA/HAT	131	20	20
5	DNA/Mariner	22	9	9
6	DNA/MuDR	607	201	158
7	DNA/Pogo	58	10	10
8	DNA/Tc1	15	5	5
9	LINE	13	2	2
10	LINE/L1	140	13	12
11	LTR/Copia	185	17	17
12	LTR/Gypsy	130	10	8
13	RathE1 cons	27	6	6
14	RathE2 cons	15	4	4
15	RathE3 cons	17	3	3
16	RC/Helitron	1653	546	357
17	SINE	14	3	3
18	Unassigned	4	0	0

B



Supplemental Figure 6: Relationship between TE-SF AT content and enrichment of TE-SF in S/MARs. (A) The overlap of the various annotated TE-SFs with S/MARs was determined. Shown are the total numbers of each TE-SF for chr4, the number of TE-SF in S/MARs, and the number of S/MARs that overlap with each TE-SF. **(B)** The sequence coverage of each of the 18 annotated TE-SFs for chr4 and for the S/MAR was determined. As a measure of enrichment in S/MARs, the ratio of coverage in S/MARs to coverage in chr4 was calculated. The mean AT content for each TE-SF was also determined. A simple linear model was used to determine the correlation between AT content and S/MAR enrichment for each TE-SF ($R^2 = 0.56$, p value = 0.0003). Statistical significance for any observed enrichment was estimated by permutation, and enriched TE-SFs are indicated in red ($p \leq 0.01$), with the single depleted TE-SF indicated in blue ($p \leq 0.01$).

Supplemental Data. Pascuzzi et al. Plant Cell (2014). 10.1105/tpc.113.121194

Supplemental Table 1. Distribution and density of genes, TEs and S/MARs in chr4 regions.

Region	Low Coord	High Coord	Length	AT content (%)	Gene Count	Gene Density (count/Mb)	TE Count	TE Density (count/Mb)	S/MAR Count	S/MAR Density (count/Mb)	S/MAR distance (kb)
Distal Short	1	1592652	1592652	63.85	452	284	404	254	141	89	11.2
Proximal Short	2310328	2811478	501151	65.55	124	247	260	519	42	84	11.9
Proximal Long	5266604	9200060	3933457	65.35	1020	259	1559	396	325	83	12
Distal Long	9200061	18585056	9384996	63.66	2748	293	1013	108	850	91	11
Combined	NA	NA	15412256	64.17	4344	282	3236	210	1358	88	11.4

Supplemental Table 2. Comparison of mapped and pS/MARs.

		mapped S/MAR						
		0	1					
pS/MAR	0	-	498					
	1	2032	860	29.7				
			63.3					
				S/MAR Cluster				
			Not Identified	**A**	**B**	**C**	**D**	**E**
predicted	**NO**	498	0	114	15	117	181	71
	YES	860	2032	310	261	168	50	71
	%	63.3		73.1	94.6	58.9	21.6	50.0

Supplemental Table 3. Summary of k-means clustering results

Cluster	Gene	Exon	TE	Count	WithinSS	WithinSS/Count	Unannotated	Intron	AT_percent
A	3.24	2.93	7.38	424	105326.46	248.41	89.15	0.31	69.99
B	5.75	4.77	80.54	276	160962.87	583.20	14.87	0.98	72.91
C	47.67	36.83	7.12	285	158016.00	554.44	45.82	10.84	65.88
D	89.94	82.21	4.21	231	129215.36	559.37	9.44	7.73	58.69
E	93.78	28.84	14.21	142	162702.89	1145.79	5.11	64.94	67.12

Supplemental Data. Pascuzzi et al. Plant Cell (2014). 10.1105/tpc.113.121194

Supplemental Table 4. Data for TE superfamily enrichment analysis.

TE_SF	TE_SF_Count	TE_SF_BP	TE_SF_Perc	SMAR_TE_Count	SMAR_Count	SMAR_TE_BP	SMAR_TE_SF_Perc	Ratio	Pvalue	TE_AT_Perc
DNA	121	43528	0.2824	31	29	8047	0.6771	2.398	0.00576	74.97243154
DNA/En-Spm	46	22365	0.1451	7	7	1055	0.08877	0.6118	1	62.48602727
DNA/Harbinger	39	22751	0.1476	15	15	7002	0.5892	3.992	0.00108	67.83438091
DNA/HAT	131	48968	0.3177	20	20	6738	0.5669	1.785	0.48996	66.8109786
DNA/Mariner	22	4964	0.0322	9	9	1914	0.161	5.001	0.00036	80.60032232
DNA/MuDR	607	363159	2.356	201	158	62417	5.252	2.229	0	67.53047563
DNA/Pogo	58	22872	0.1484	10	10	2585	0.2175	1.466	1	79.52955579
DNA/Tc1	15	3379	0.02192	5	5	1121	0.09432	4.303	0.05904	81.23705238
LINE?	13	3167	0.02054	2	2	133	0.01119	0.5447	1	65.51941901
LINE/L1	140	82568	0.5356	13	12	2011	0.1692	0.3159	0.25128	59.02407712
LTR/Copia	185	209256	1.357	17	17	2120	0.1784	0.1314	0	61.51364835
LTR/Gypsy	130	111791	0.7252	10	8	4949	0.4164	0.5742	1	57.01174513
RathE1_cons	27	3541	0.02297	6	6	631	0.05309	2.311	1	63.31544761
RathE2_cons	15	2287	0.01484	4	4	428	0.03601	2.427	1	63.7516397
RathE3_cons	17	2452	0.01591	3	3	199	0.01674	1.053	1	64.96737357
RC/Helitron	1653	873336	5.665	546	357	194615	16.38	2.89	0	76.46035027
SINE	14	4631	0.03004	3	3	404	0.03399	1.132	1	61.56337724
Unassigned	4	6879	0.04462	0	0	0	0	0	1	59.36909435

Supplemental Table 5. Binomial tests of gene activity.								
		TSS-associated Genes						
	Cluster	on	total	pvalue	est	confint_lov	confint_hig	null
	A	112	239	0.01334174	0.4686192	0.4040002	0.5340262	0.5498975
	B	58	145	0.00031245	0.4	0.3196149	0.484558	0.5498975
	C	167	243	1.4119E-05	0.6872428	0.624864	0.7449846	0.5498975
	D	83	137	0.1984159	0.6058394	0.5187991	0.6882005	0.5498975
	E	57	99	0.615439	0.5757576	0.4723362	0.6745238	0.5498975
		TTS-associated genes						
	Cluster	on	total	pvalue	est	confint_lov	confint_hig	null
	A	69	124	0.9281978	0.5564516	0.4645328	0.6456112	0.5498975
	B	32	66	0.3227705	0.4848485	0.3599108	0.6111897	0.5498975
	C	35	72	0.2882447	0.4861111	0.3665003	0.6069	0.5498975
	D	47	74	0.1609528	0.6351351	0.5150697	0.7440246	0.5498975
	E	9	21	0.2805597	0.4285714	0.2181969	0.6597937	0.5498975
		TFs and Non-TFs						
	Cluster	on	total	pvalue	est	confint_lov	confint_hig	null
TFs	A	12	35	0.60593449	0.3428571	0.1913241	0.52211	0.3956044
	B	2	8	0.49158875	0.25	0.031854	0.6508558	0.3956044
	C	12	19	0.05730479	0.6315789	0.3835779	0.8371141	0.3956044
	D	3	11	0.54366332	0.2727273	0.0602177	0.6097426	0.3956044
	E	4	8	0.72037381	0.5	0.1570128	0.8429872	0.3956044
Non-TFs	A	9	13	0.75261848	0.6923077	0.3857383	0.9090796	0.739726
	B	6	9	0.70443005	0.6666667	0.2992951	0.9251454	0.739726
	C	13	18	0.79372956	0.7222222	0.465198	0.9030508	0.739726
	D	8	12	0.52341314	0.6666667	0.3488755	0.9007539	0.739726
	E	5	6	1	0.8333333	0.3587654	0.9957893	0.739726

Supplemental Table 6. NDR content of S/MARs.

	NDR Count								NDR Positive (%)	NDR Coverage (%)
S/MAR Cluster	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**		
A	8	144	186	62	12	9	2	1	98.1	65.1
B	4	81	104	56	20	9	0	2	98.6	67.5
C	11	101	115	40	12	5	0	1	96.1	53.7
D	75	106	36	12	2	0	0	0	67.5	30.2
E	7	56	51	26	2	0	0	0	95.1	53.9

Supplemental Table 7. Comparison of S/MAR-positive and S/MAR-negative NDRs.					
	S/MAR+	**S/MAR-**	**t**	**d.f.**	**p value**
N	2407	12677	-	-	-
Length	278	283	-1.23	3015.2	0.2181
AT content	70.92	68.41	18.31	3225.4	< 2.2E-16
AT content (enrichment)	10.53	6.61	18.31	3225.4	< 2.2E-16
Poly(dA:dT) content	18.0	13.18	27.8	2929.6	< 2.2E-16
Poly(dA:dT) content (enrichment)	85.8	36.4	27.8	2929.6	< 2.2E-16

Welch Two-Sample T-test used for comparisons

Poly(dA:dT) Content determined using tracts of 4 bp or longer and allowing no mismatches.

Enrichment for AT and Poly(dA:dT) content was calculated to make direct comparisons easier.

This enrichment was calculated relative to the means for chr4, 64.1% for AT content and 9.66% for poly(dA:dT) content.