Cell Line	Sub-cellular location	Poly-A-	Poly-A+	Total
GM12878	Cytosolic	18,211,686	---	---
	Nucleolar	---	---	26,792,181
	Nuclear	27,652,635	---	---
H1HESC	Whole Cell	28,801,912	---	---
HEPG2	Cytosolic	19,645,027	---	---
	Nucleolar	---	---	35,803,226
	Nuclear	16,792,966	---	---
HUVEC	Cytosolic	19,837,471	---	---
K562	Cytosolic	20,273,886	18,769,778	---
	Nucleolar	---	---	9,527,032
	Nucleoplasmic	---	---	14,826,128
NHEK	Nuclear	25,989,950	20,648,810	---
	Cytosolic	23,312,041	---	---
	Nuclear	68757727	---	---

Supplementary Table S1: Number of CAGE tag mappings from each cell line, sub-cellular location and poly-adenylation state. CAGE tag mappings from the GM12878, H1HESC, HEPG2, HUVEC and K562 and NHEK cell lines were downloaded from the ENCODE repository on the UCSC genome browser.

Cell Line	Sub-cellular location	Poly-A-	Poly-A+	Total
GM12878	Cytosolic	407,021	---	---
	Nucleolar	---	---	2,458,566
	Nuclear	1,087,671	---	---
H1HESC	Whole Cell	903,838	---	---
HEPG2	Cytosolic	668,040	---	---
	Nucleolar	---	---	2,888,807
	Nuclear	4,188,848	---	---
HUVEC	Cytosolic	857,093	---	---
K562	Cytosolic	4,096,071	525,177	---
	Nucleolar	---	---	3,503,588
	Nucleoplasmic	---	---	4,617,119
NHEK	Nuclear	6,829,025	2,244,742	---
	Cytosolic	1,730,893	---	---
	Nuclear	3,082,557	---	---

Supplementary Table S2: CAGE clusters identified in each cell line, sub-cellular location and poly-adenylation state. Overlapping CAGE tag mappings (Table S1) from the ENCODE cell lines and in the same sub-cellular locations and poly-adenylation states were grouped together and desiganted as clusters.

Modification	Tags Mapped					
	GM12878	H1HESC	HEPG2	HUVEC	K562	NHEK
Control	7,436,431	11,908,617	11,039,784	16,836,245	13,240,739	10,666,985
H3K4Me1	14,069,086	9,713,507	---	14,524,897	---	11,260,426
H3K4Me2	9,163,434	14,479,372	17,293,347	12,005,596	12,454,360	11,031,009
H3K4me3	10,218,953	7,072,374	10,289,145	12,497,262	15,989,323	10,296,574
H3K9Ac	12,022,891	16,477,468	7,351,567	8,670,429	17,281,199	12,454,536
H3K9Me1	---	---	---	10,658,052	15,905,405	10,731,385
H3K27Ac	10,770,731	---	8,856,877	16,833,005	15,871,535	12,788,055
H3K27Me3	14,430,662	7,160,479	---	11,652,289	12,412,831	9,141,036
H3K36Me3	15,195,406	14,680,520	13,579,529	9,818,236	14,950,529	9,182,104
H4K20Me1	12,224,195	16,605,685	10,356,633	16,664,745	13,685,630	12,380,840
Pol2b	---	---	---	9,860,160	10,822,295	10,175,792

Supplementary Table S3: ChIP-seq reads mapped for each histone modification and cell line. ChIP-seq data from the GM12878 and K562 cell lines were downloaded from the ENCODE repository on the UCSC genome browser. Reads were mapped using bowtie, keeping the best hits with ties broken by quality. Ambiguously mapped reads were resolved using GibbsAM

a

Counts		cis-NAT		
		High	Not	
Promoter	High	5663	20069	25732
	Not	12872	125847	138719
		18535	145916	164451

Fractions		cis-NAT		
		High	Not	
Promoter	High	0.034	0.122	0.156
	Not	0.078	0.765	0.844
		0.113	0.887	1

Measures		
	Confidence	Interest
C(N->P)	0.306	1.953
C(P->N)	0.220	

b

		cis-NAT		
		High	Not	
Promoter	Low	4427	83723	88150
	Not	14108	62193	76301
		18535	145916	164451

		cis-NAT		
		High	Not	
Promoter	Low	0.027	0.509	0.536
	Not	0.086	0.378	0.464
		0.113	0.887	1

	Confidence	Interest
C(N->P)	0.239	0.446
C(P->N)	0.050	

c

		cis-NAT		
		Low	Not	
Promoter	High	10110	15622	25732
	Not	99123	39596	138719
		109233	55218	164451

		cis-NAT		
		Low	Not	
Promoter	High	0.061	0.095	0.156
	Not	0.603	0.241	0.844
		0.664	0.336	1

	Confidence	Interest
C(N->P)	0.093	0.592
C(P->N)	0.393	

d

		cis-NAT		
		Low	Not	
Promoter	Low	74788	13362	88150
	Not	34445	41856	76301
		109233	55218	164451

		cis-NAT		
		Low	Not	
Promoter	Low	0.455	0.081	0.536
	Not	0.209	0.255	0.464
		0.664	0.336	1

	Confidence	Interest
C(N->P)	0.685	1.277
C(P->N)	0.848	

Supplementary Table S4. Association of cis-NAT and promoter activity. Cis-NAT promoters were identified in GM12878, H1HESC, HEPG2, HUVEC, K562 and NHEK cells using CAGE data. Genic promoters were characterized as 'high' in a cell type if they were in the top 10% most active of detected genic promoters in that cell type. Genic promoters were characterized as 'low' in a cell type if they were in the bottom 10% least active of detected promoters in that cell type or there were not detected in that cell type. The same was done for the sum of cis-NAT promoter activity corresponding to each genic promoter. The co-occurrence of (a) high promoter and high cis-NAT activity (b) low promoter and high cis-NAT activity (c) high promoter and low cis-NAT activity and (d) low promoter and low cis-NAT activity was found across the six cell types.

a

Counts		cis-NAT		
		High	Not	
Promoter	High	5120	20507	25627
	Not	11851	125314	137165
		16971	145821	162792

Fractions		cis-NAT		
		High	Not	
Promoter	High	0.031	0.126	0.157
	Not	0.073	0.770	0.843
		0.104	0.896	1

Measures		
	Confidence	Interest
C(N->P)	0.302	1.916
C(P->N)	0.200	

b

		cis-NAT		
		High	Not	
Promoter	Low	4138	82526	86664
	Not	12833	63295	76128
		16971	145821	162792

		cis-NAT		
		High	Not	
Promoter	Low	0.025	0.507	0.532
	Not	0.079	0.389	0.468
		0.104	0.896	1

	Confidence	Interest
C(N->P)	0.244	0.458
C(P->N)	0.048	

c

		cis-NAT		
		Low	Not	
Promoter	High	11428	14199	25627
	Not	101017	36148	137165
		112445	50347	162792

		cis-NAT		
		Low	Not	
Promoter	High	0.070	0.087	0.157
	Not	0.621	0.222	0.843
		0.691	0.309	1

	Confidence	Interest
C(N->P)	0.102	0.646
C(P->N)	0.446	

d

		cis-NAT		
		Low	Not	
Promoter	Low	74341	12323	86664
	Not	38104	38024	76128
		112445	50347	162792

		cis-NAT		
		Low	Not	
Promoter	Low	0.457	0.076	0.532
	Not	0.234	0.234	0.468
		0.691	0.309	1

	Confidence	Interest
C(N->P)	0.661	1.242
C(P->N)	0.858	

Supplementary Table S5. Association of cis-NAT and promoter activity for cis-NAT promoters distal to the genic promoter. Cis-NAT promoters were identified in GM12878, H1HESC, HEPG2, HUVEC, K562 and NHEK cells using CAGE data. Genic promoters were characterized as 'high' in a cell type if they were in the top 10% most active of detected genic promoters in that cell type. Genic promoters were characterized as 'low' in a cell type if they were in the bottom 10% least active of detected promoters in that cell type or there were not detected in that cell type. The same was done for the sum of distal (> 2500bp downstream from the genic promoter) cis-NAT promoter activity corresponding to each genic promoter. The co-occurrence of (a) high promoter and high cis-NAT activity (b) low promoter and high cis-NAT activity (c) high promoter and low cis-NAT activity and (d) low promoter and low cis-NAT activity was found across the six cell types.

a

Counts		cis-NAT		
		High	Not	
Promoter	High	2233	22534	24767
	Not	4409	119273	123682
		6642	141807	148449

Fractions		cis-NAT		
		High	Not	
Promoter	High	0.015	0.152	0.167
	Not	0.030	0.803	0.833
		0.045	0.955	1

Measures		
	Confidence	Interest
C(N->P)	0.336	2.015
C(P->N)	0.090	

b

		cis-NAT		
		High	Not	
Promoter	Low	1406	73696	75102
	Not	5236	68111	73347
		6642	141807	148449

		cis-NAT		
		High	Not	
Promoter	Low	0.009	0.496	0.506
	Not	0.035	0.459	0.494
		0.045	0.955	1

	Confidence	Interest
C(N->P)	0.212	0.418
C(P->N)	0.019	

c

		cis-NAT		
		Low	Not	
Promoter	High	18768	5999	24767
	Not	109932	13750	123682
		128700	19749	148449

		cis-NAT		
		Low	Not	
Promoter	High	0.126	0.040	0.167
	Not	0.741	0.093	0.833
		0.867	0.133	1

	Confidence	Interest
C(N->P)	0.146	0.874
C(P->N)	0.758	

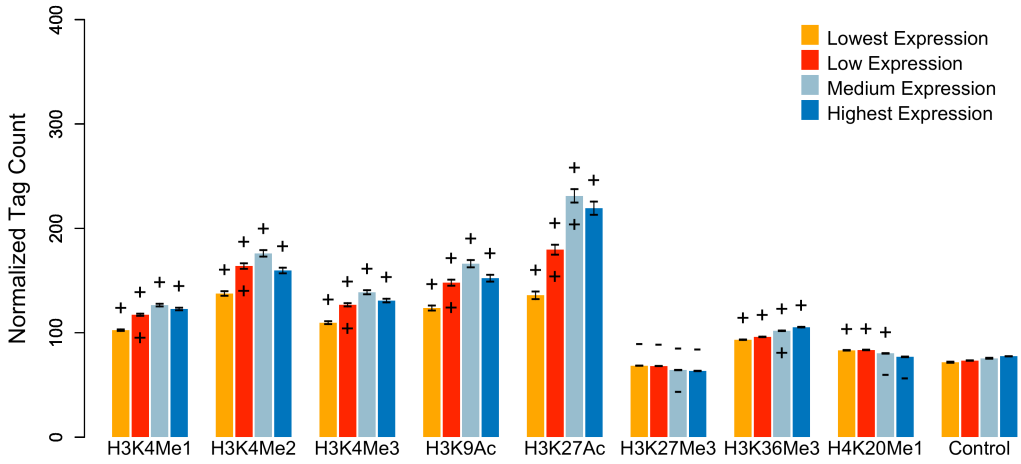
d

		cis-NAT		
		Low	Not	
Promoter	Low	70760	4342	75102
	Not	57940	15407	73347
		128700	19749	148449

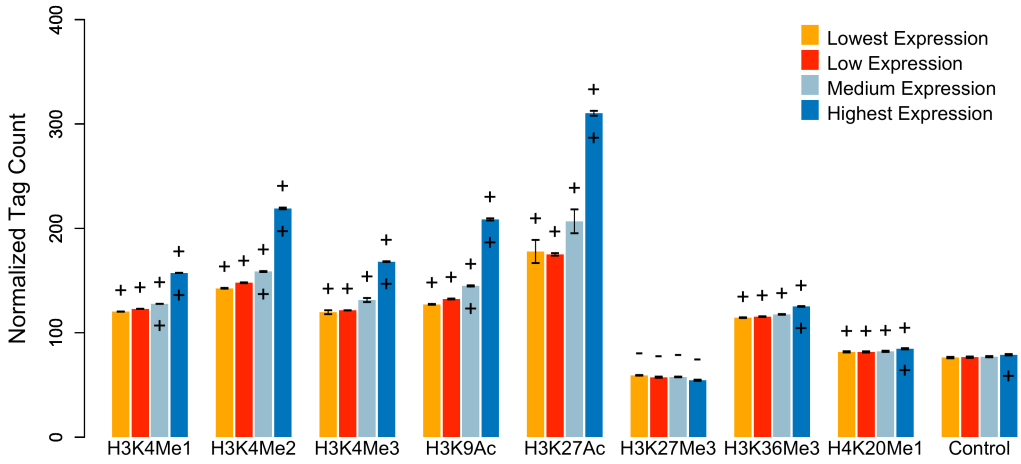
		cis-NAT		
		Low	Not	
Promoter	Low	0.477	0.029	0.506
	Not	0.390	0.104	0.494
		0.867	0.133	1

	Confidence	Interest
C(N->P)	0.550	1.087
C(P->N)	0.942	

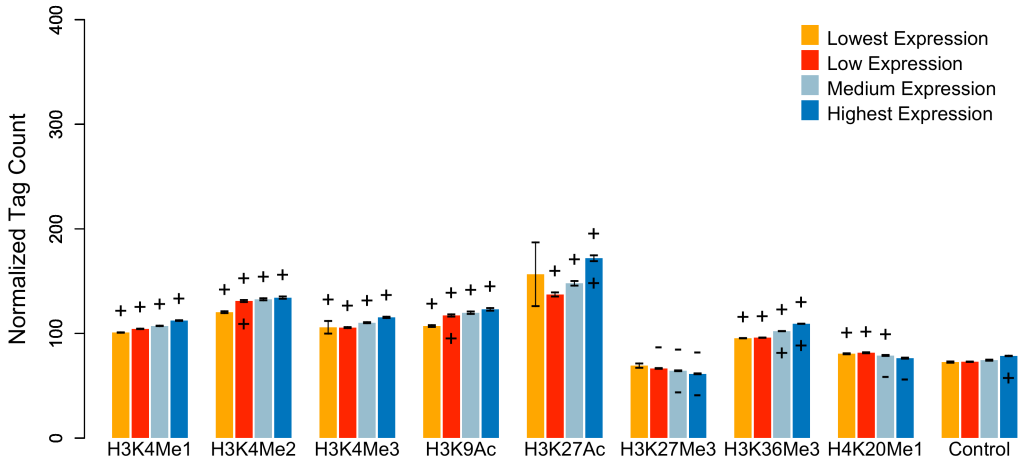
Supplementary Table S6. Association of cis-NAT and promoter activity for cis-NAT promoters proximal to the genic promoter. Cis-NAT promoters were identified in GM12878, H1HESC, HEPG2, HUVEC, K562 and NHEK cells using CAGE data. Genic promoters were characterized as 'high' in a cell type if they were in the top 10% most active of detected genic promoters in that cell type. Genic promoters were characterized as 'low' in a cell type if they were in the bottom 10% least active of detected promoters in that cell type or there were not detected in that cell type. The same was done for the sum of proximal (< 2500bp downstream from the genic promoter) cis-NAT promoter activity corresponding to each genic promoter. The co-occurrence of (a) high promoter and high cis-NAT activity (b) low promoter and high cis-NAT activity (c) high promoter and low cis-NAT activity and (d) low promoter and low cis-NAT activity was found across the six cell types.



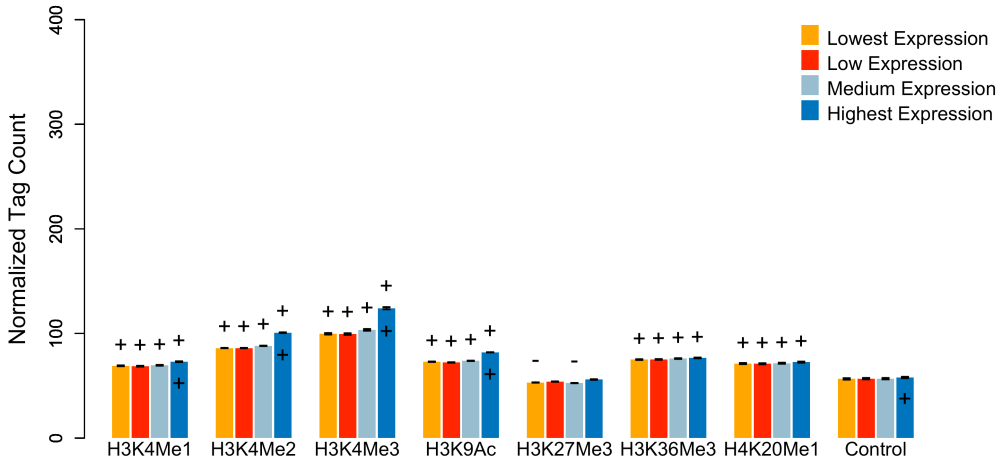
Supplementary Figure S1. Enrichment of chromatin modifications at cis-NAT promoters in GM12878. Cis-NAT promoters in the GM12878 cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis-NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



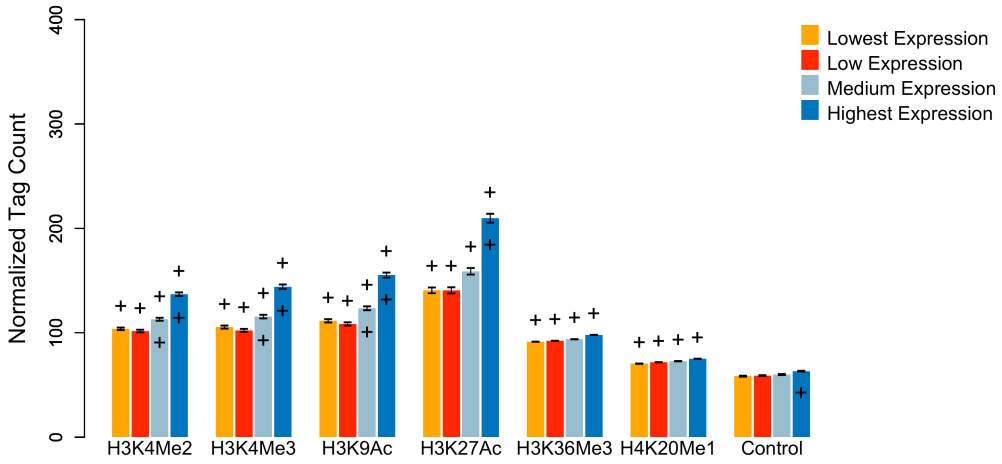
Supplementary Figure S2. Enrichment of chromatin modifications at cis-NAT promoters in GM12878. Cis-NAT promoters in the GM12878 cell type were identified using CAGE data from total RNA from nucleolus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis-NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



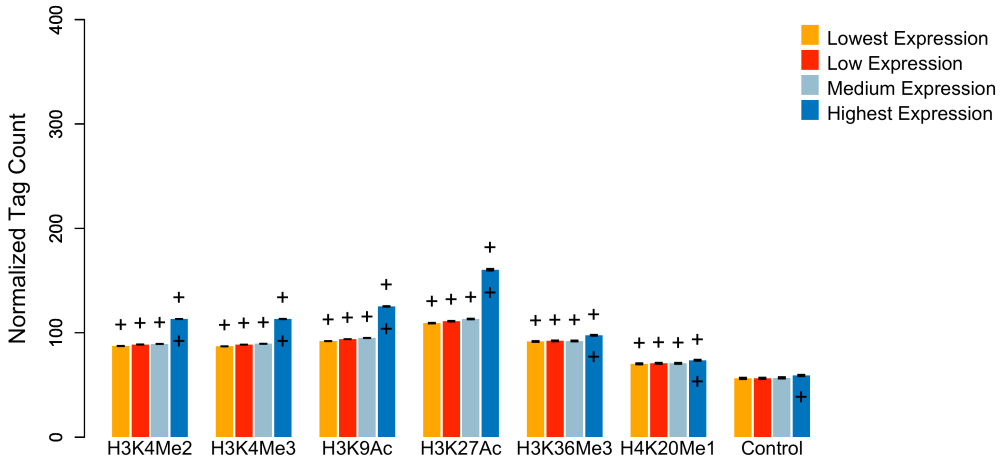
Supplementary Figure S3. Enrichment of chromatin modifications at cis-NAT promoters in GM12878. Cis-NAT promoters in the GM12878 cell type were identified using CAGE data from non-polyadenylated RNA from nucleus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis-NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



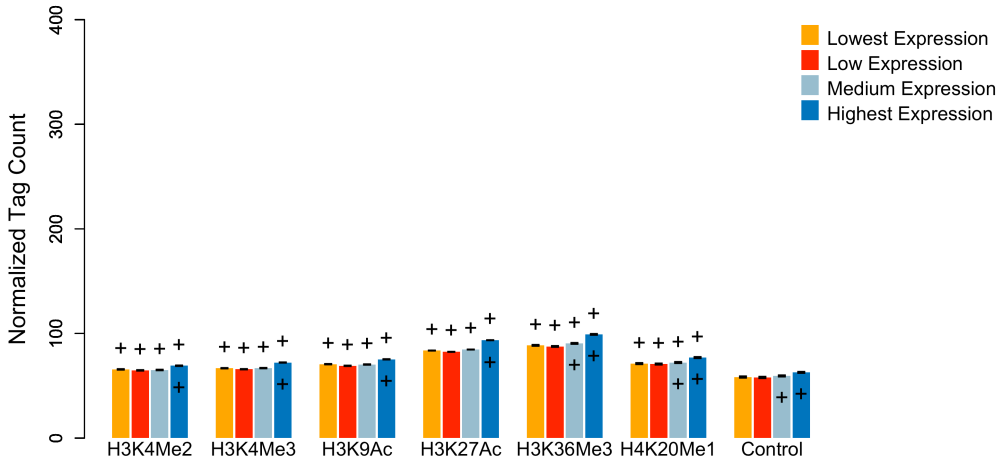
Supplementary Figure S4. Enrichment of chromatin modifications at cis-NAT promoters in H1HESC. Cis-NAT promoters in the H1HESC cell type were identified using CAGE data from total RNA from whole-cell isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis-NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



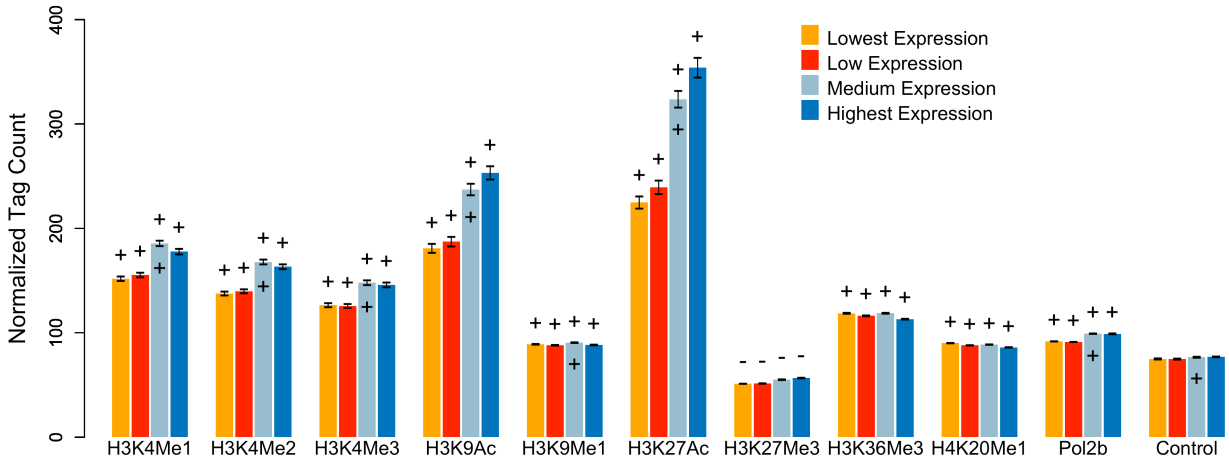
Supplementary Figure S5. Enrichment of chromatin modifications at cis-NAT promoters in HepG2. Cis-NAT promoters in the HepG2 cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis- NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



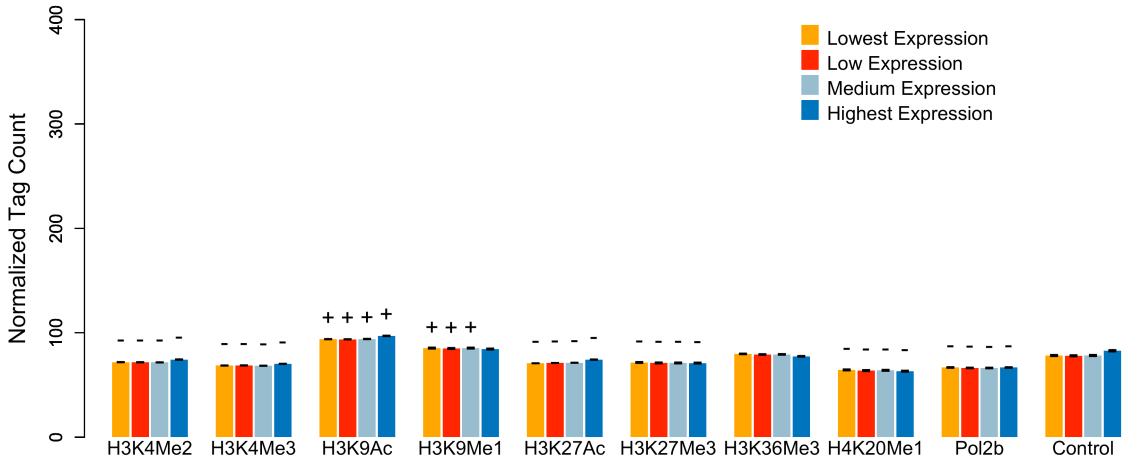
Supplementary Figure S6. Enrichment of chromatin modifications at cis-NAT promoters in HepG2. Cis-NAT promoters in the HepG2 cell type were identified using CAGE data from total RNA from nucleolus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis-NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



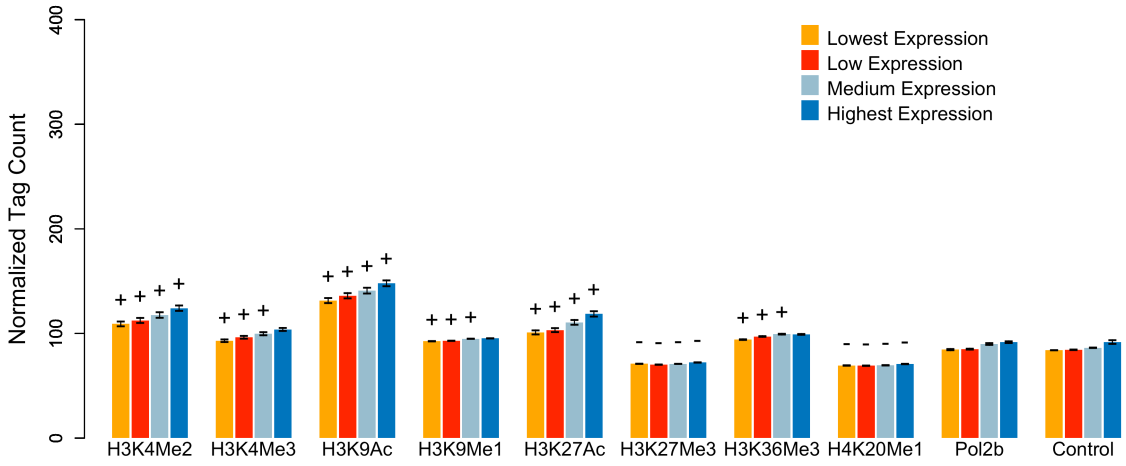
Supplementary Figure S7. Enrichment of chromatin modifications at cis-NAT promoters in HepG2. Cis-NAT promoters in the HepG2 cell type were identified using CAGE data from non-polyadenylated RNA from nucleus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis- NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



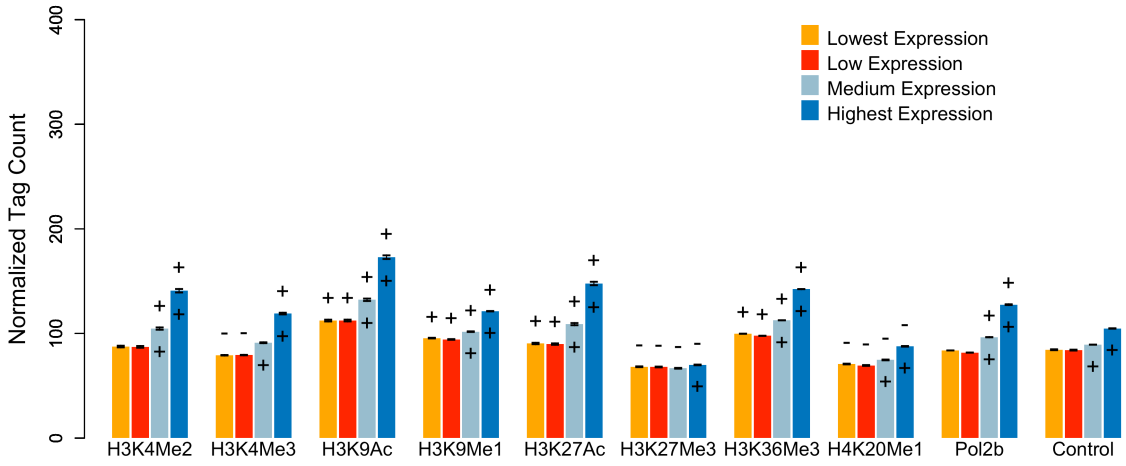
Supplementary Figure S8. Enrichment of chromatin modifications at cis-NAT promoters in HUVEC. Cis-NAT promoters in the HUVEC cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis- NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



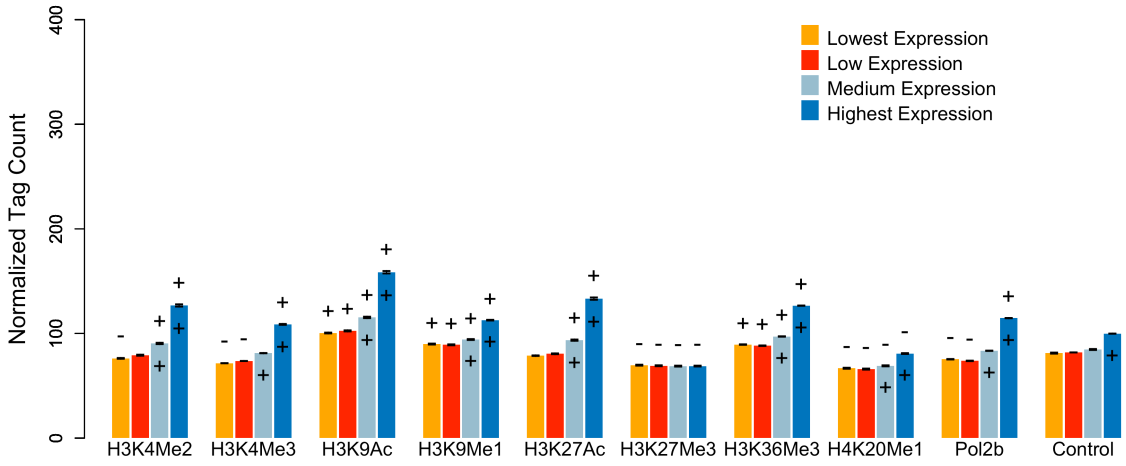
Supplementary Figure S9. Enrichment of chromatin modifications at cis-NAT promoters in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis- NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



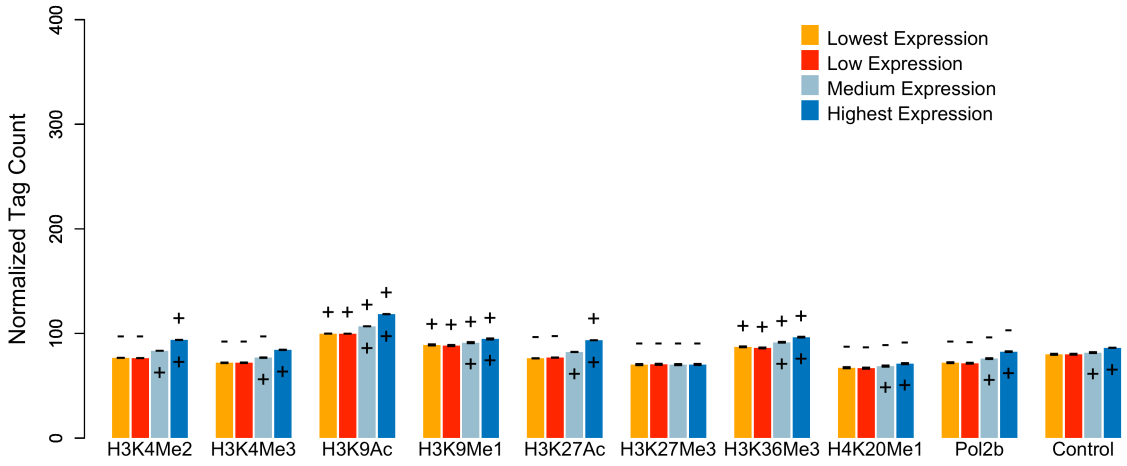
Supplementary Figure S10. Enrichment of chromatin modifications at cis-NAT promoters in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from polyadenylated RNA from cytosol isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis-NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



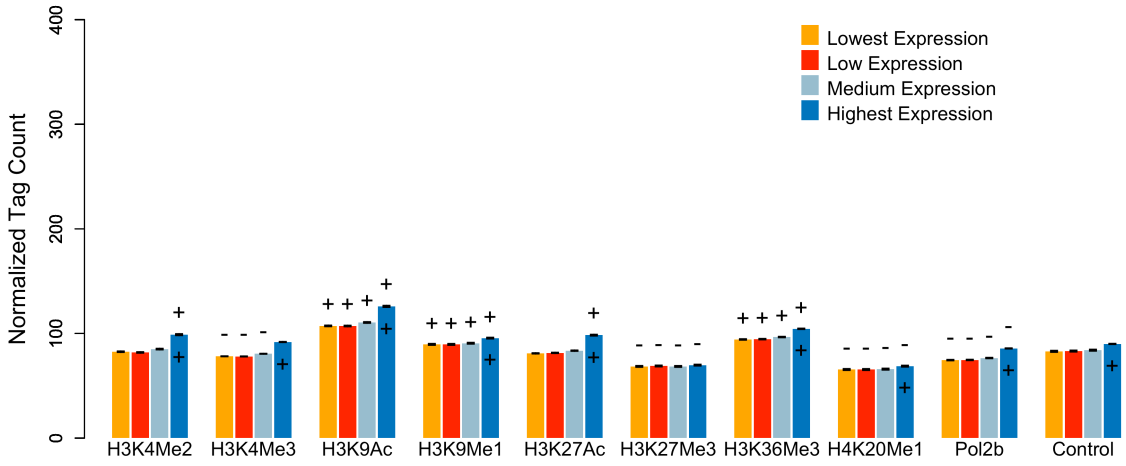
Supplementary Figure S11. Enrichment of chromatin modifications at cis-NAT promoters in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from total RNA from nucleolus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis-NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



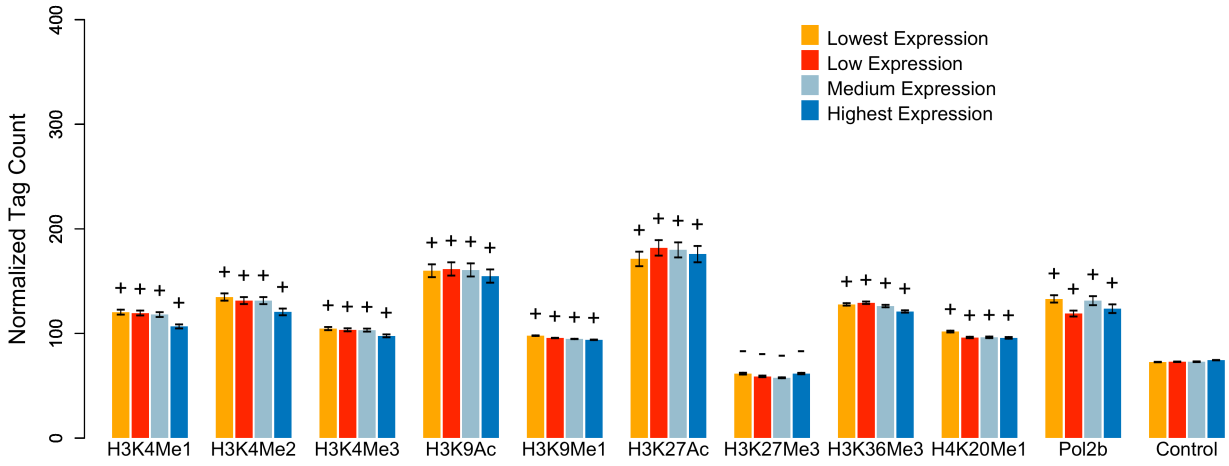
Supplementary Figure S12. Enrichment of chromatin modifications at cis-NAT promoters in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from total RNA from nucleoplasm isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis-NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



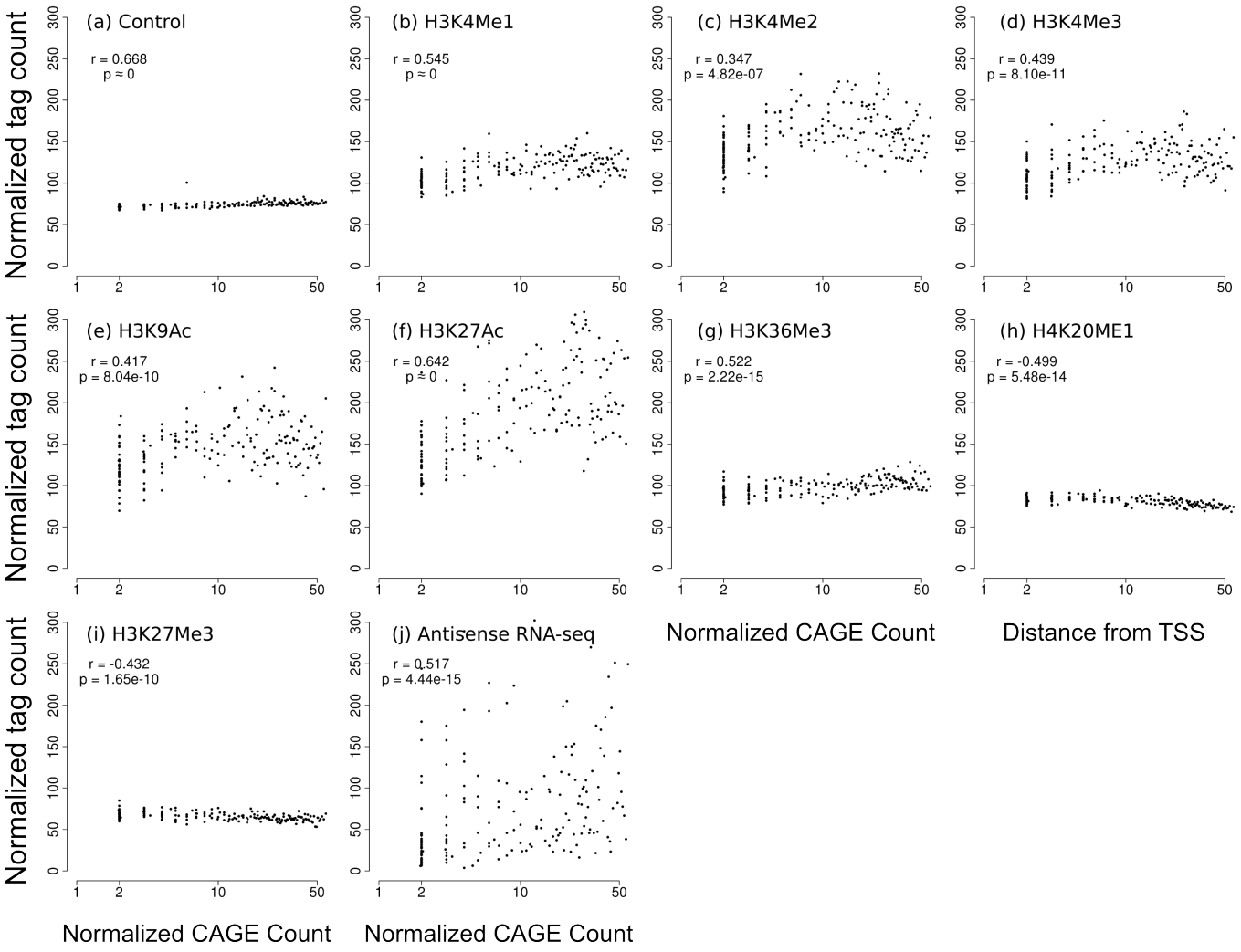
Supplementary Figure S13. Enrichment of chromatin modifications at cis-NAT promoters in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from non-polyadenylated RNA from nucleus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis- NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



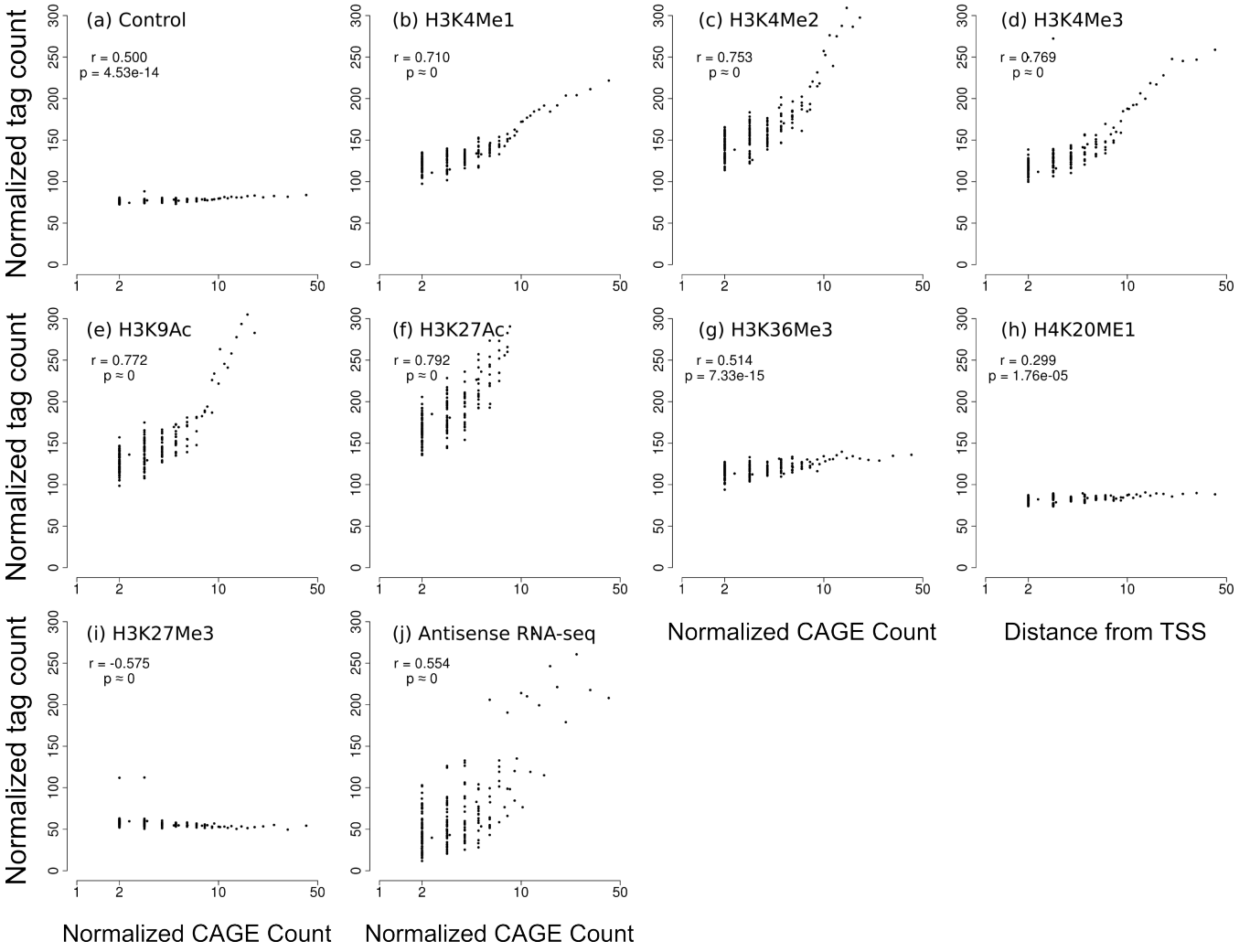
Supplementary Figure S14. Enrichment of chromatin modifications at cis-NAT promoters in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from polyadenylated RNA from nucleus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis- NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



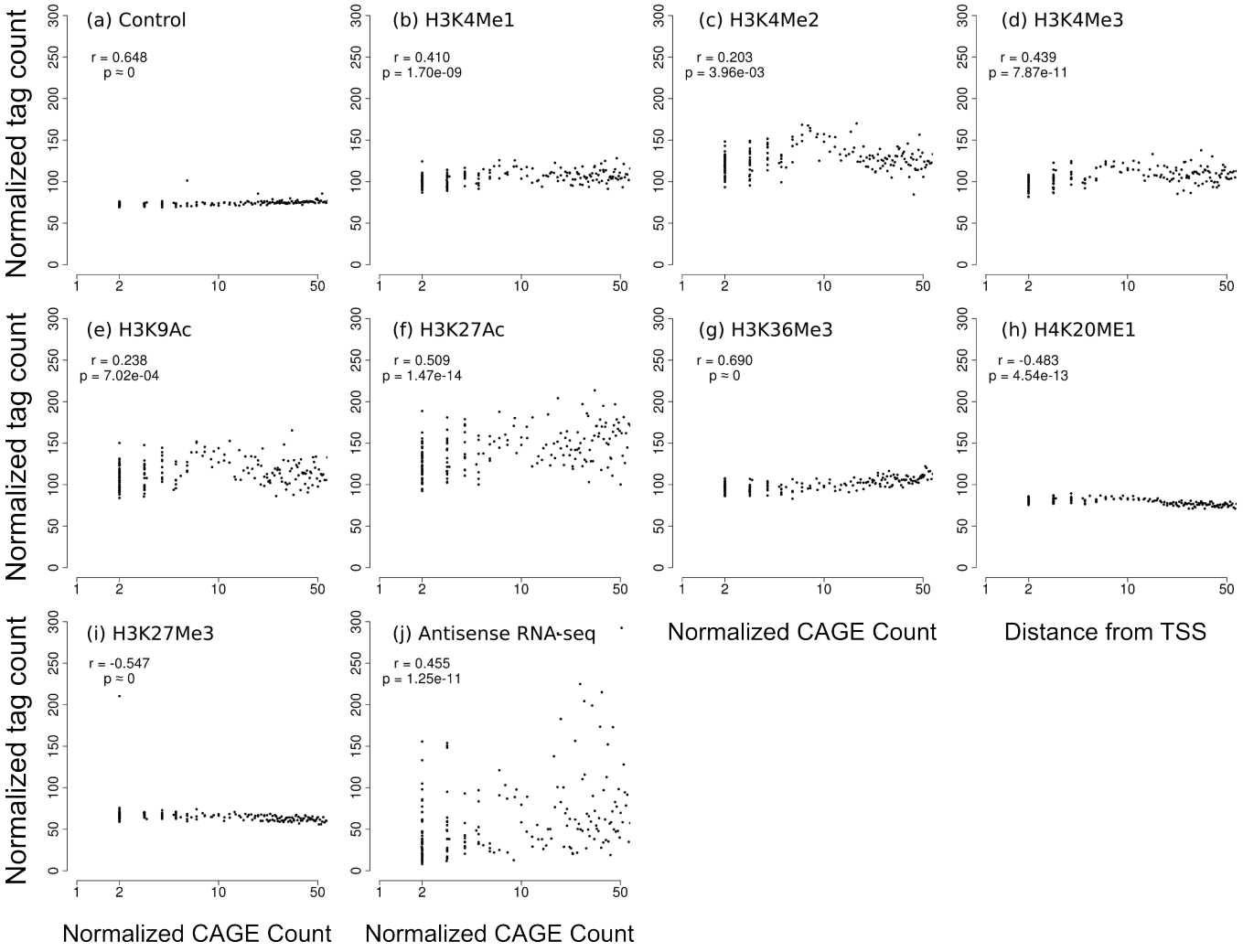
Supplementary Figure S15. Enrichment of chromatin modifications at cis-NAT promoters in NHEK. Cis-NAT promoters in the NHEK cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads from each histone modification +/- 5 kb of the cis- NAT TSS were calculated for each bin. A '+' or '-' above a bar indicates that the number of ChIP-seq reads for that bin and modification is significantly higher or lower, respectively, than the control for that bin ($P < 0.001$). A '+' or '-' within a bar indicates that a bin is significantly enriched or depleted, respectively, for the histone modification compared to the next lowest activity bin ($P < 0.001$). Error bars shown are the standard error of the mean.



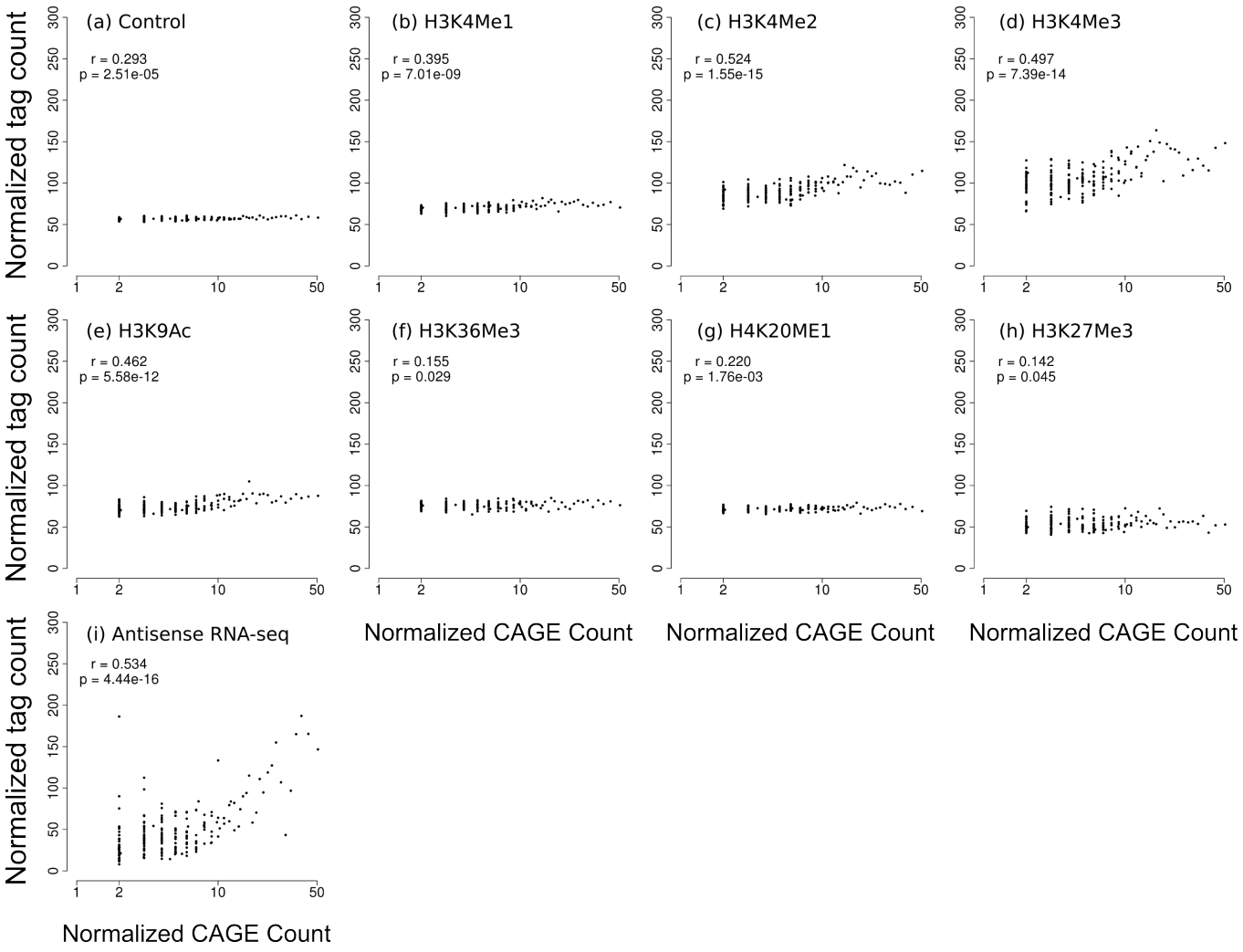
Supplementary Figure S16. Correlation between cis-NAT promoter activity and local chromatin environment in GM12878. Cis-NAT promoters in the GM12878 cell type were identified using CAGE data from non-polyadenylated cytosol isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis-NAT promoter activity and local histone modifications.



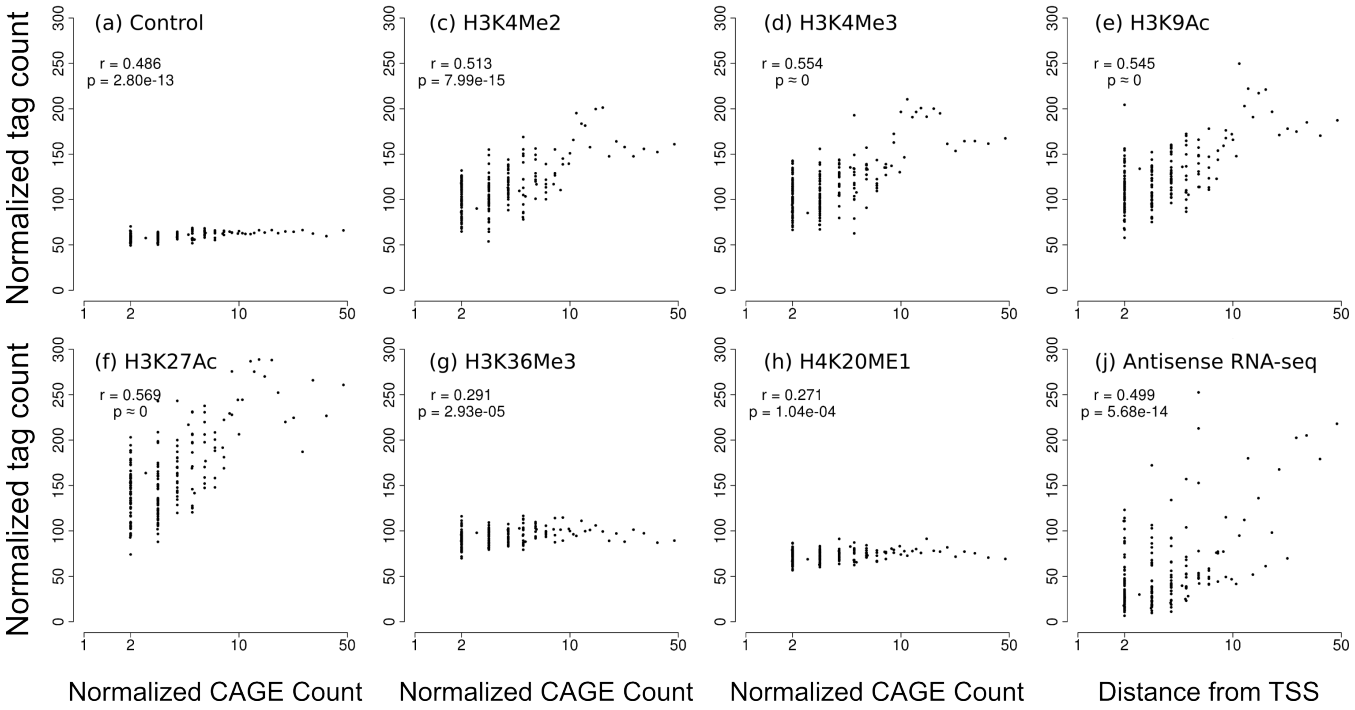
Supplementary Figure S17. Correlation between cis-NAT promoter activity and local chromatin environment in GM12878. Cis-NAT promoters in the GM12878 cell type were identified using CAGE data from total nucleolus isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis- NAT promoter activity and local histone modifications.



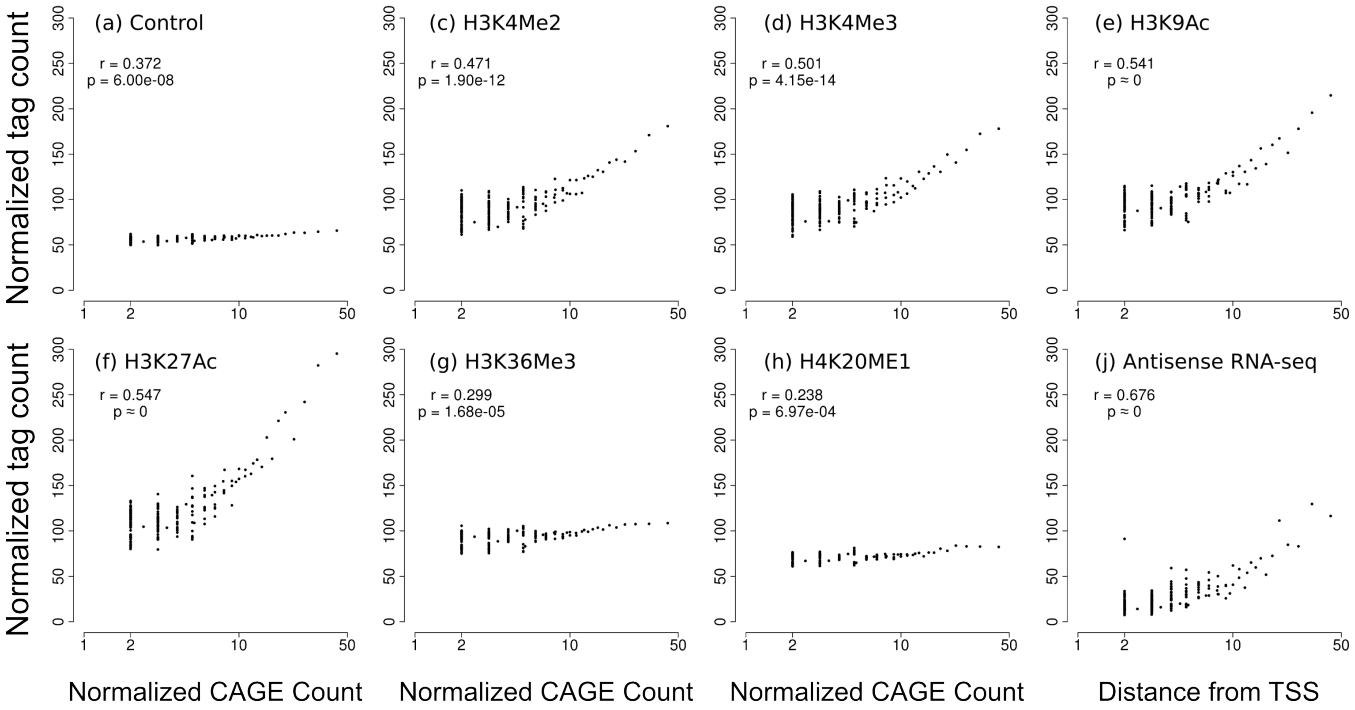
Supplementary Figure S18. Correlation between cis-NAT promoter activity and local chromatin environment in GM12878. Cis-NAT promoters in the GM12878 cell type were identified using CAGE data from non-polyadenylated nucleus isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis-NAT promoter activity and local histone modifications.



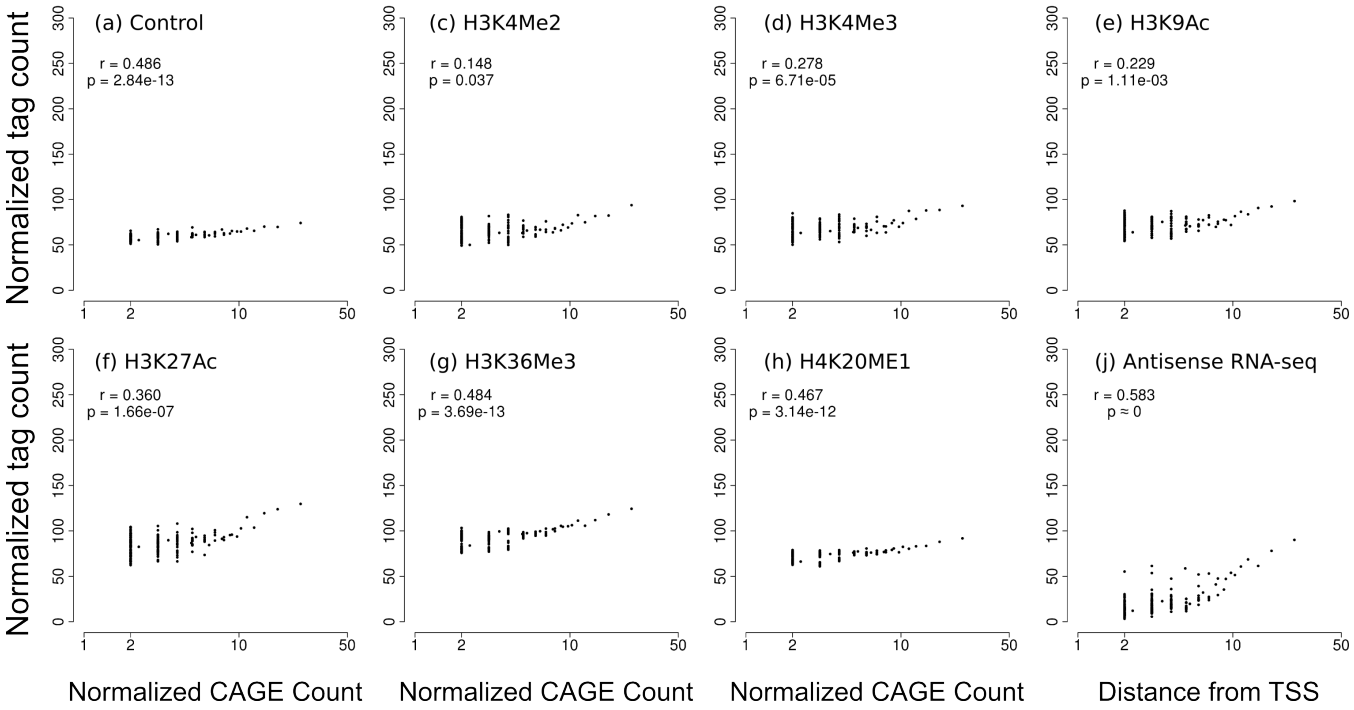
Supplementary Figure S19. Correlation between cis-NAT promoter activity and local chromatin environment in H1HESC. Cis-NAT promoters in the H1HESC cell type were identified using CAGE data from non-polyadenylated whole-cell isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis-NAT promoter activity and local histone modifications**.**



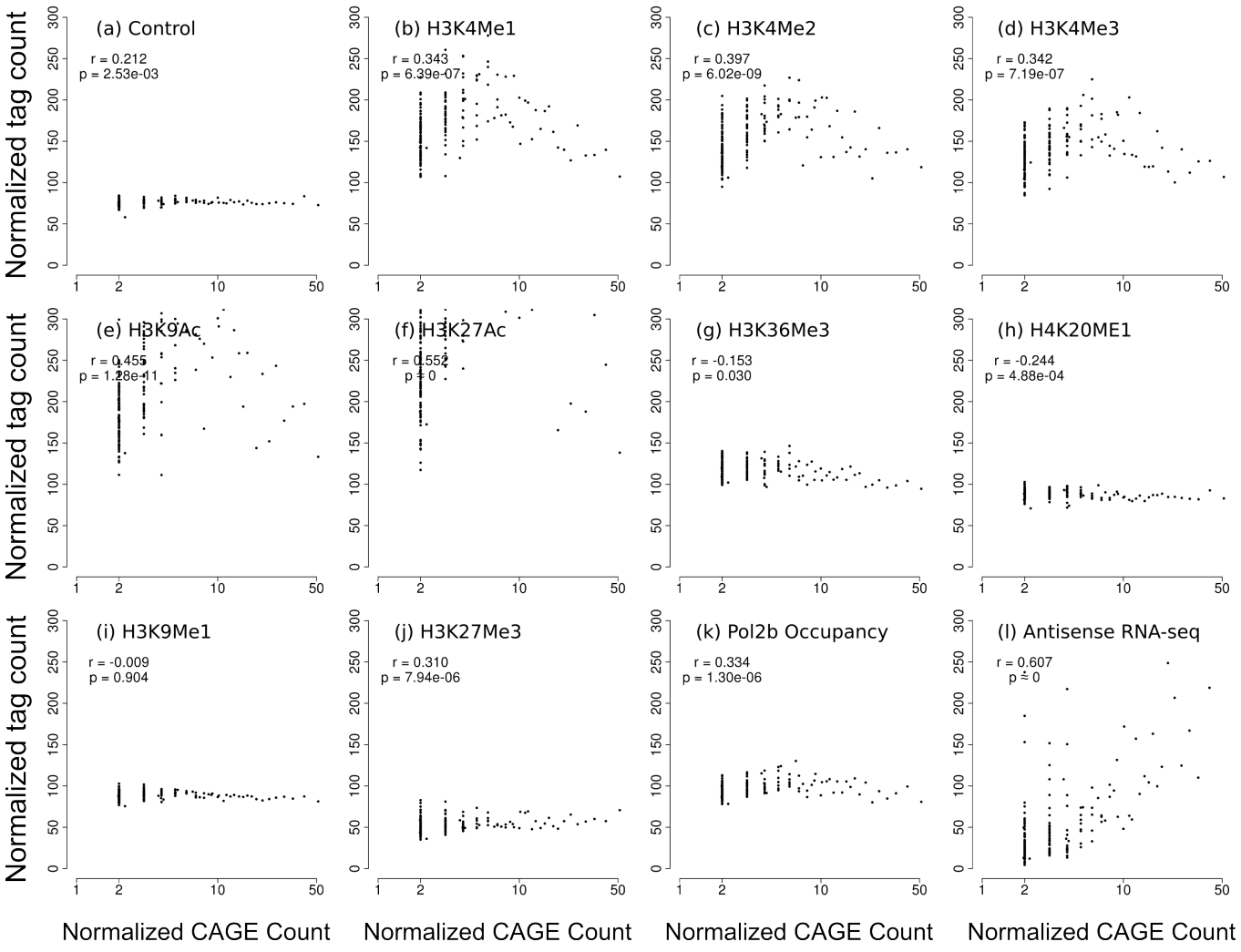
Supplementary Figure S20. Correlation between cis-NAT promoter activity and local chromatin environment in HepG2. Cis-NAT promoters in the HepG2 cell type were identified using CAGE data from non-polyadenylated cytosol isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis- NAT promoter activity and local histone modifications.



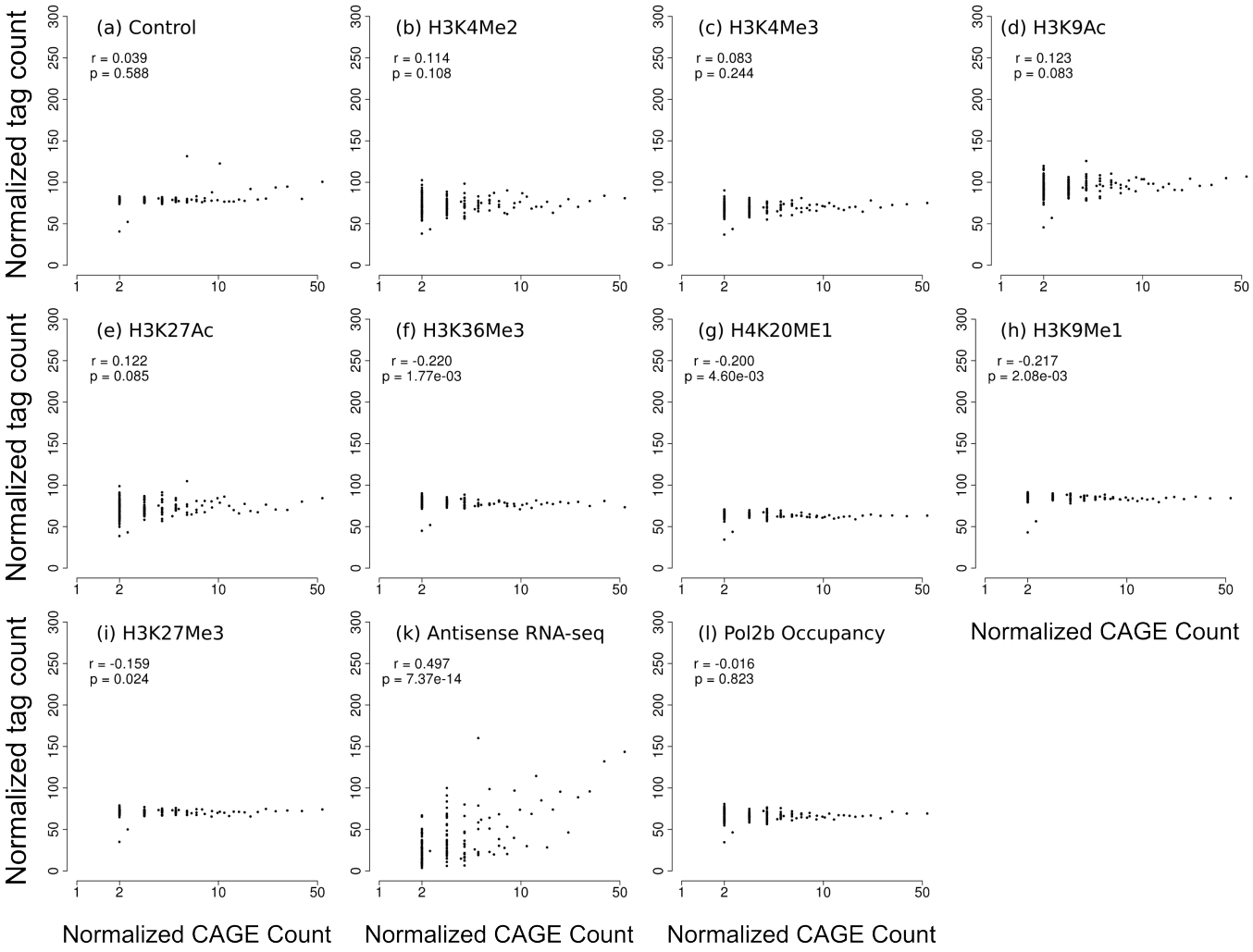
Supplementary Figure S21. Correlation between cis-NAT promoter activity and local chromatin environment in HepG2. Cis-NAT promoters in the HepG2 cell type were identified using CAGE data from total nucleolus isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis- NAT promoter activity and local histone modifications.



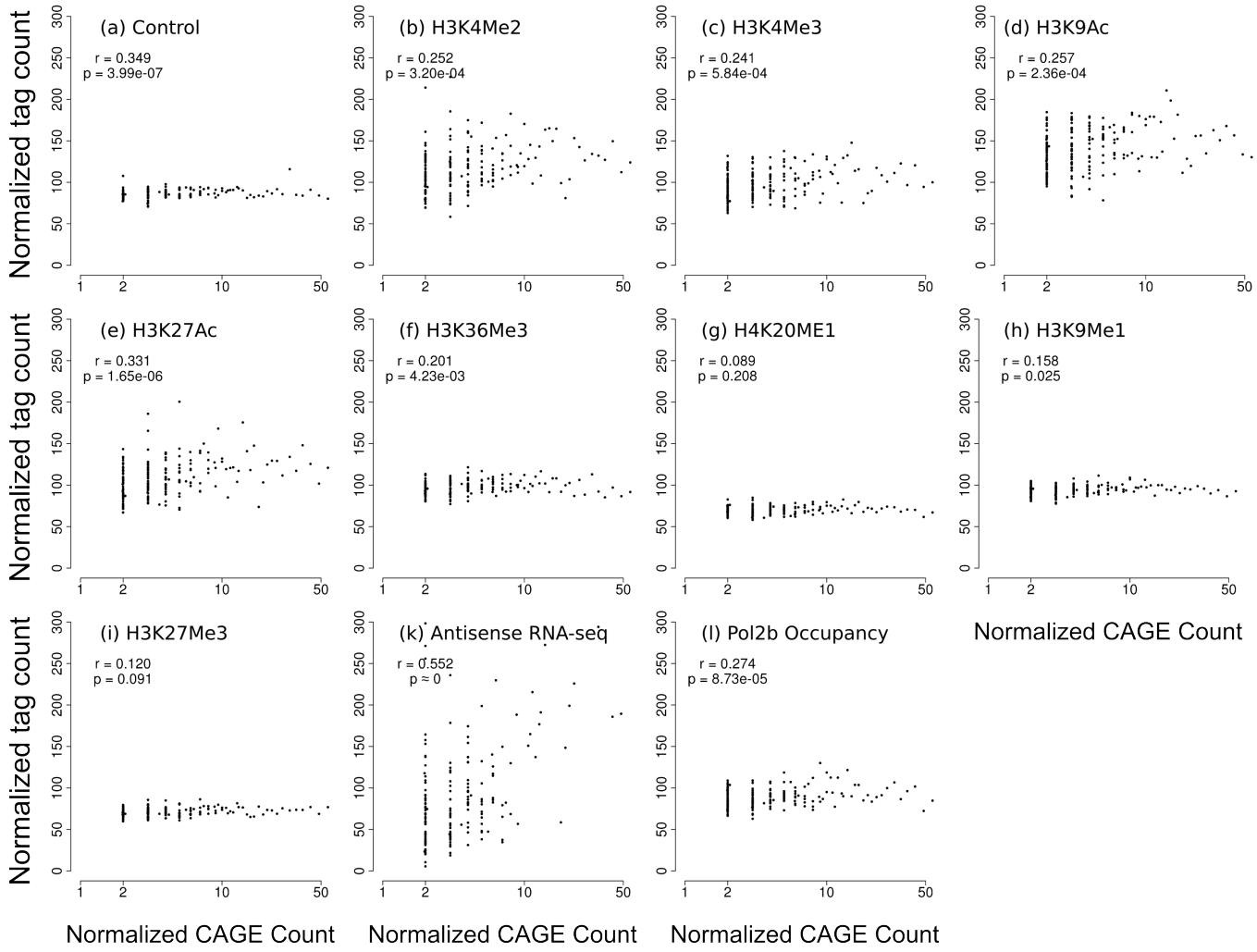
Supplementary Figure S22. Correlation between cis-NAT promoter activity and local chromatin environment in HepG2. Cis-NAT promoters in the HepG2 cell type were identified using CAGE data from non-polyadenylated nucleus isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis- NAT promoter activity and local histone modifications.



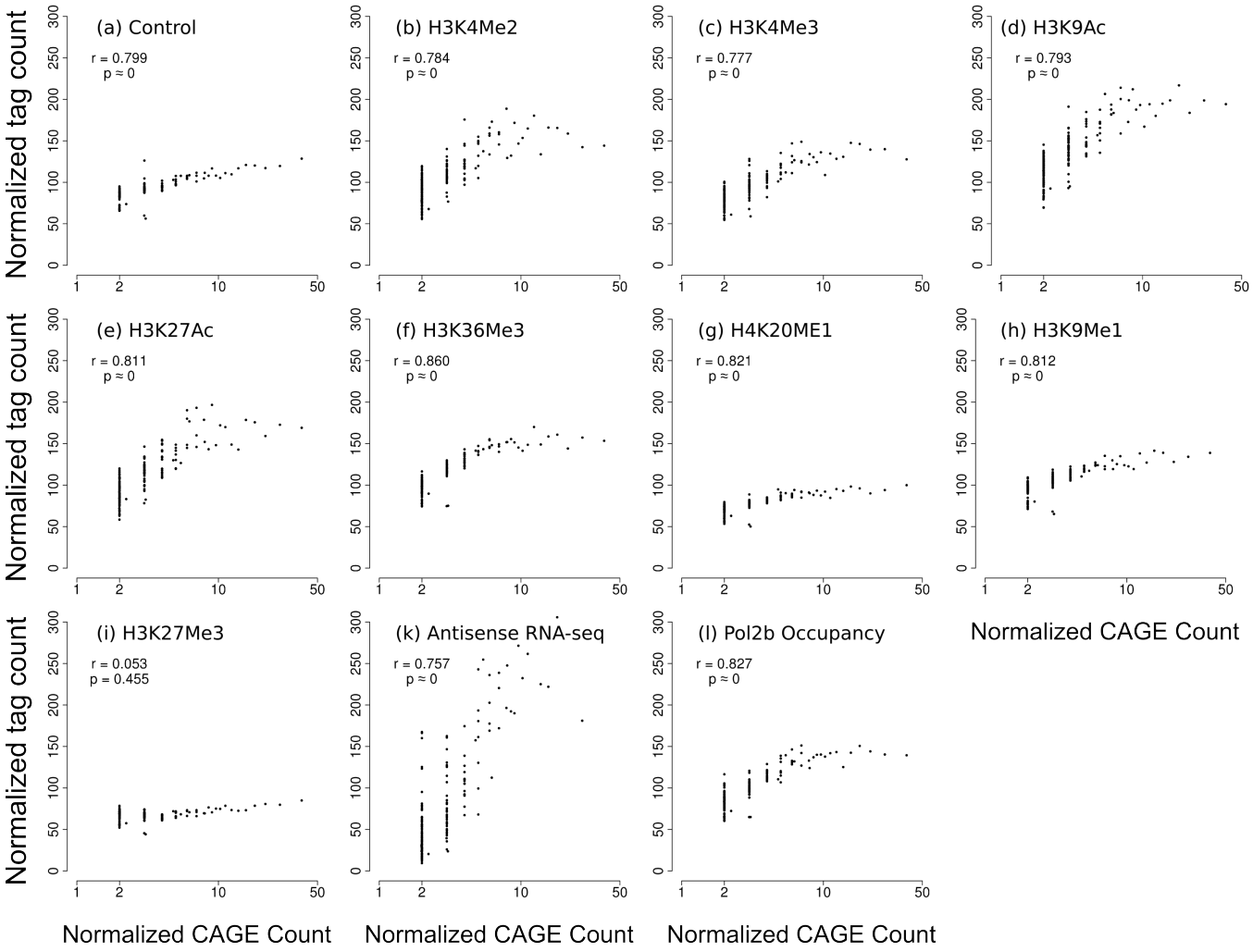
Supplementary Figure S23. Correlation between cis-NAT promoter activity and local chromatin environment in HUVEC. Cis-NAT promoters in the HUVEC cell type were identified using CAGE data from non-polyadenylated cytosol isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications and RNA Pol II binding +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis-NAT promoter activity and local histone modifications or RNA Pol II occupancy.



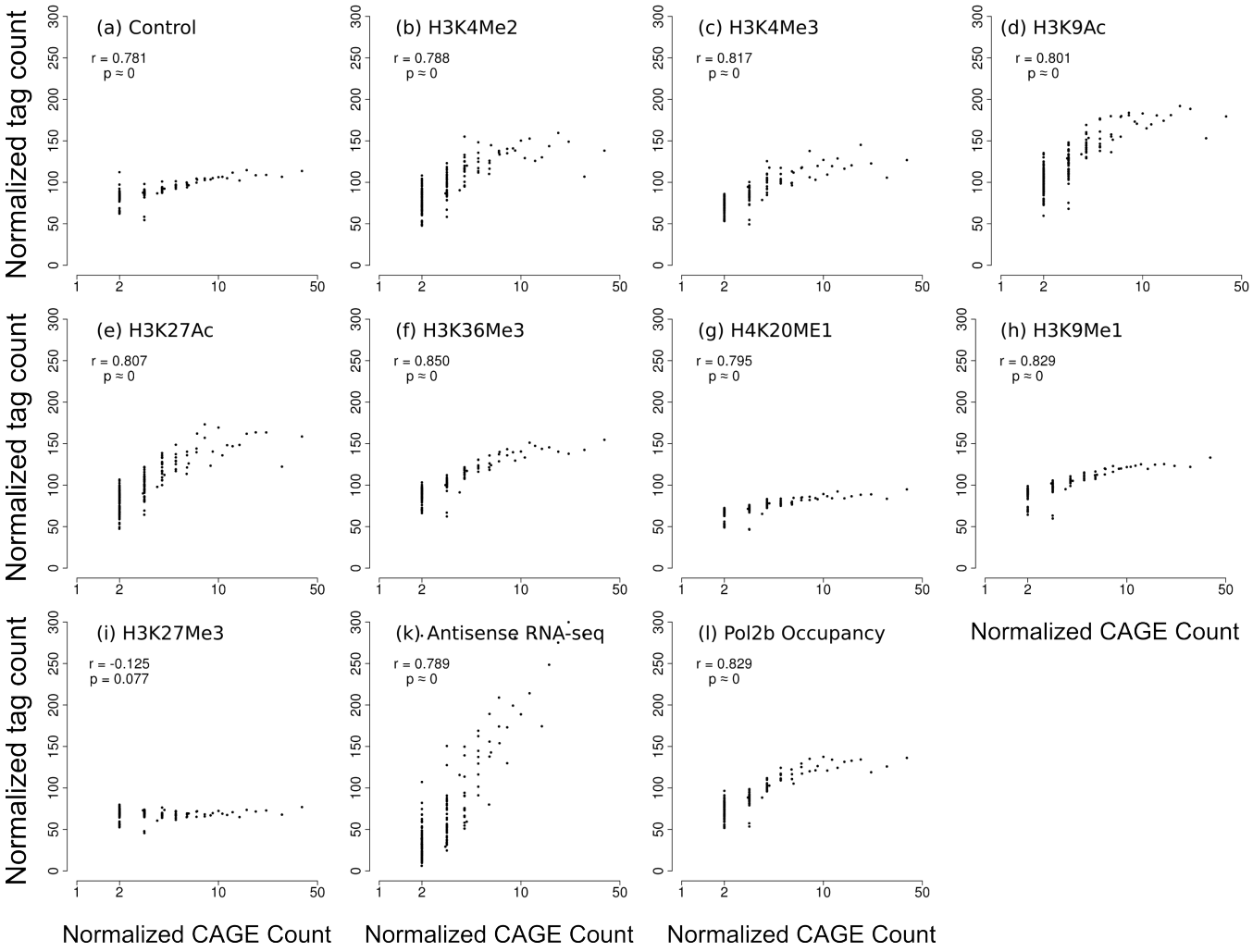
Supplementary Figure S24. Correlation between cis-NAT promoter activity and local chromatin environment in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from non-polyadenylated cytosol isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications and RNA Pol II binding +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis-NAT promoter activity and local histone modifications or RNA Pol II occupancy.



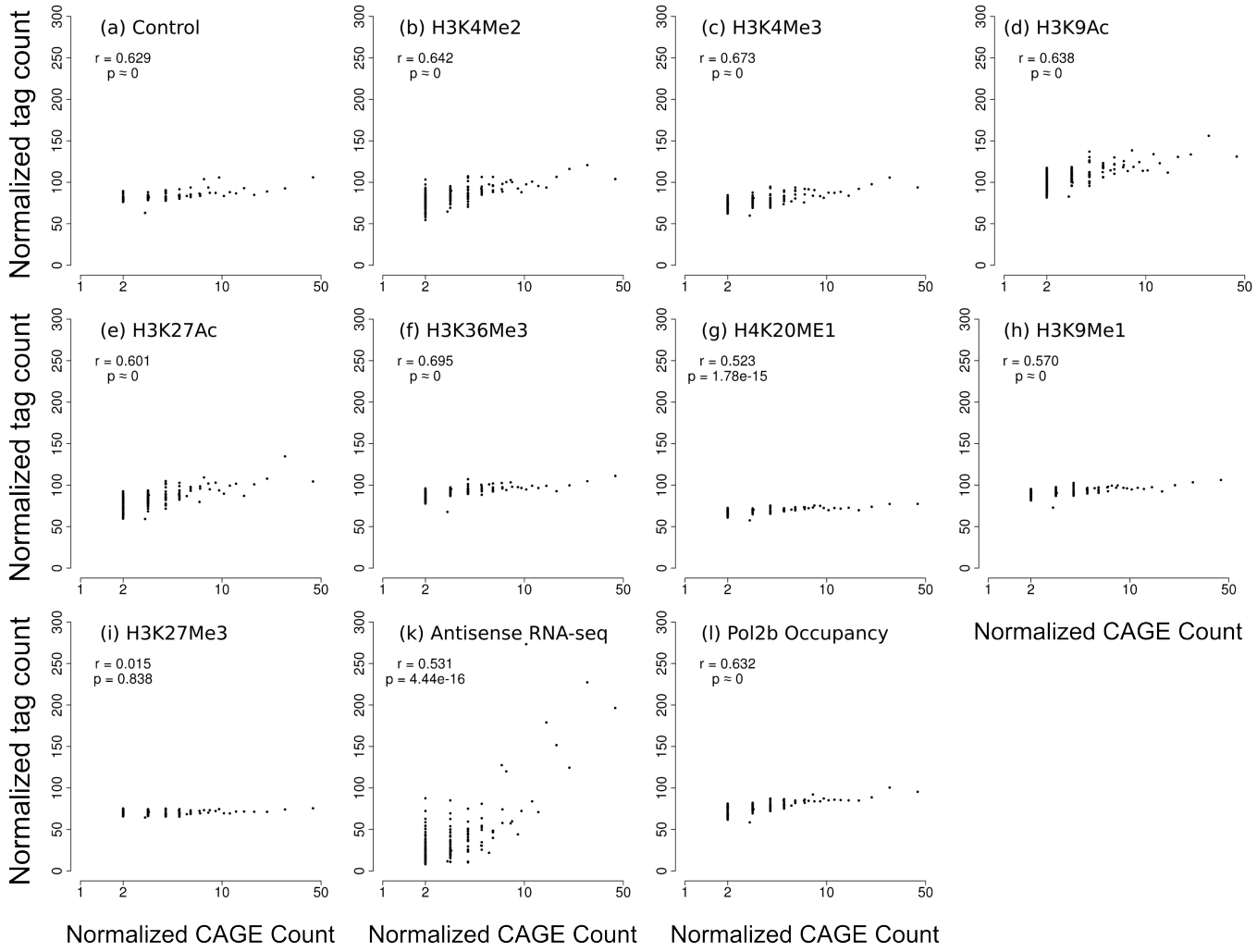
Supplementary Figure S25. Correlation between cis-NAT promoter activity and local chromatin environment in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from polyadenylated cytosol isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications and RNA Pol II binding +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis-NAT promoter activity and local histone modifications or RNA Pol II occupancy.



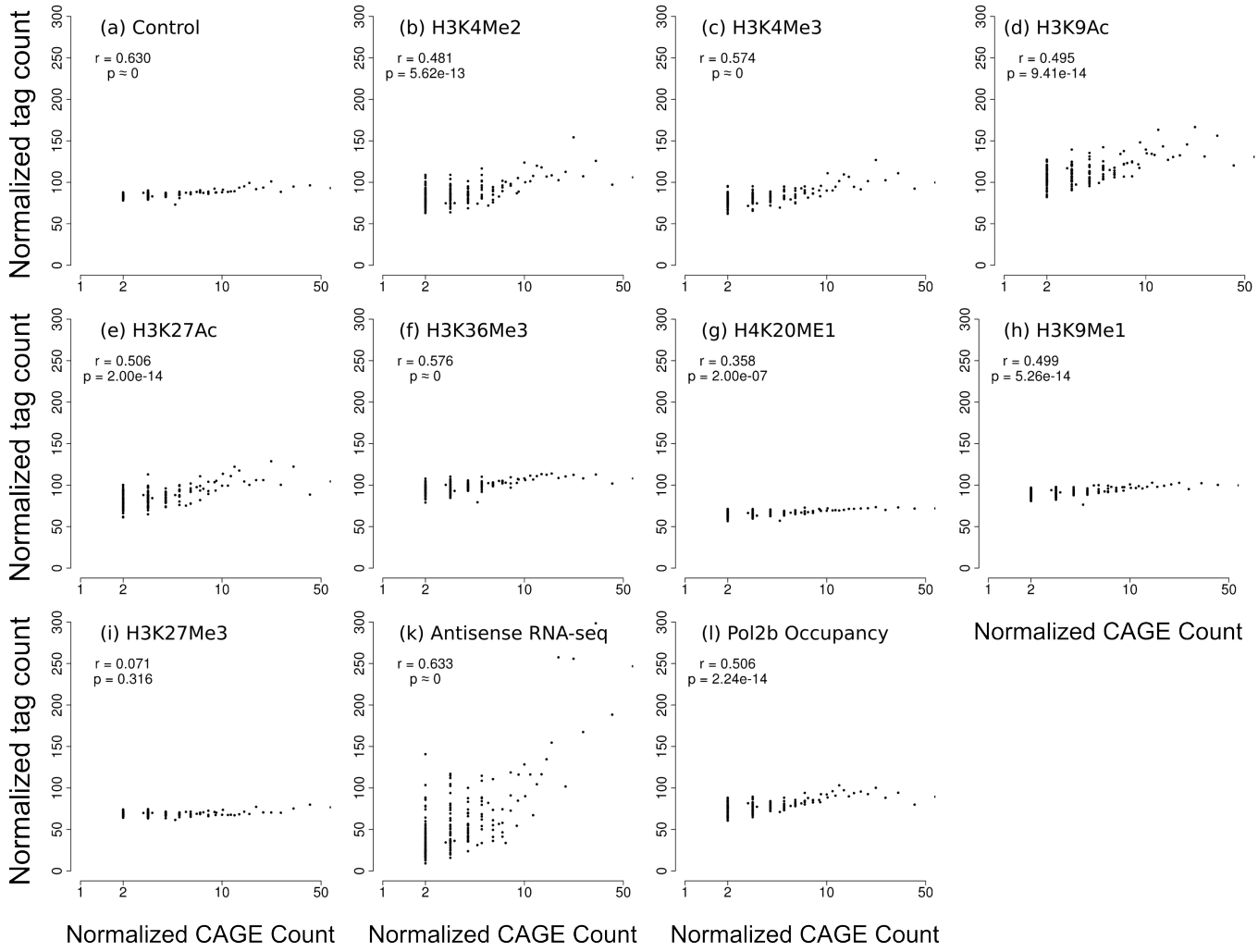
Supplementary Figure S26. Correlation between cis-NAT promoter activity and local chromatin environment in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from total nucleolus isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications and RNA Pol II binding +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis- NAT promoter activity and local histone modifications or RNA Pol II occupancy.



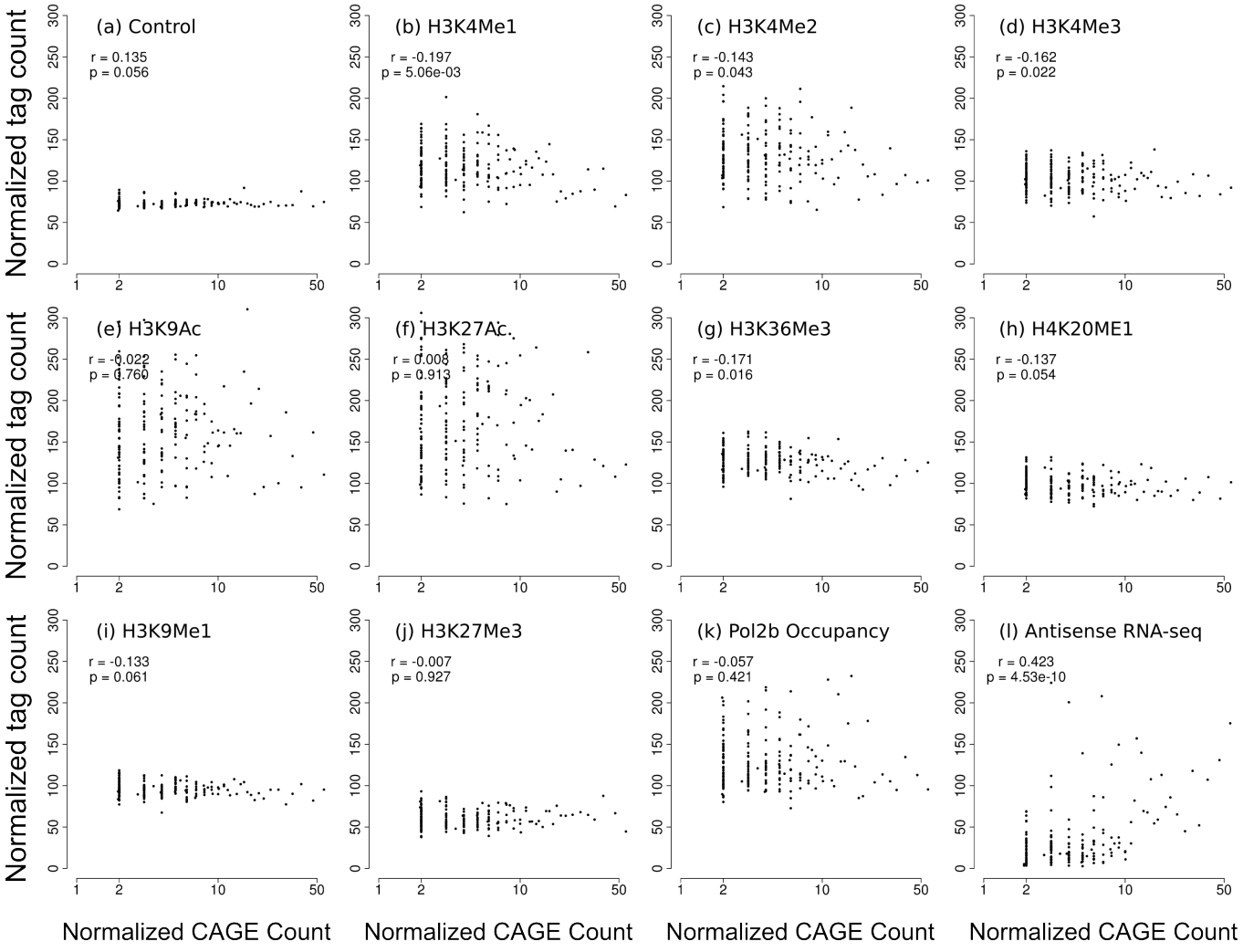
Supplementary Figure S27. Correlation between cis-NAT promoter activity and local chromatin environment in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from total nucleoplasm isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications and RNA Pol II binding +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis-NAT promoter activity and local histone modifications or RNA Pol II occupancy.



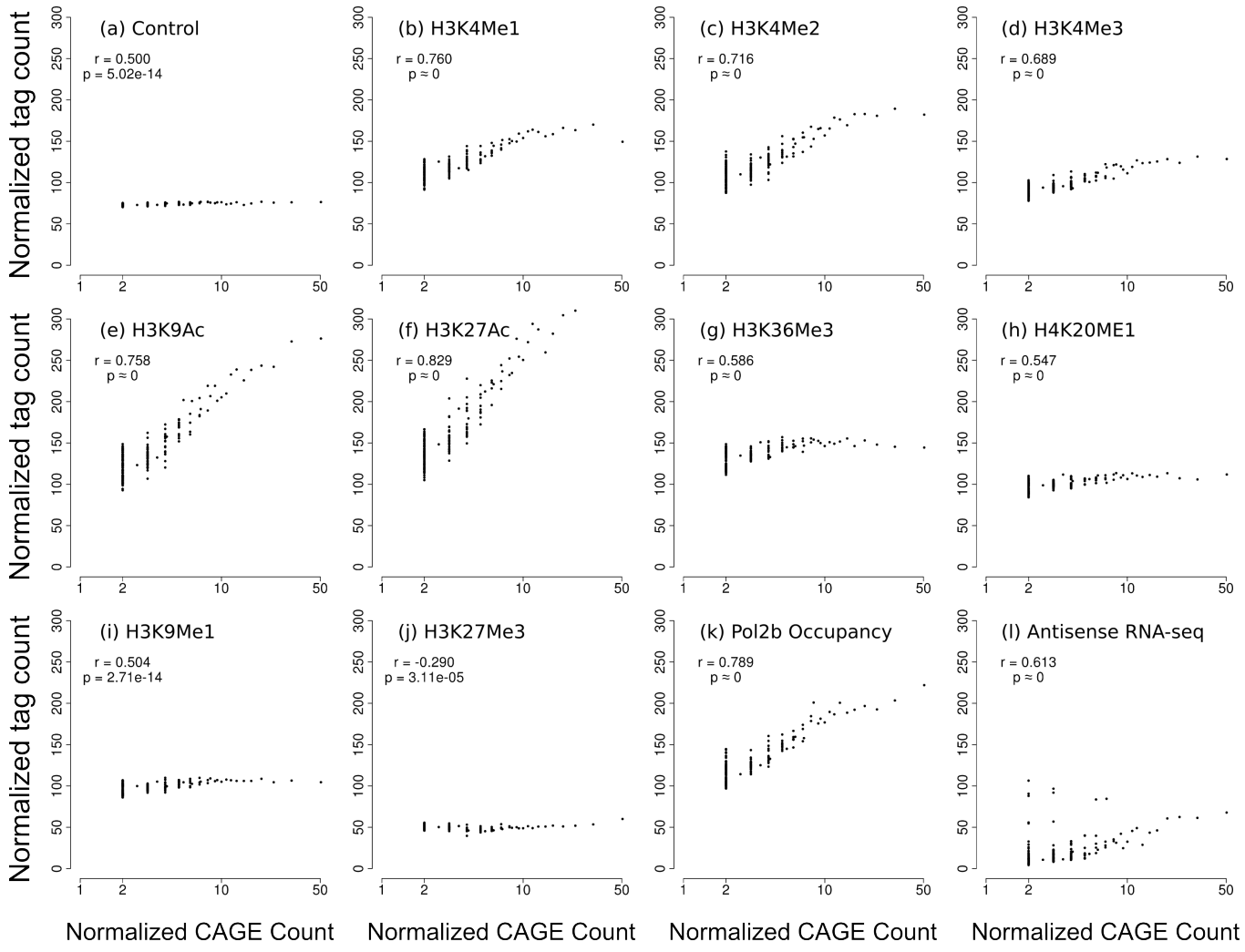
Supplementary Figure S28. Correlation between cis-NAT promoter activity and local chromatin environment in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from non-polyadenylated nucleus isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications and RNA Pol II binding +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis-NAT promoter activity and local histone modifications or RNA Pol II occupancy.



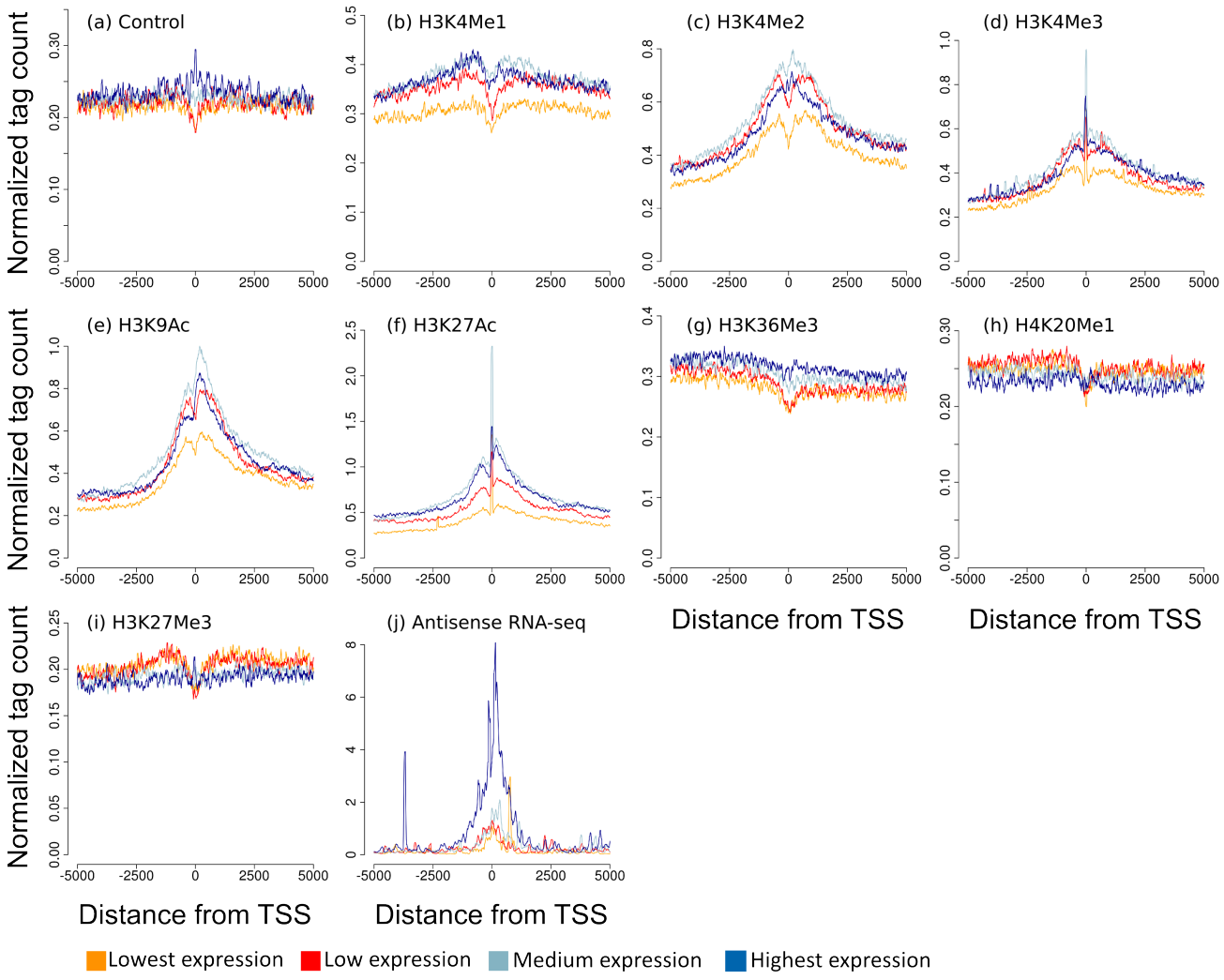
Supplementary Figure S29. Correlation between cis-NAT promoter activity and local chromatin environment in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from polyadenylated nucleus isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications and RNA Pol II binding +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis-NAT promoter activity and local histone modifications or RNA Pol II occupancy.



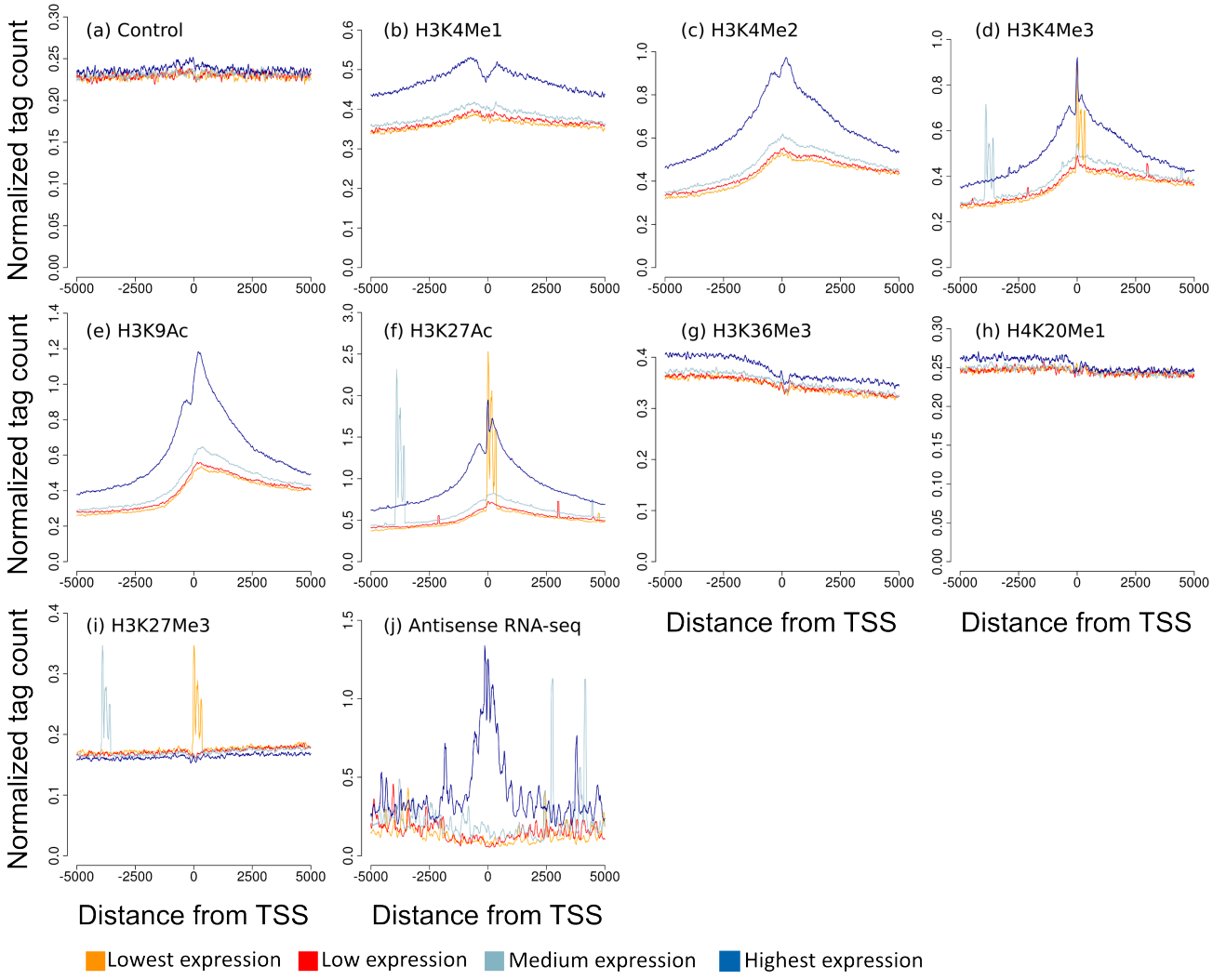
Supplementary Figure S30. Correlation between cis-NAT promoter activity and local chromatin environment in NHEK. Cis-NAT promoters in the NHEK cell type were identified using CAGE data from non-polyadenylated cytosol isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications and RNA Pol II binding +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis-NAT promoter activity and local histone modifications or RNA Pol II occupancy.



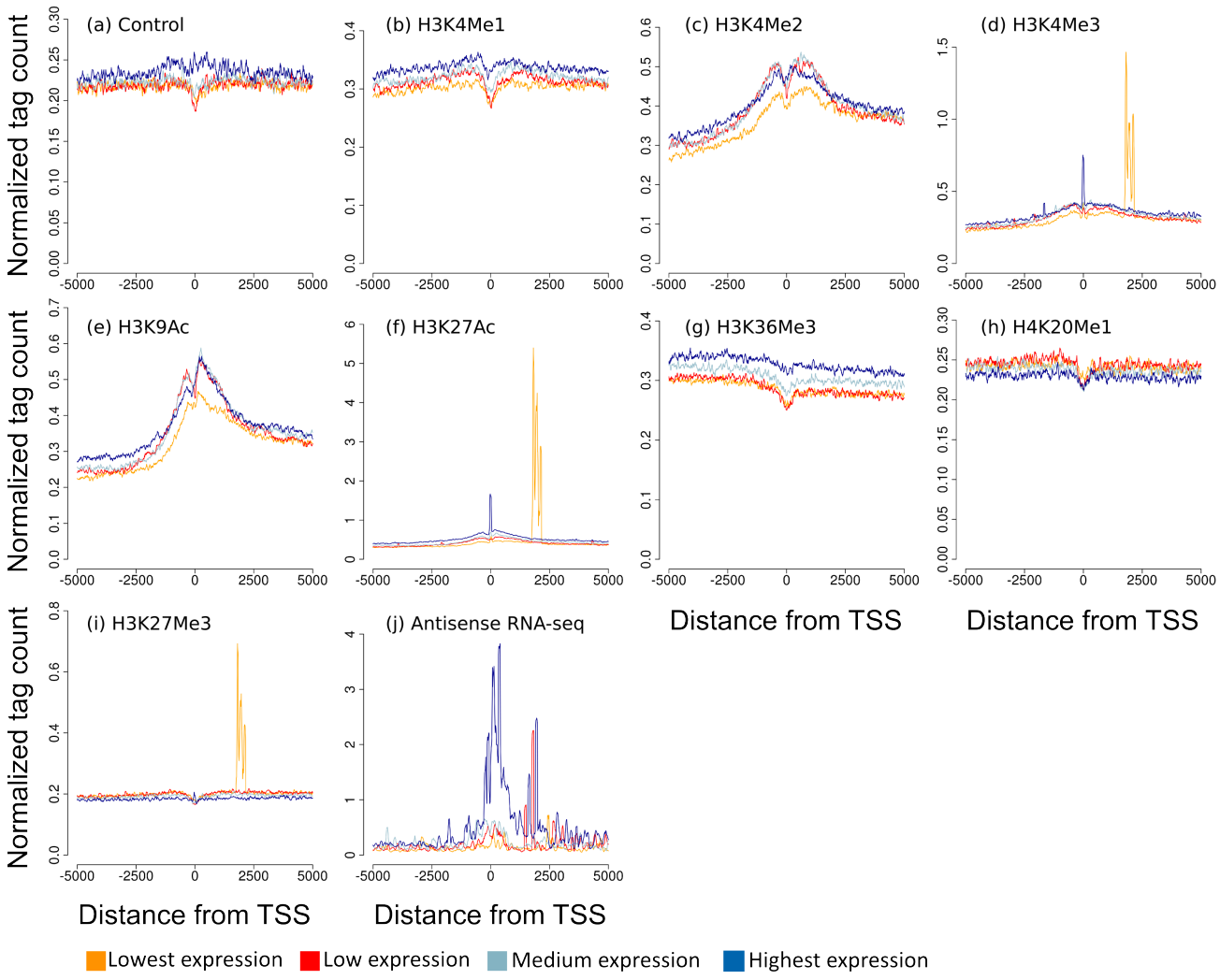
Supplementary Figure S31. Correlation between cis-NAT promoter activity and local chromatin environment in NHEK. Cis-NAT promoters in the NHEK cell type were identified using CAGE data from non-polyadenylated nucleus isolates. Cis-NAT promoters were divided into 200 bins based on activity, as measured by CAGE tags counts, and the normalized average numbers of ChIP-seq reads for histone modifications and RNA Pol II binding +/- 5kb of the cis-NAT TSS were calculated. Spearman rank correlations were used to determine the relationship between cis-NAT promoter activity and local histone modifications or RNA Pol II occupancy.



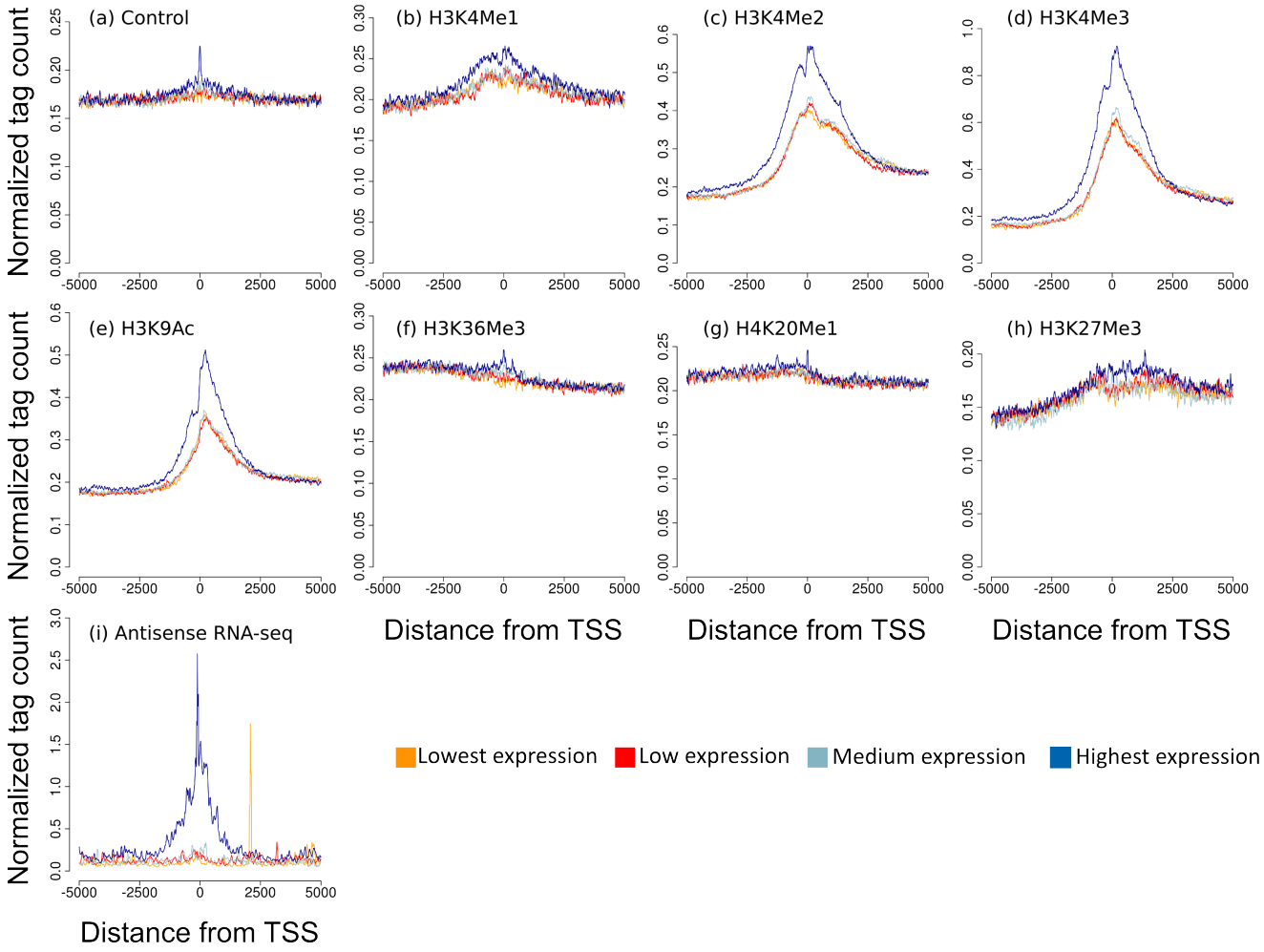
Supplementary Figure S32. Chromatin modification environment around cis-NAT promoters in GM12878. Cis-NAT promoters in the GM12878 cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



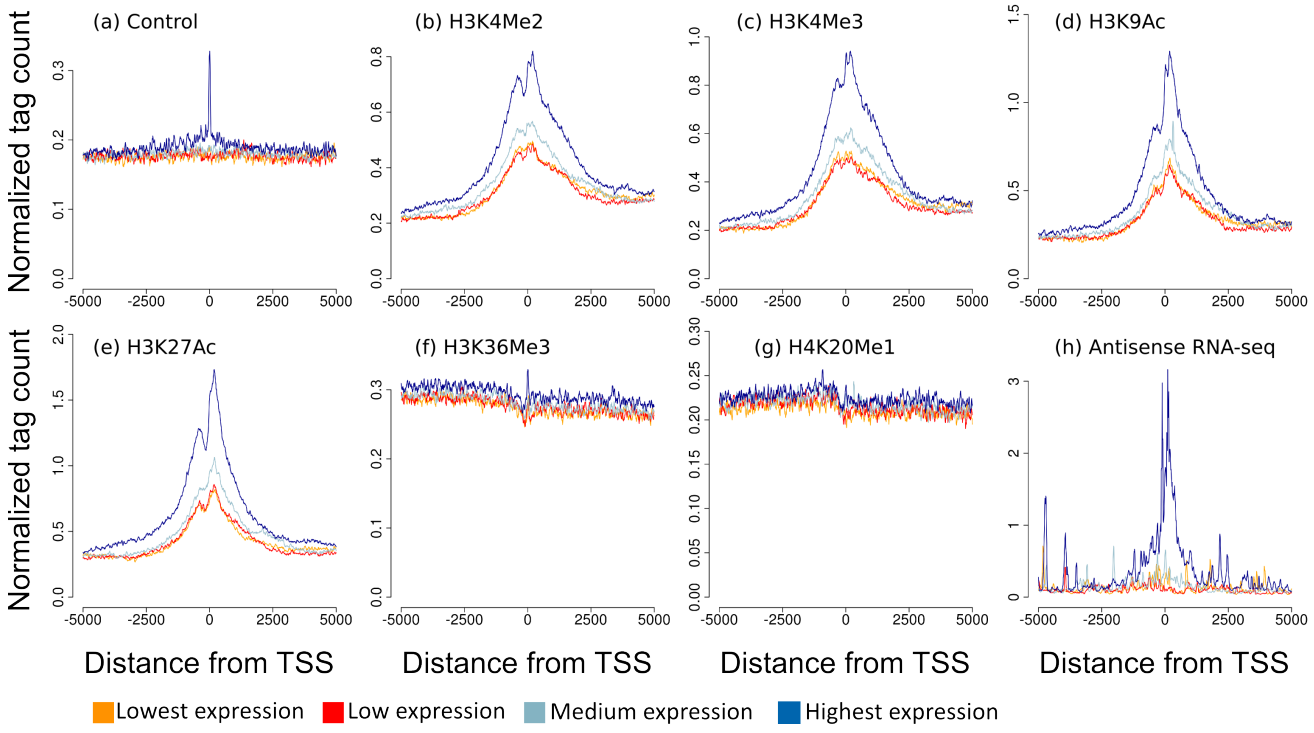
Supplementary Figure S33. Chromatin modification environment around cis-NAT promoters in GM12878. Cis-NAT promoters in the GM12878 cell type were identified using CAGE data from total RNA from nucleolus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



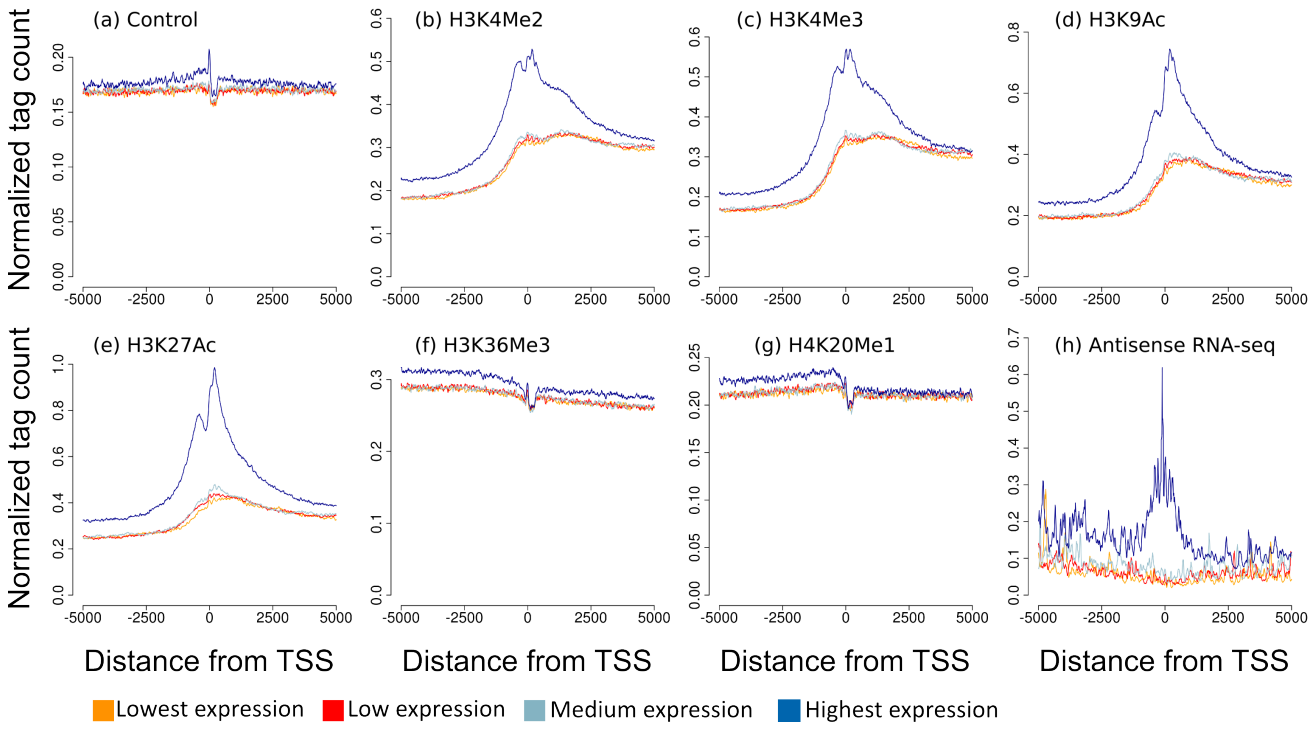
Supplementary Figure S34. Chromatin modification environment around cis-NAT promoters in GM12878. Cis-NAT promoters in the GM12878 cell type were identified using CAGE data from non-polyadenylated RNA from nucleus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



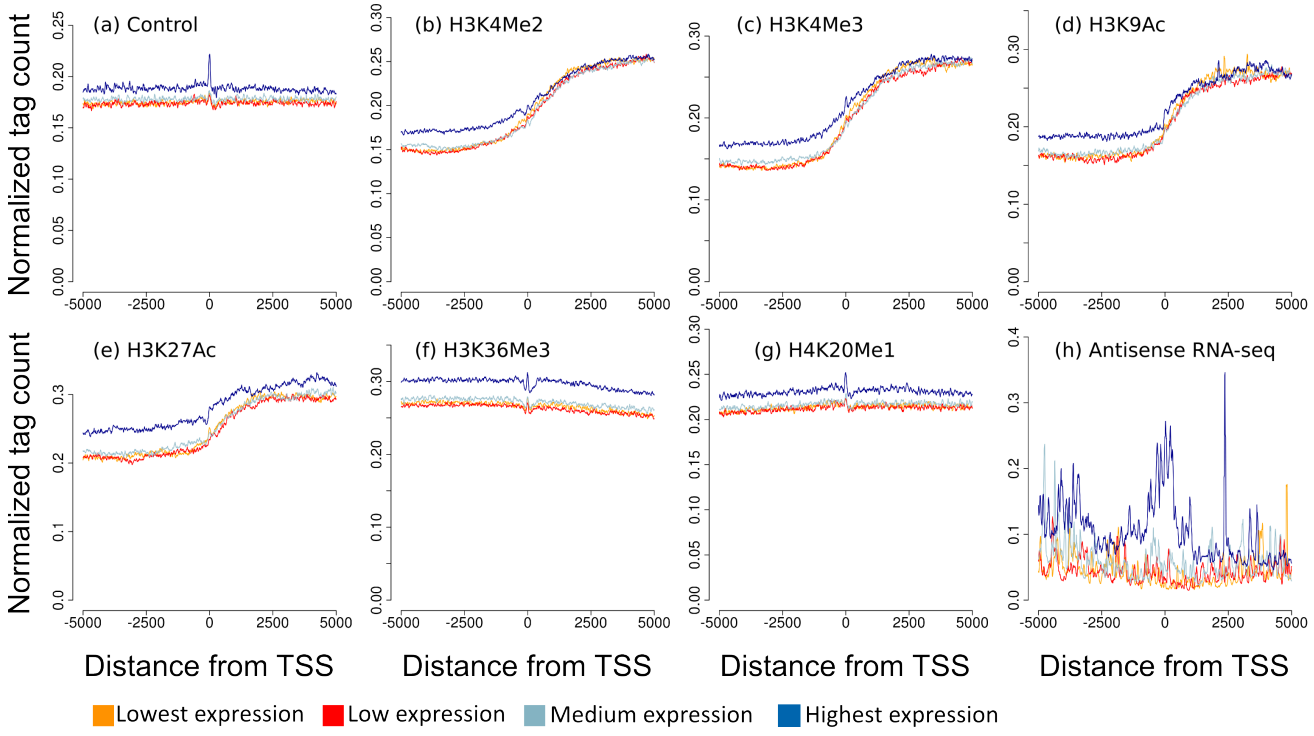
Supplementary Figure S35. Chromatin modification environment around cis-NAT promoters in H1HESC. Cis-NAT promoters in the H1HESC cell type were identified using CAGE data from non-polyadenylated RNA from whole-cell isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



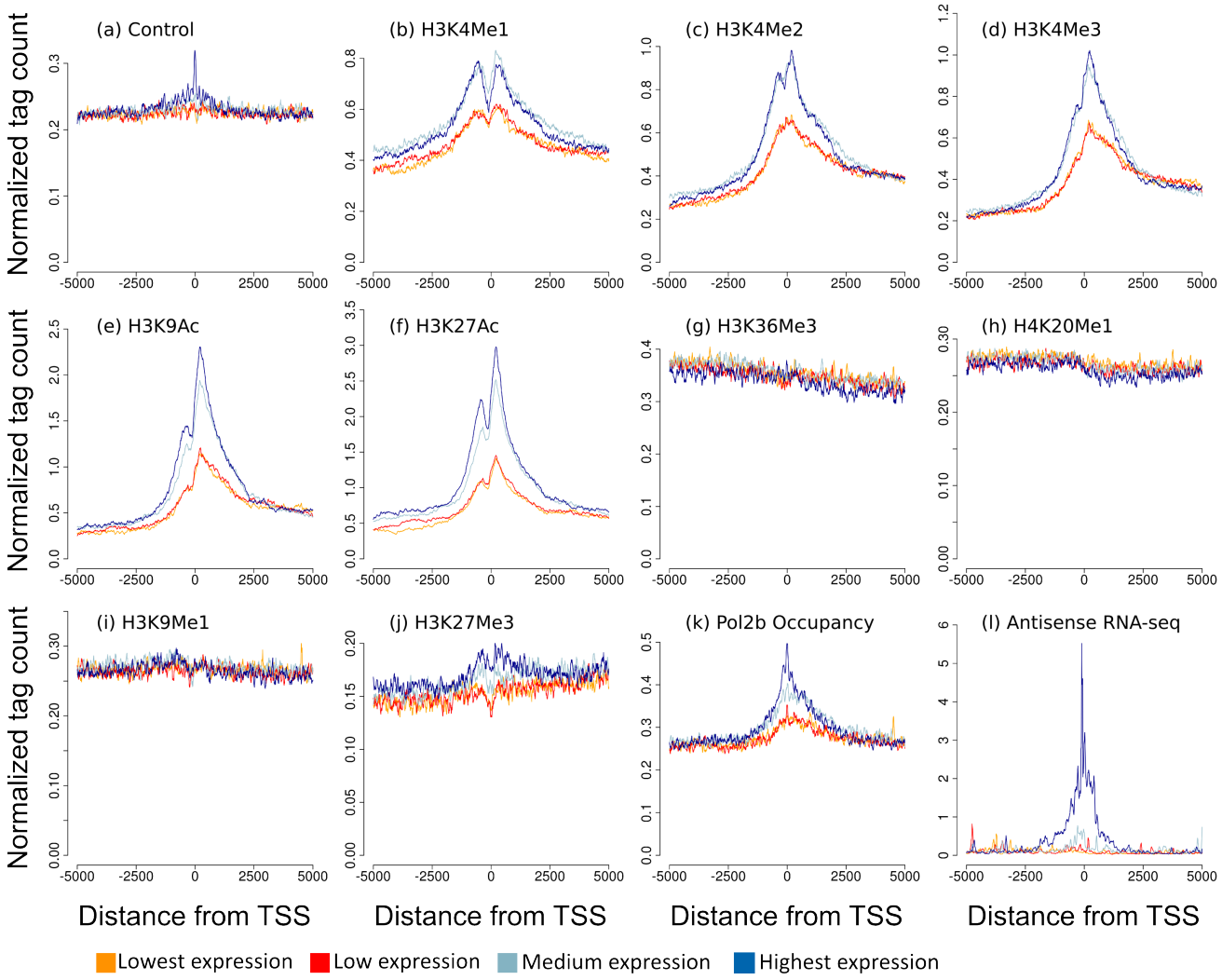
Supplementary Figure S36. Chromatin modification environment around cis-NAT promoters in HepG2. Cis-NAT promoters in the HepG2 cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



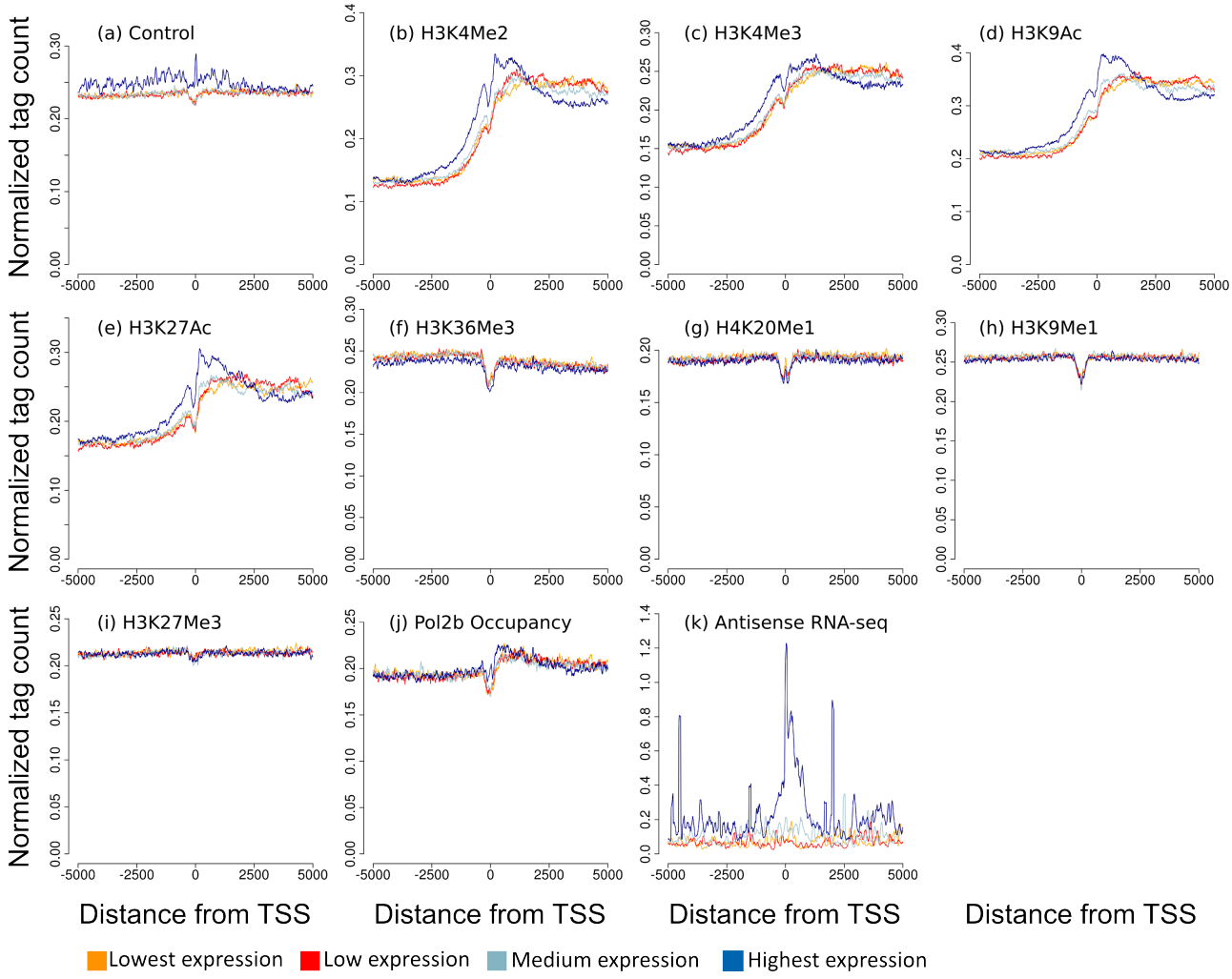
Supplementary Figure S37. Chromatin modification environment around cis-NAT promoters in HepG2. Cis-NAT promoters in the HepG2 cell type were identified using CAGE data from total RNA from nucleolus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



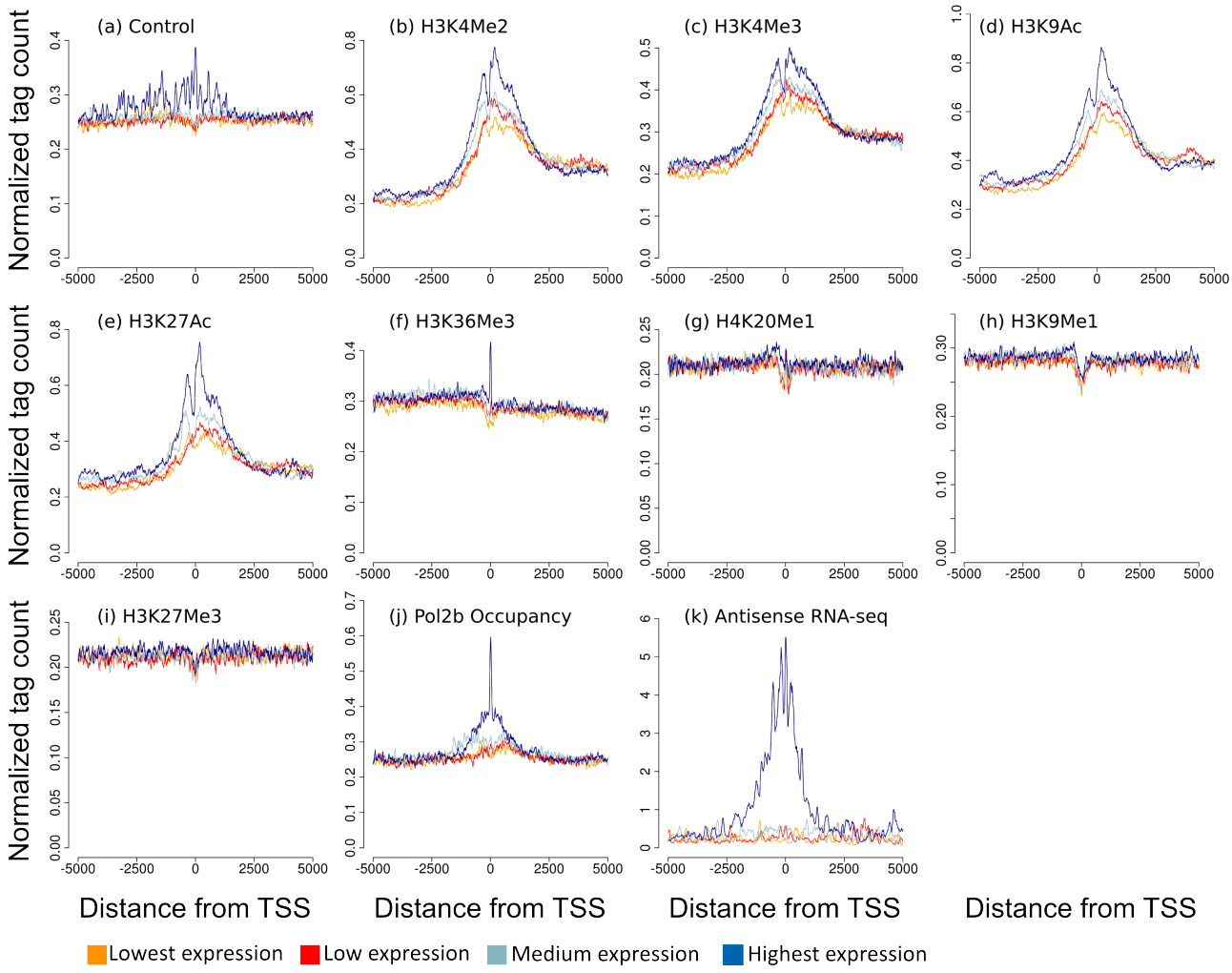
Supplementary Figure S38. Chromatin modification environment around cis-NAT promoters in HepG2. Cis-NAT promoters in the HepG2 cell type were identified using CAGE data from non-polyadenylated RNA from nucleus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



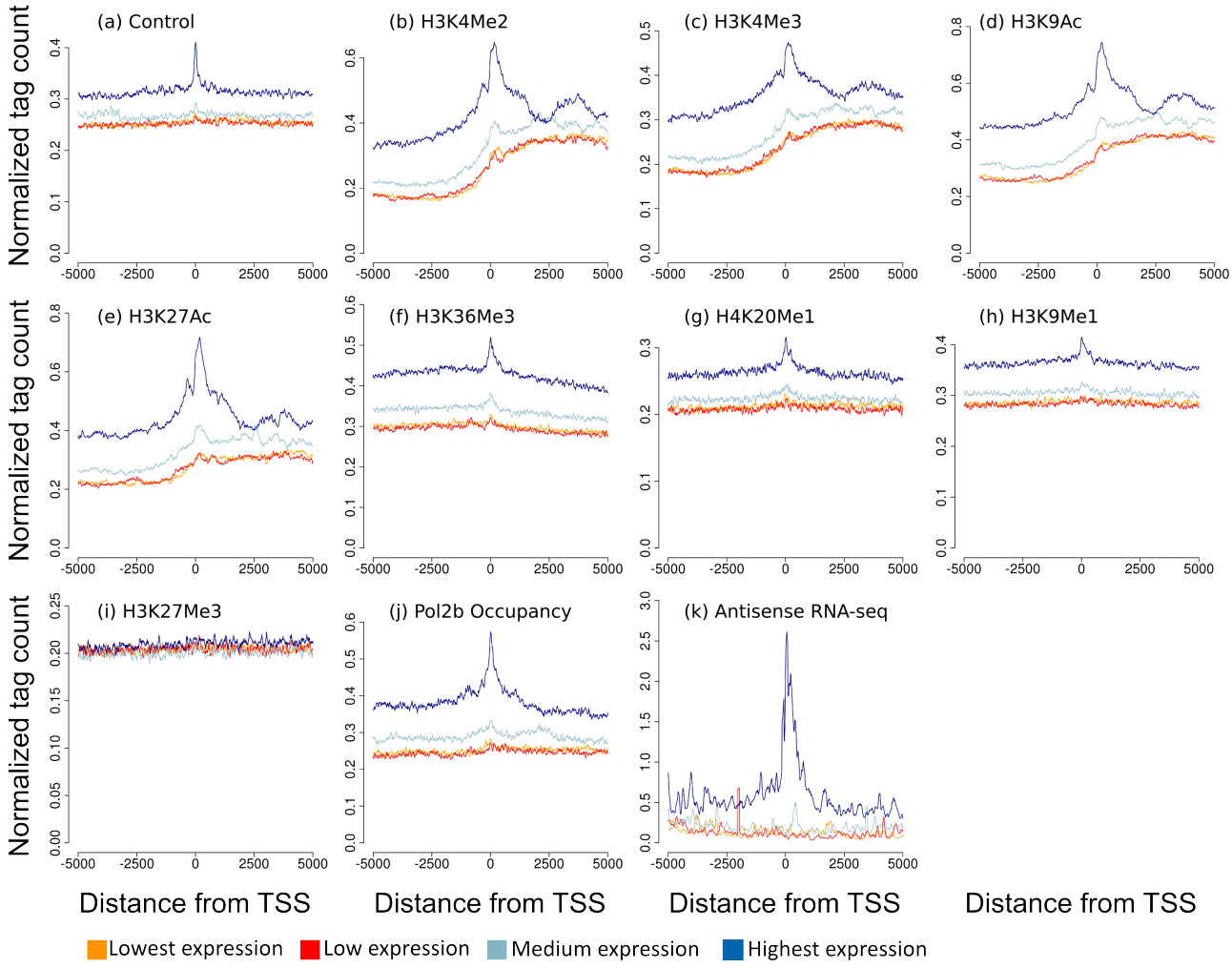
Supplementary Figure S39. Chromatin modification environment around cis-NAT promoters in HUVEC. Cis-NAT promoters in the HUVEC cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



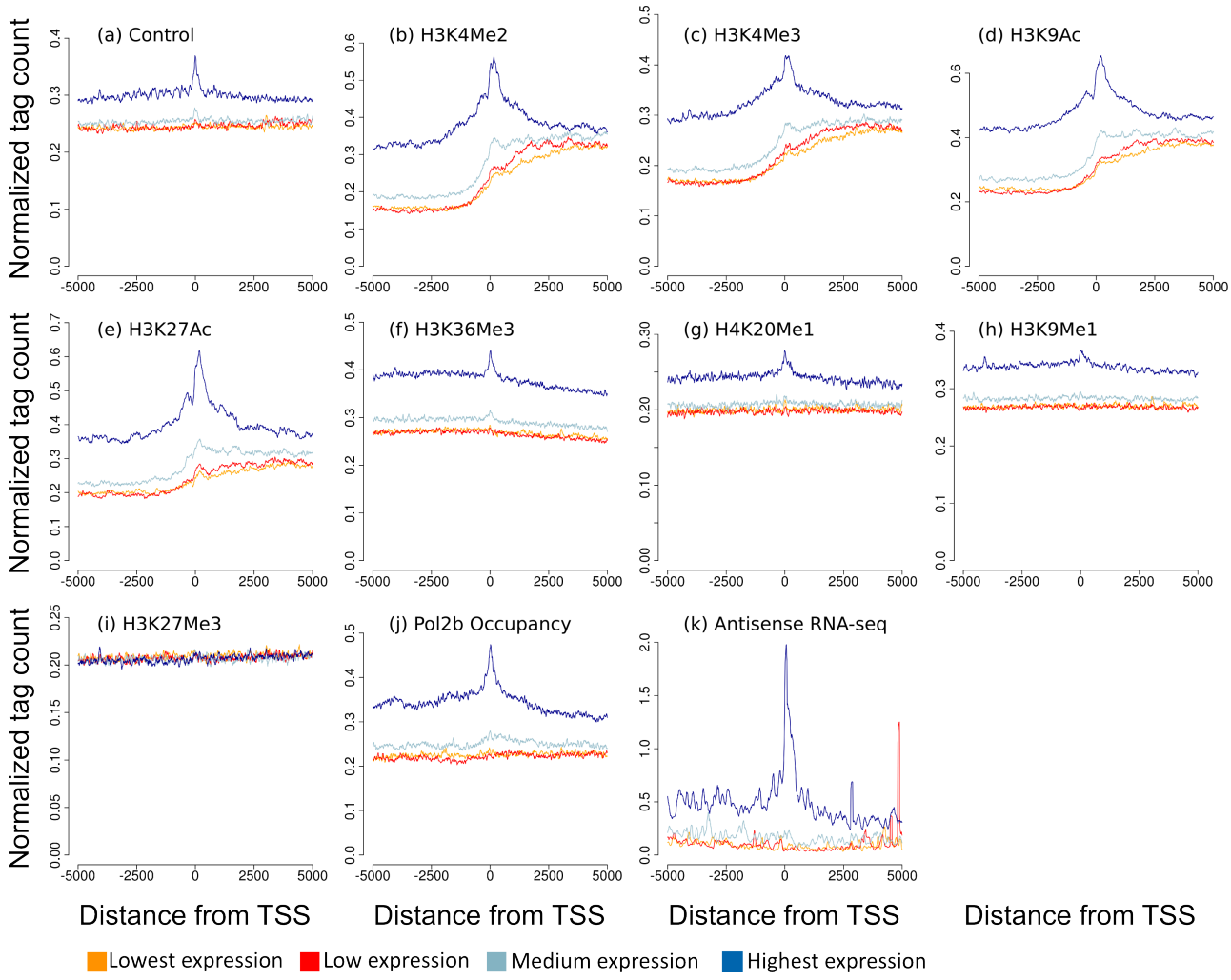
Supplementary Figure S40. Chromatin modification environment around cis-NAT promoters in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



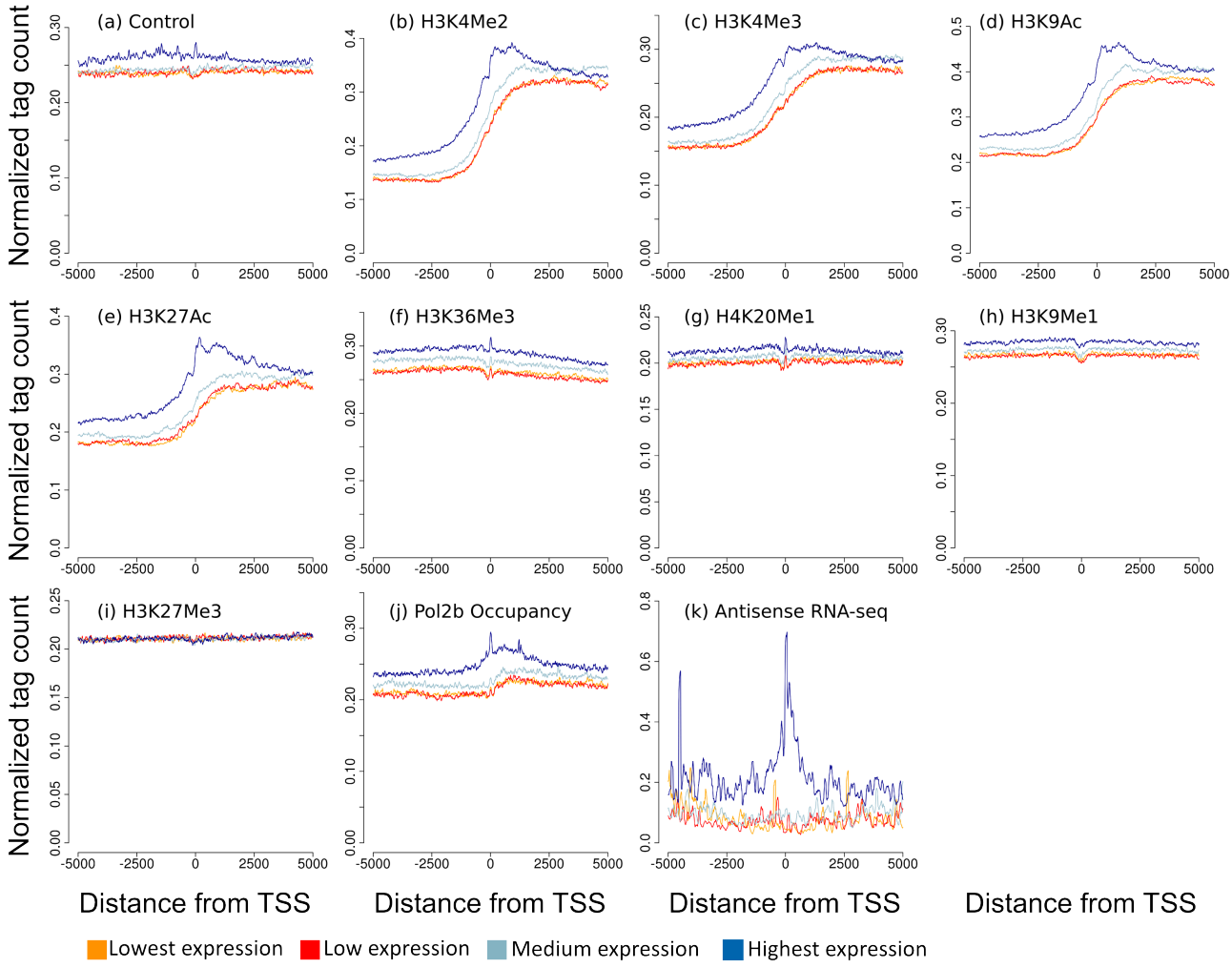
Supplementary Figure S41. Chromatin modification environment around cis-NAT promoters in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from polyadenylated RNA from cytosol isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



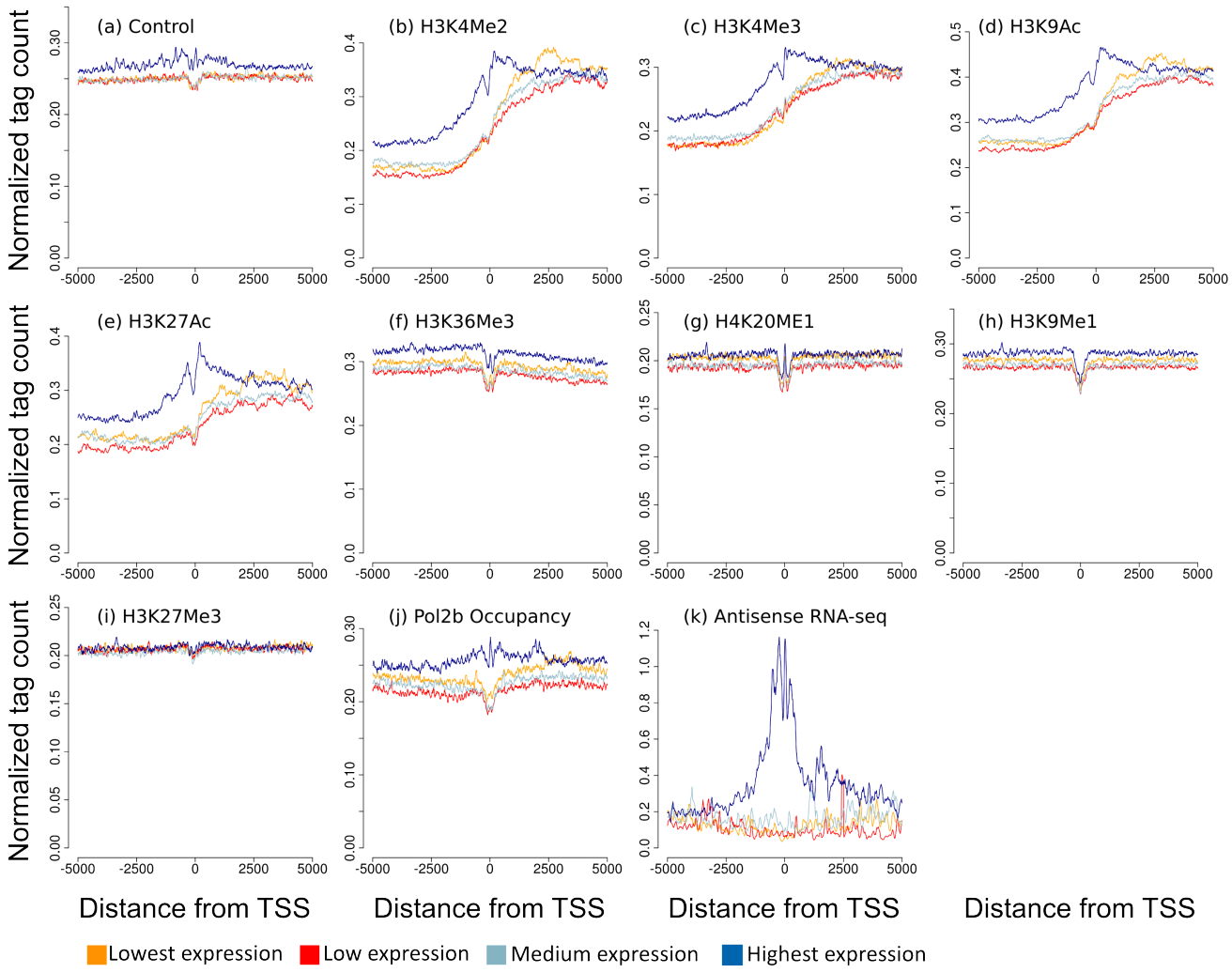
Supplementary Figure S42. Chromatin modification environment around cis-NAT promoters in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from total RNA from nucleolus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



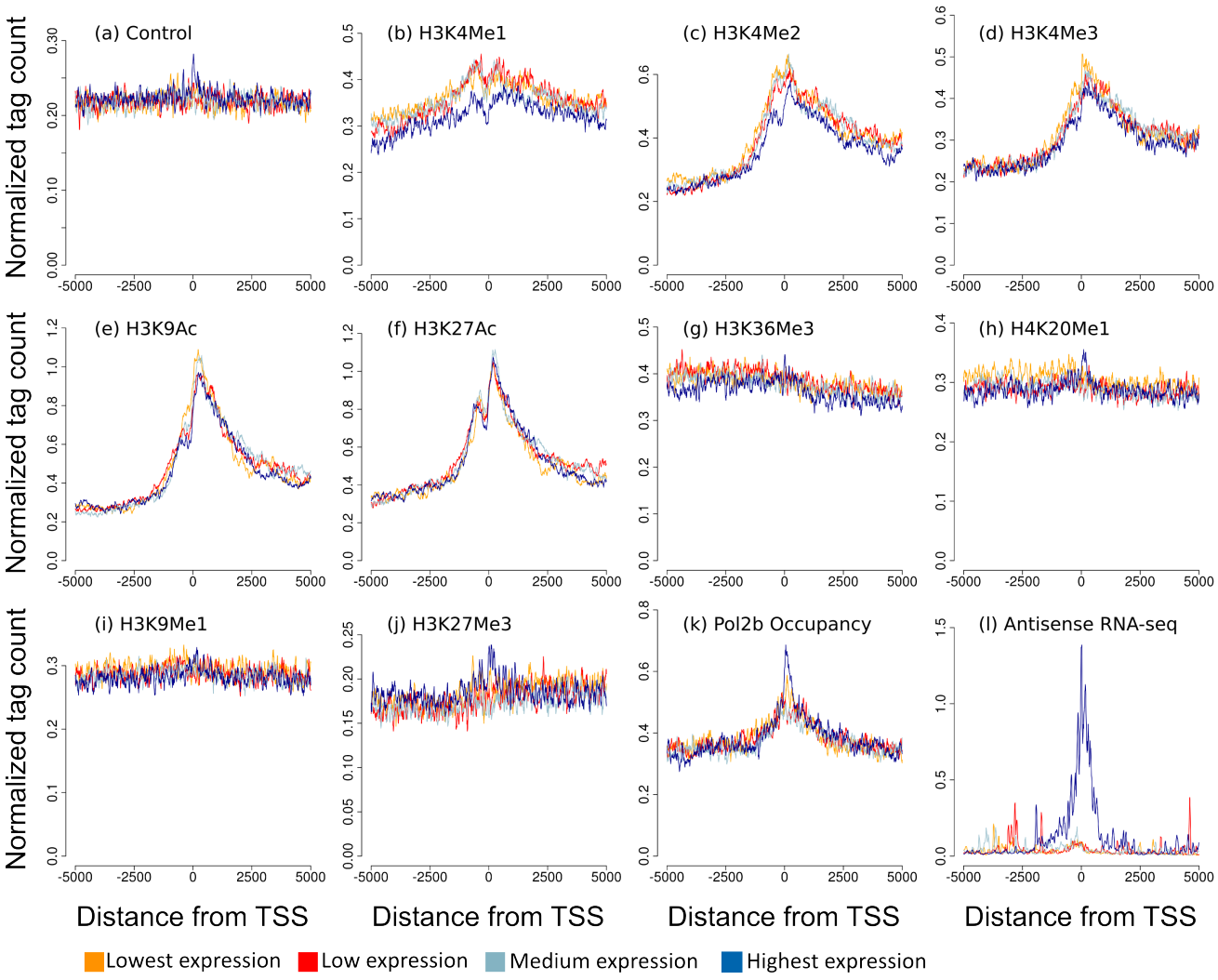
Supplementary Figure S43. Chromatin modification environment around cis-NAT promoters in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from total RNA from nucleoplasm isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



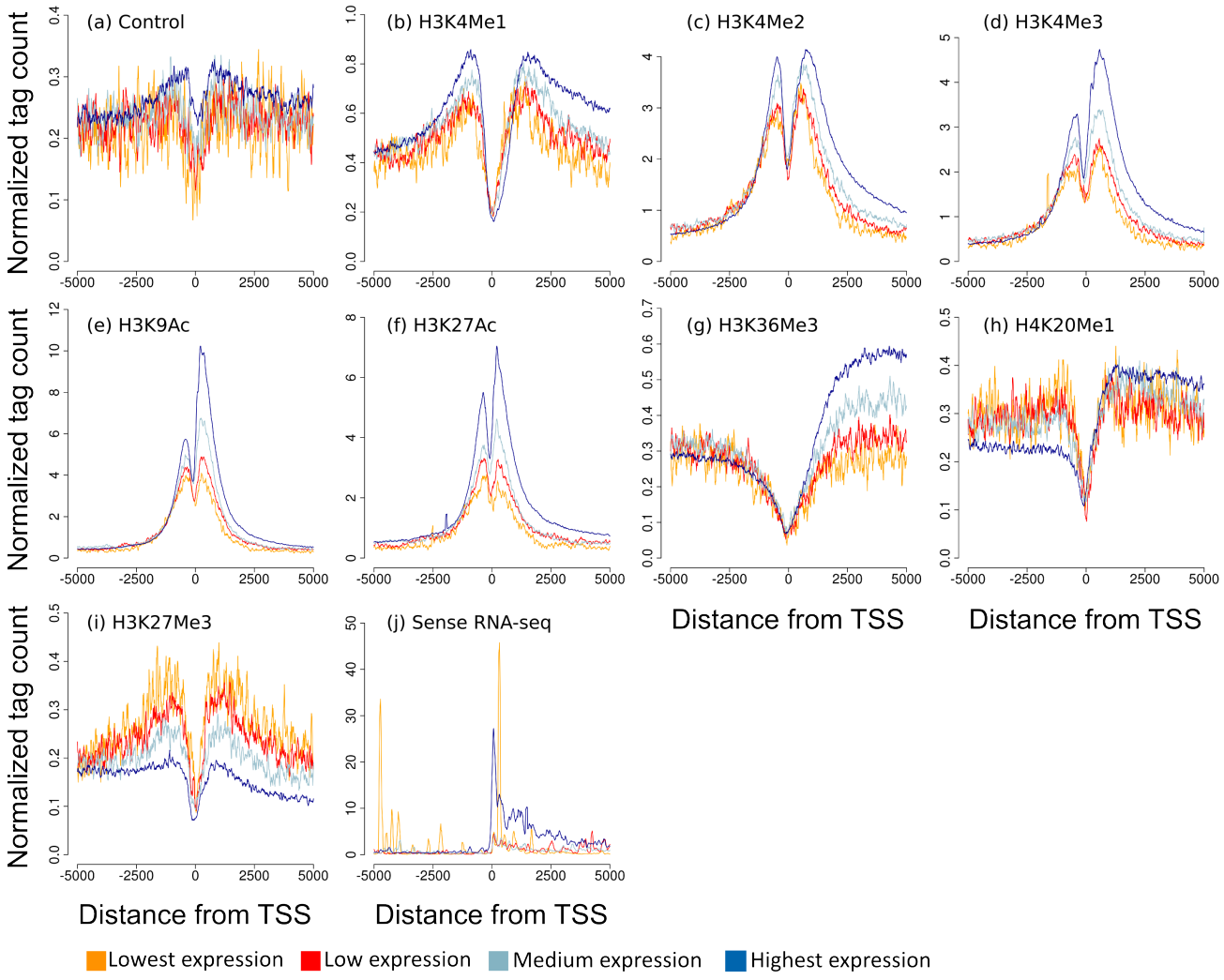
Supplementary Figure S44. Chromatin modification environment around cis-NAT promoters in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from non-polyadenylated RNA from nucleus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



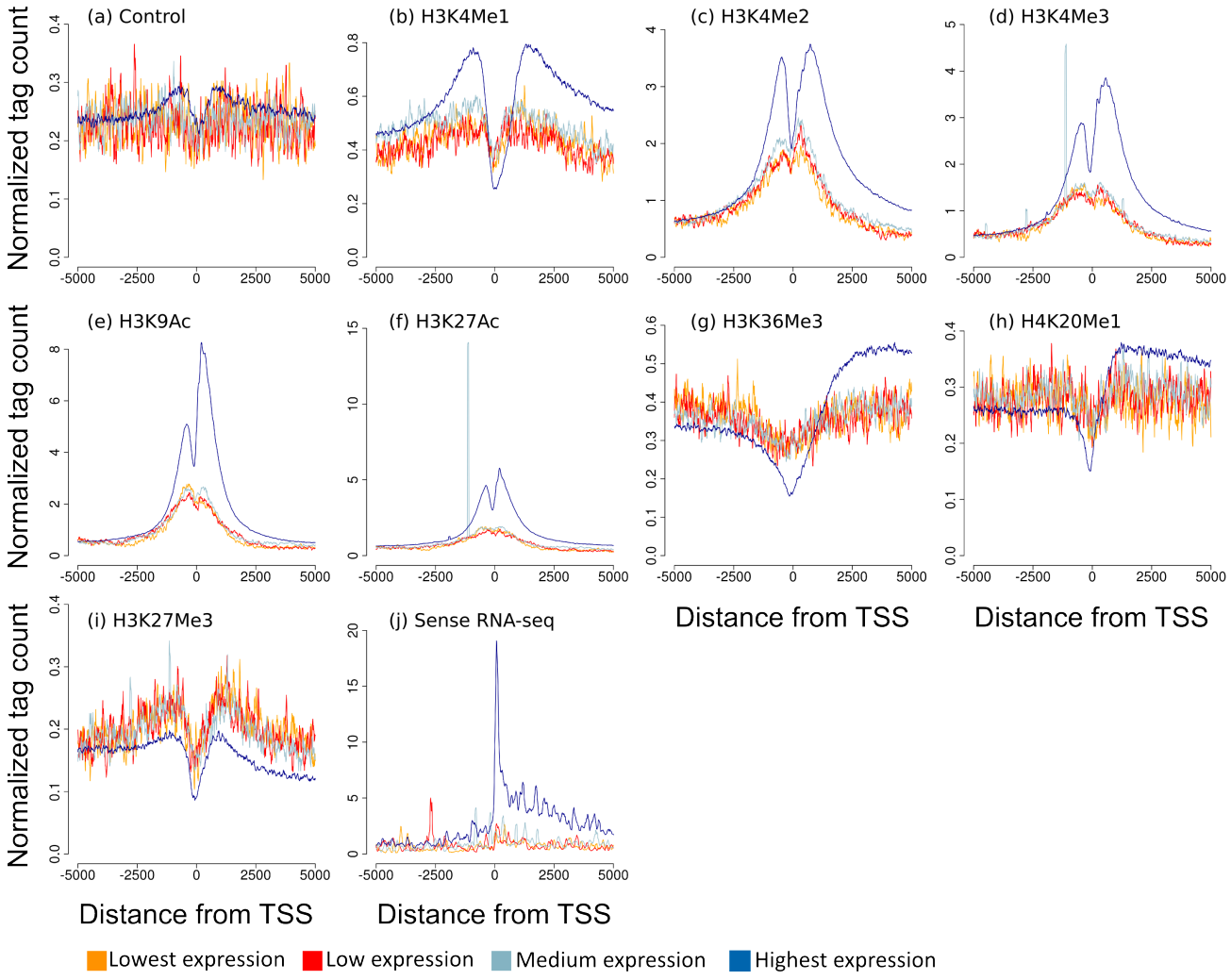
Supplementary Figure S45. Chromatin modification environment around cis-NAT promoters in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from polyadenylated RNA from nucleus isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



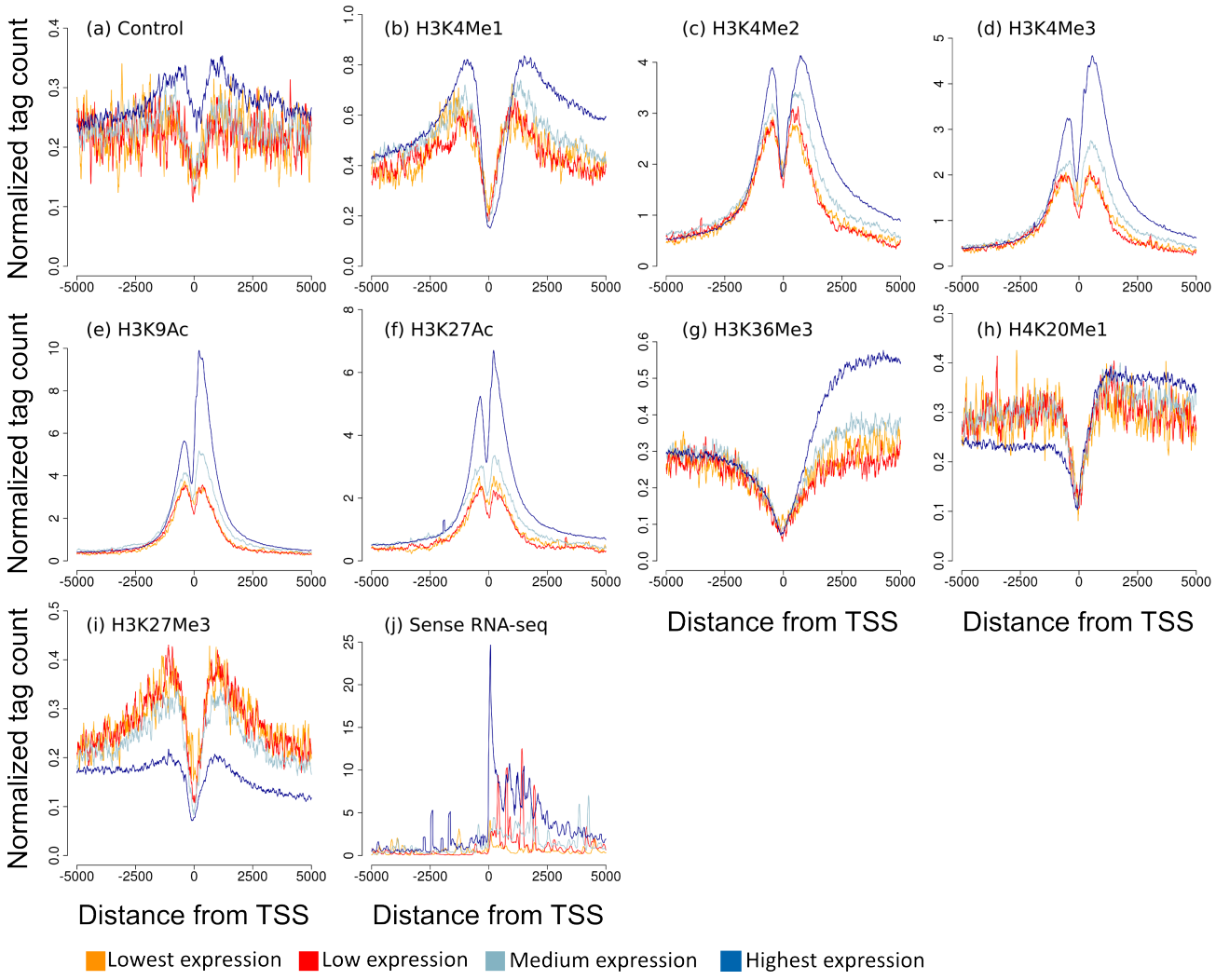
Supplementary Figure S46. Chromatin modification environment around cis-NAT promoters in NHEK. Cis-NAT promoters in the NHEK cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Cis-NAT promoters were divided into 4 bins based on their activity (lowest to highest), and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the cis-NAT TSS (at position 0) were calculated for each bin.



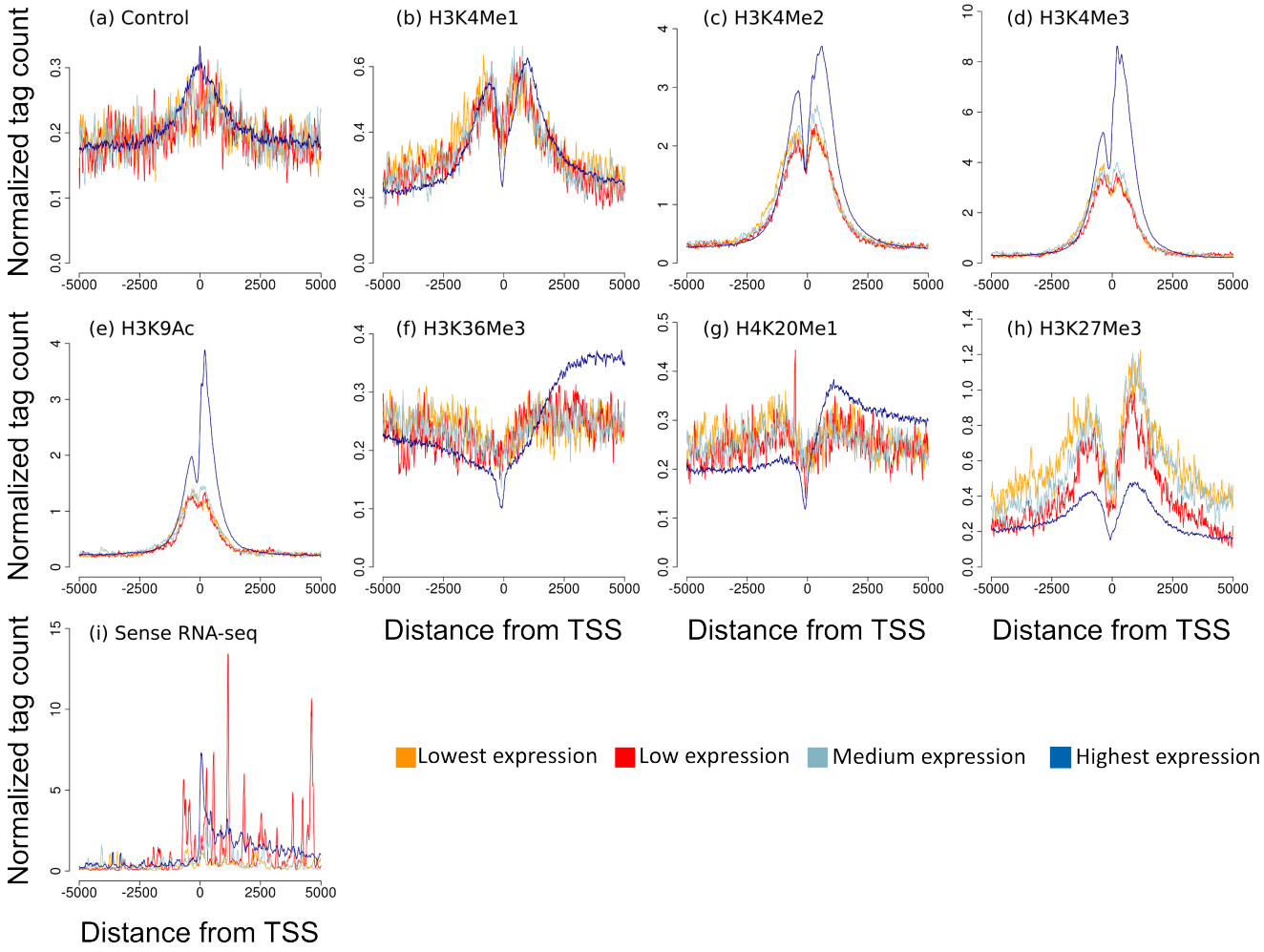
Supplementary Figure S47. Chromatin modification environment around genic promoters in GM12878. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from non-polyadenylated RNA from cytosol isolates from GM12878 cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



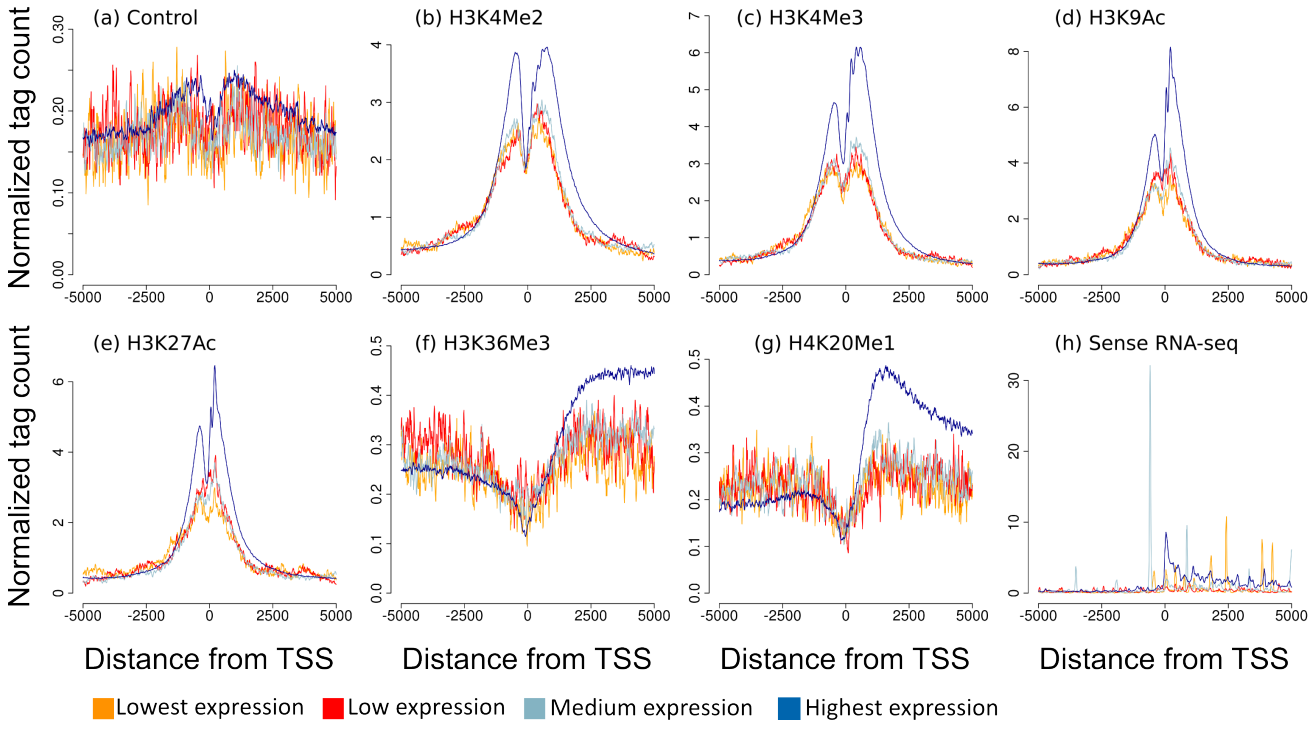
Supplementary Figure S48. Chromatin modification environment around genic promoters in GM12878. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from total RNA from nucleolar isolates from GM12878 cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



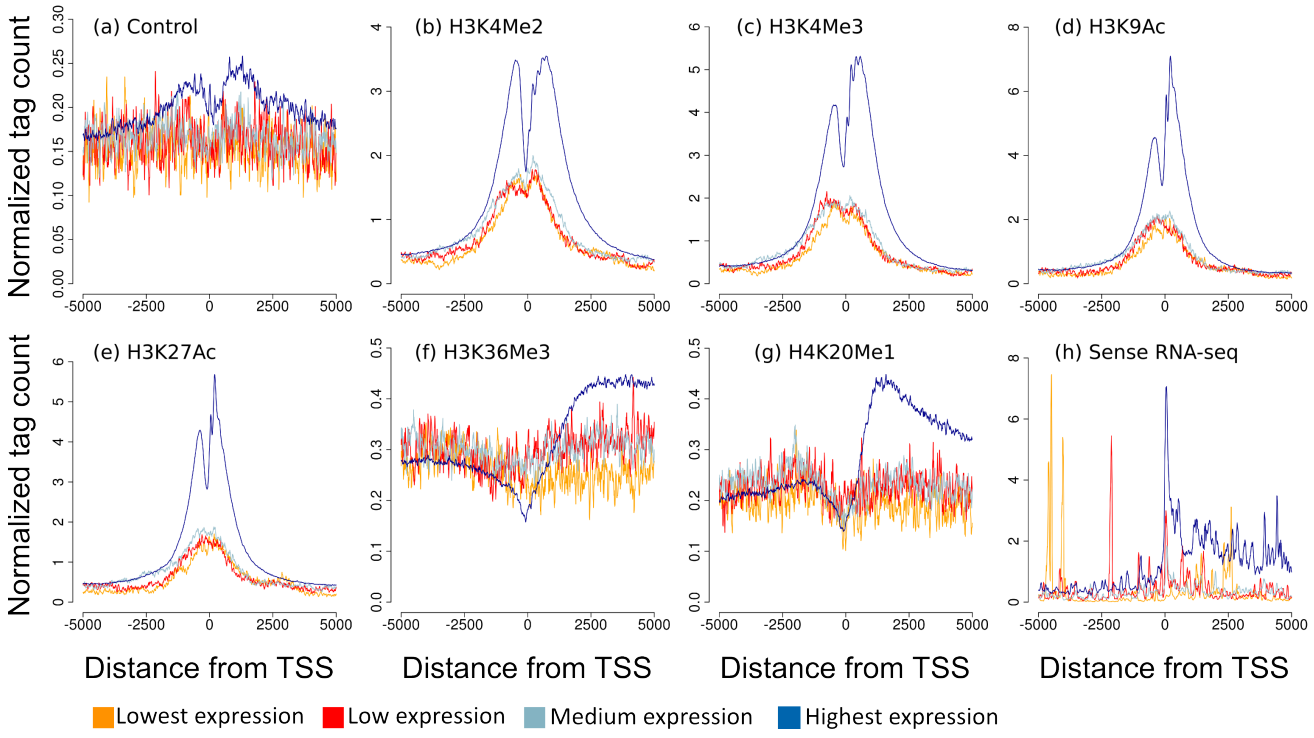
Supplementary Figure S49. Chromatin modification environment around genic promoters in GM12878. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from non-polyadenylated RNA from nucleus isolates from GM12878 cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



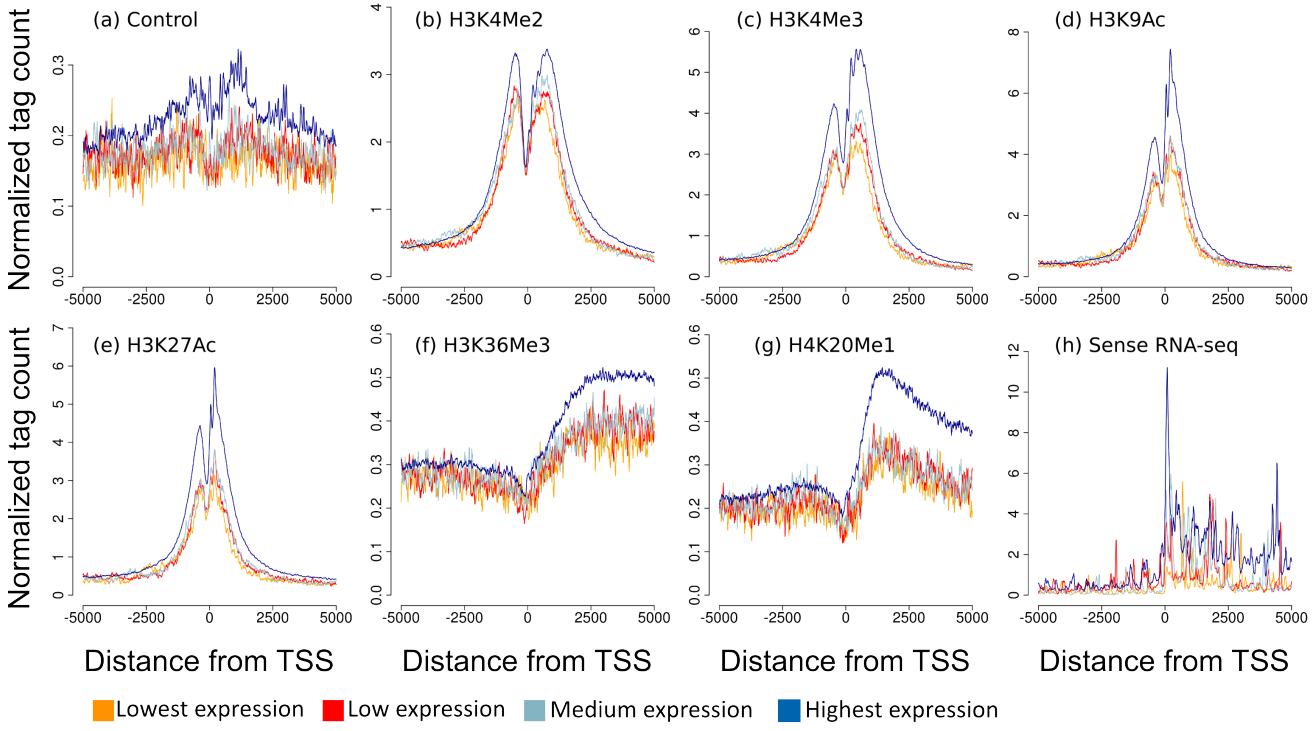
Supplementary Figure S50. Chromatin modification environment around genic promoters in H1HESC. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from non-polyadenylated RNA from whole-cell isolates from H1HESC cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



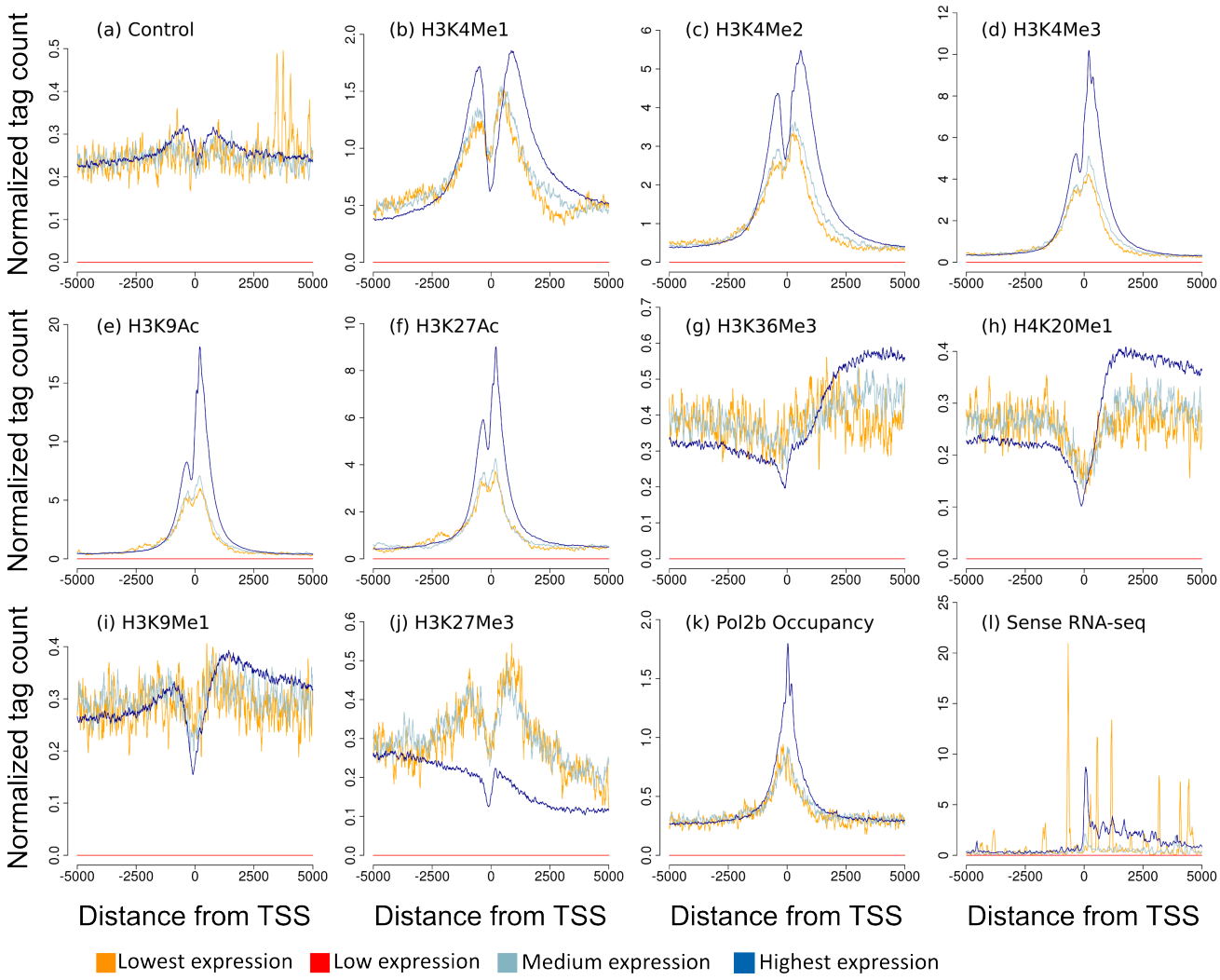
Supplementary Figure S51. Chromatin modification environment around genic promoters in HepG2. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from non-polyadenylated RNA from cytosol isolates from HepG2 cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



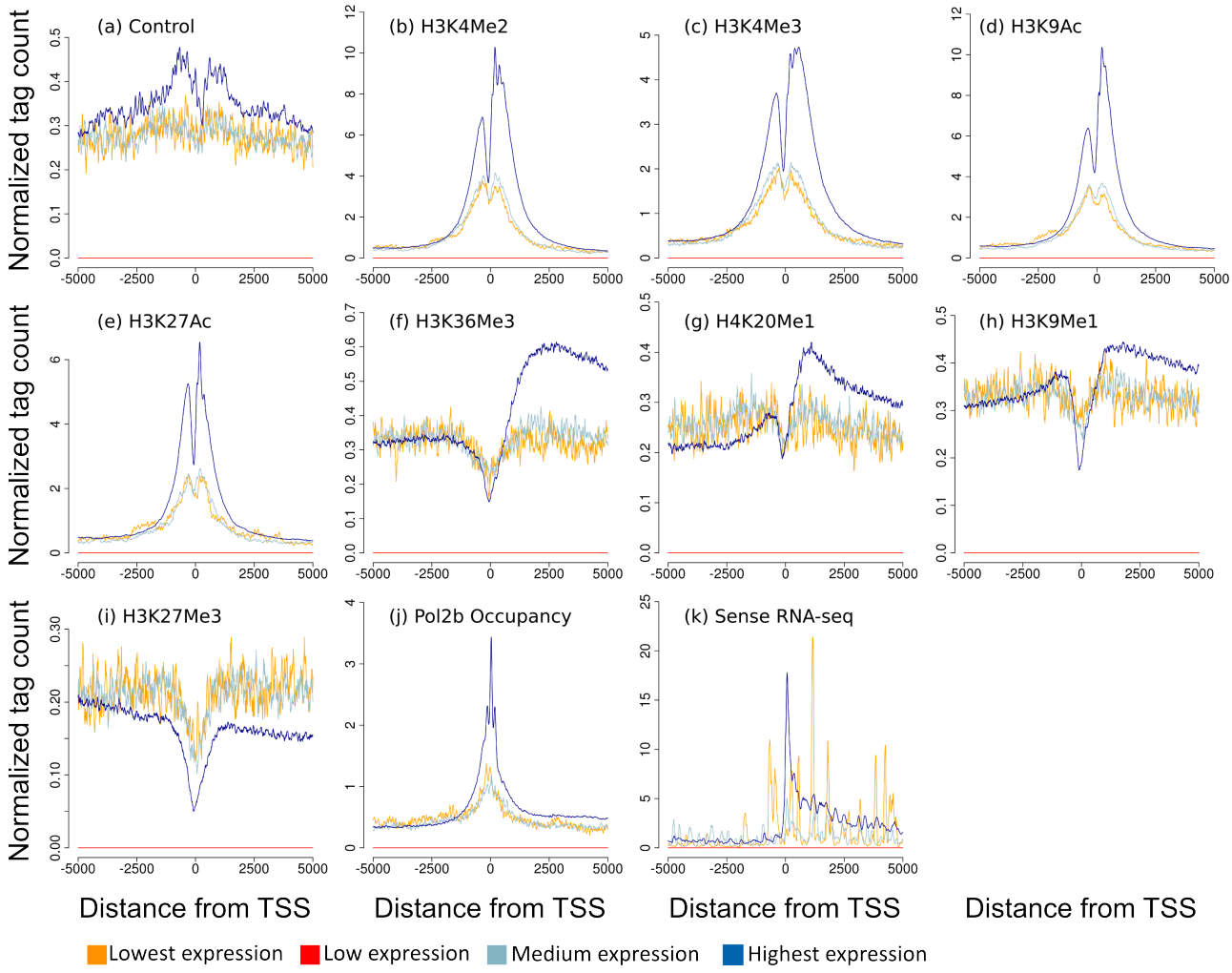
Supplementary Figure S52. Chromatin modification environment around genic promoters in HepG2. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from non-polyadenylated RNA from nucleolus isolates from HepG2 cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



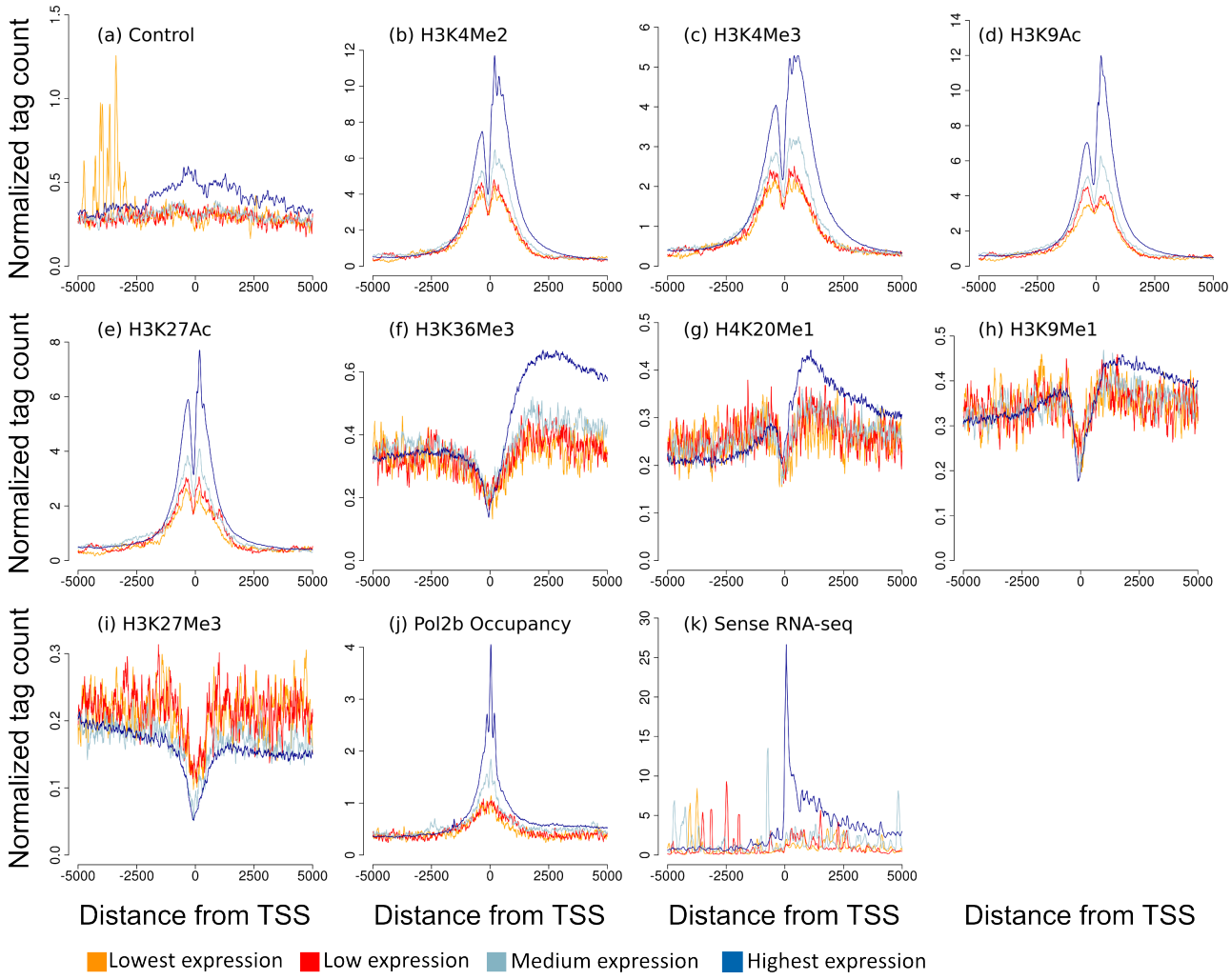
Supplementary Figure S53. Chromatin modification environment around genic promoters in HepG2. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from non-polyadenylated RNA from nucleus isolates from HepG2 cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



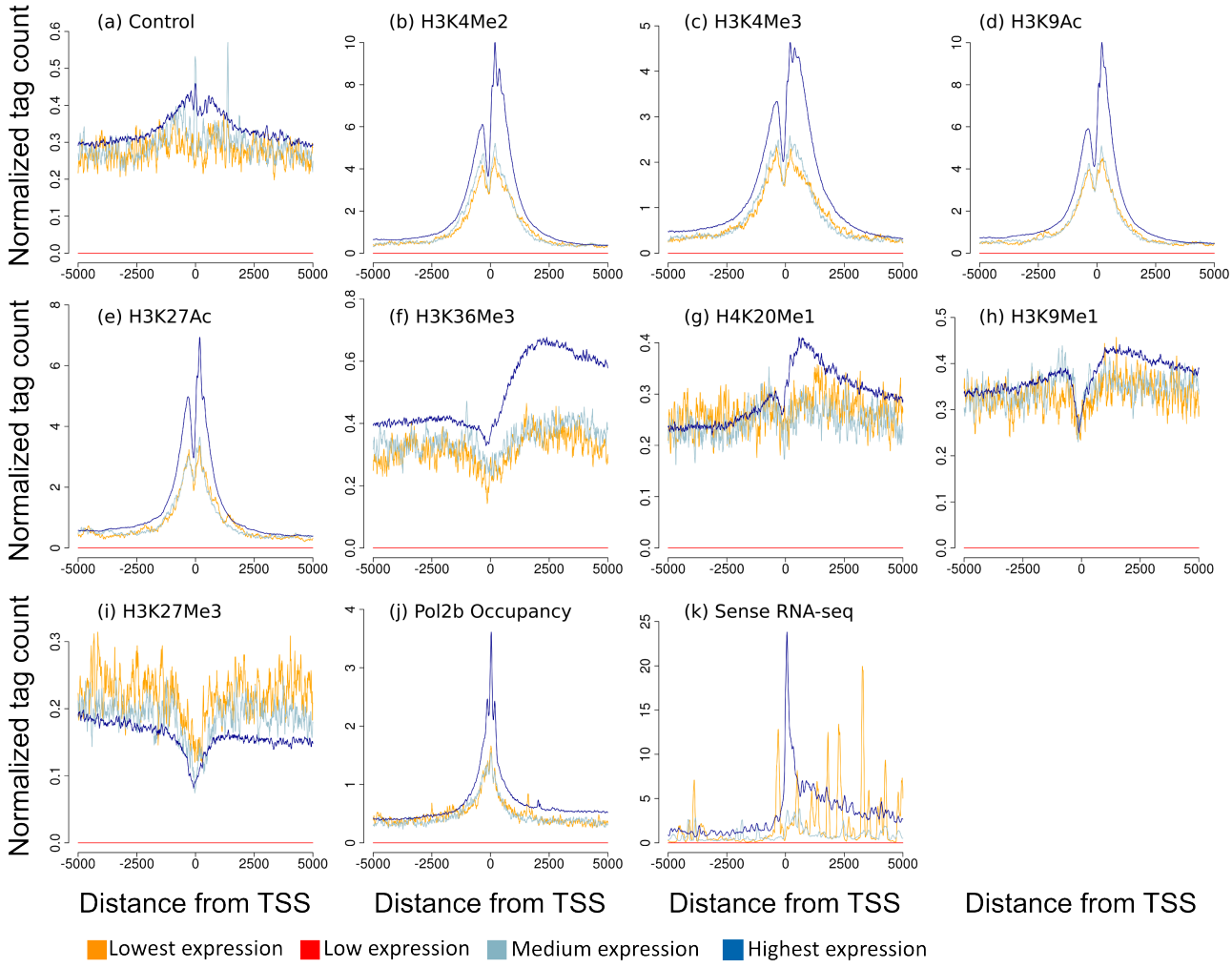
Supplementary Figure S54. Chromatin modification environment around genic promoters in HUVEC. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from non-polyadenylated RNA from cytosol isolates from HUVEC cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



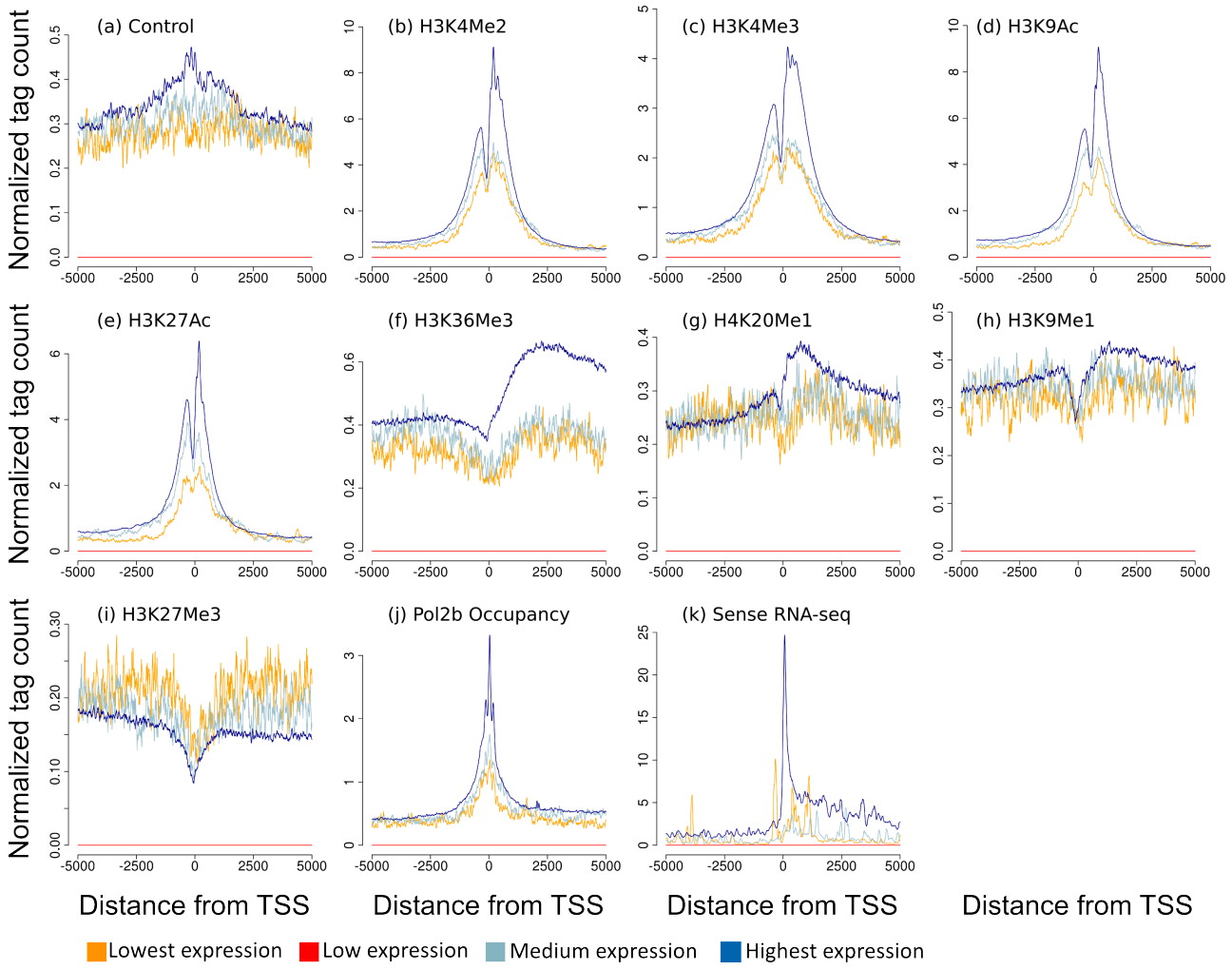
Supplementary Figure S55. Chromatin modification environment around genic promoters in HUVEC. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from non-polyadenylated RNA from cytosol isolates from HUVEC cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



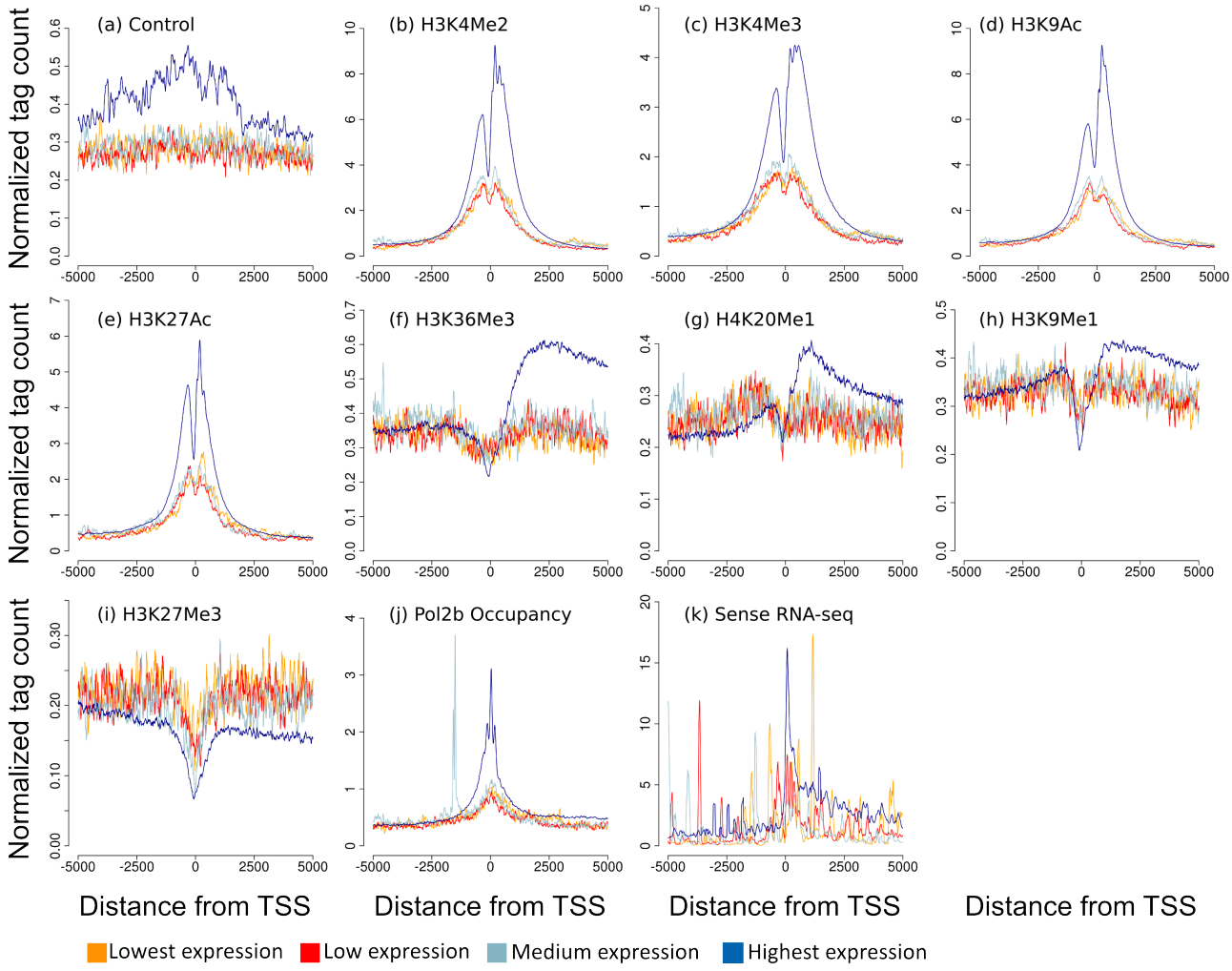
Supplementary Figure S56. Chromatin modification environment around genic promoters in K562. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from polyadenylated RNA from cytosol isolates from K562 cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



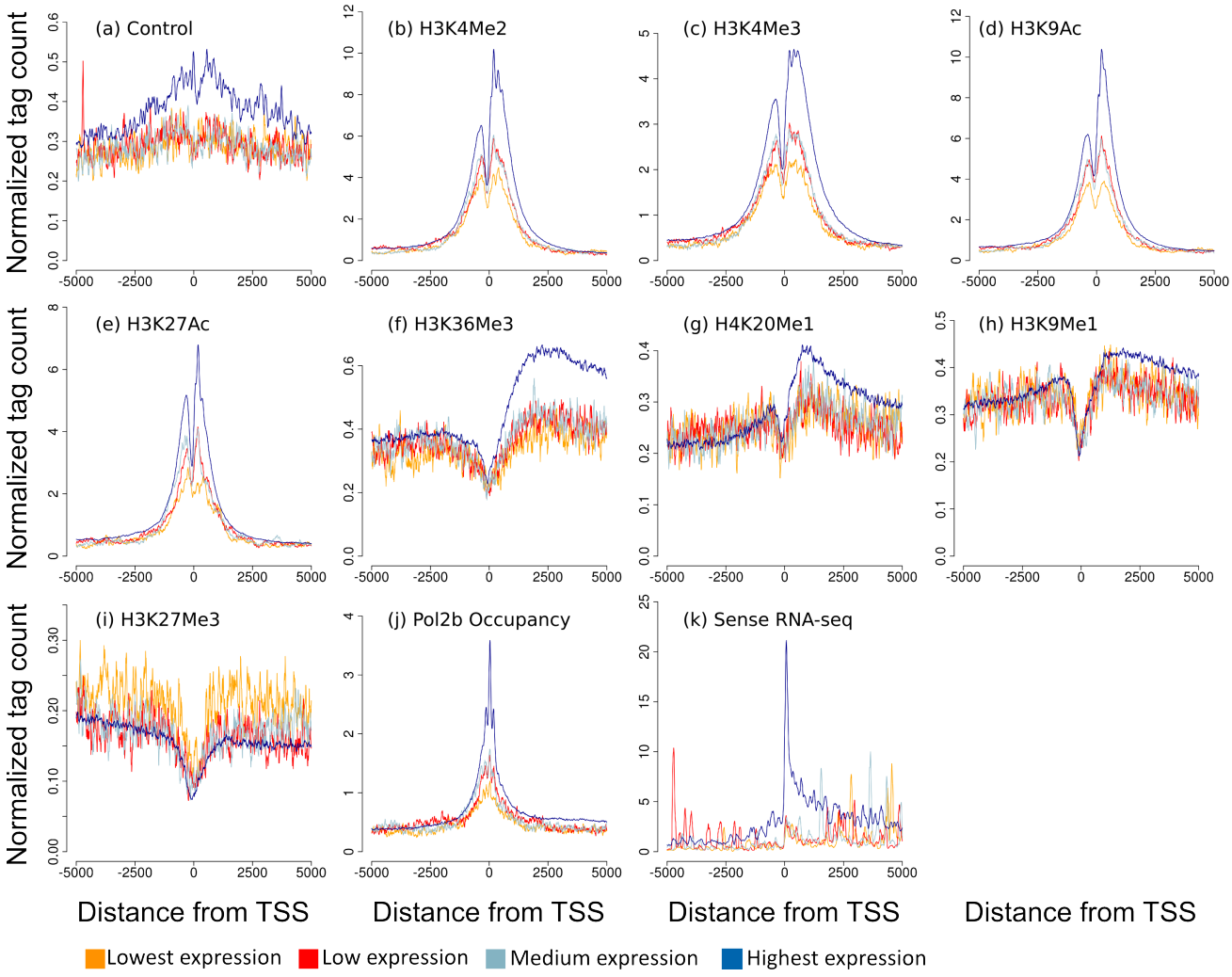
Supplementary Figure S57. Chromatin modification environment around genic promoters in K562. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from total RNA from nucleolus isolates from K562 cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



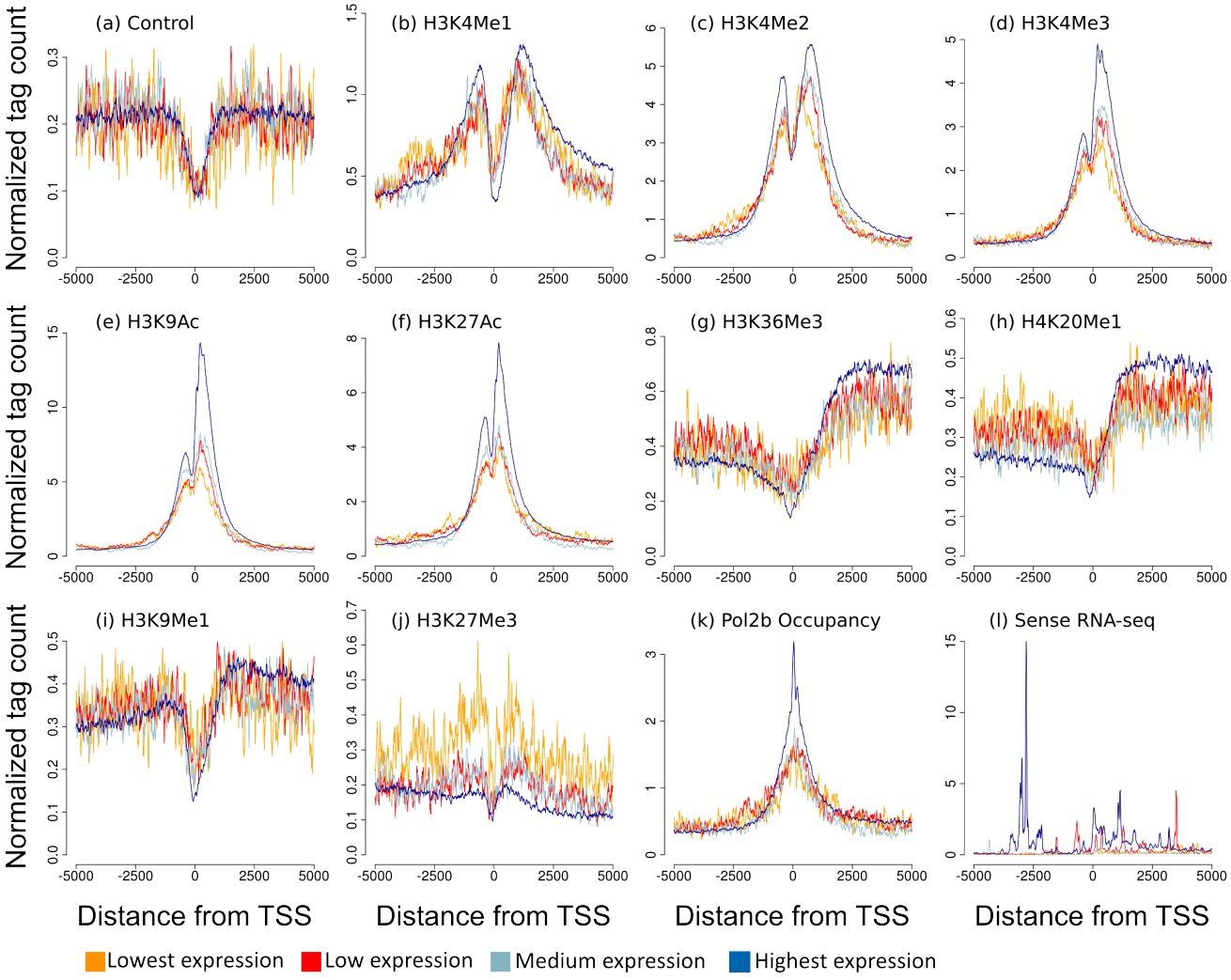
Supplementary Figure S58. Chromatin modification environment around genic promoters in K562. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from total RNA from nucleoplasm isolates from K562 cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



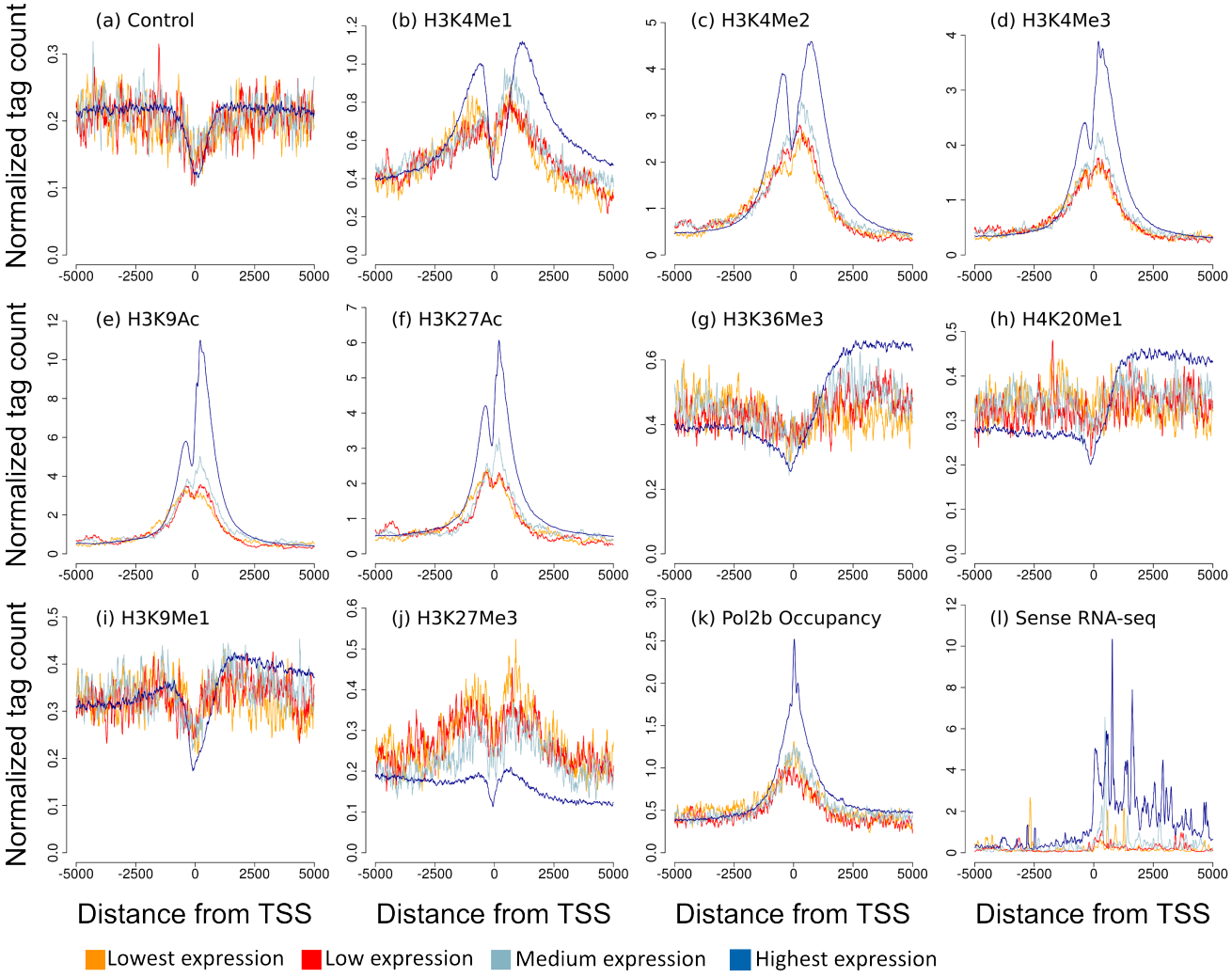
Supplementary Figure S59. Chromatin modification environment around genic promoters in K562. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from non-polyadenylated RNA from nucleus isolates from K562 cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



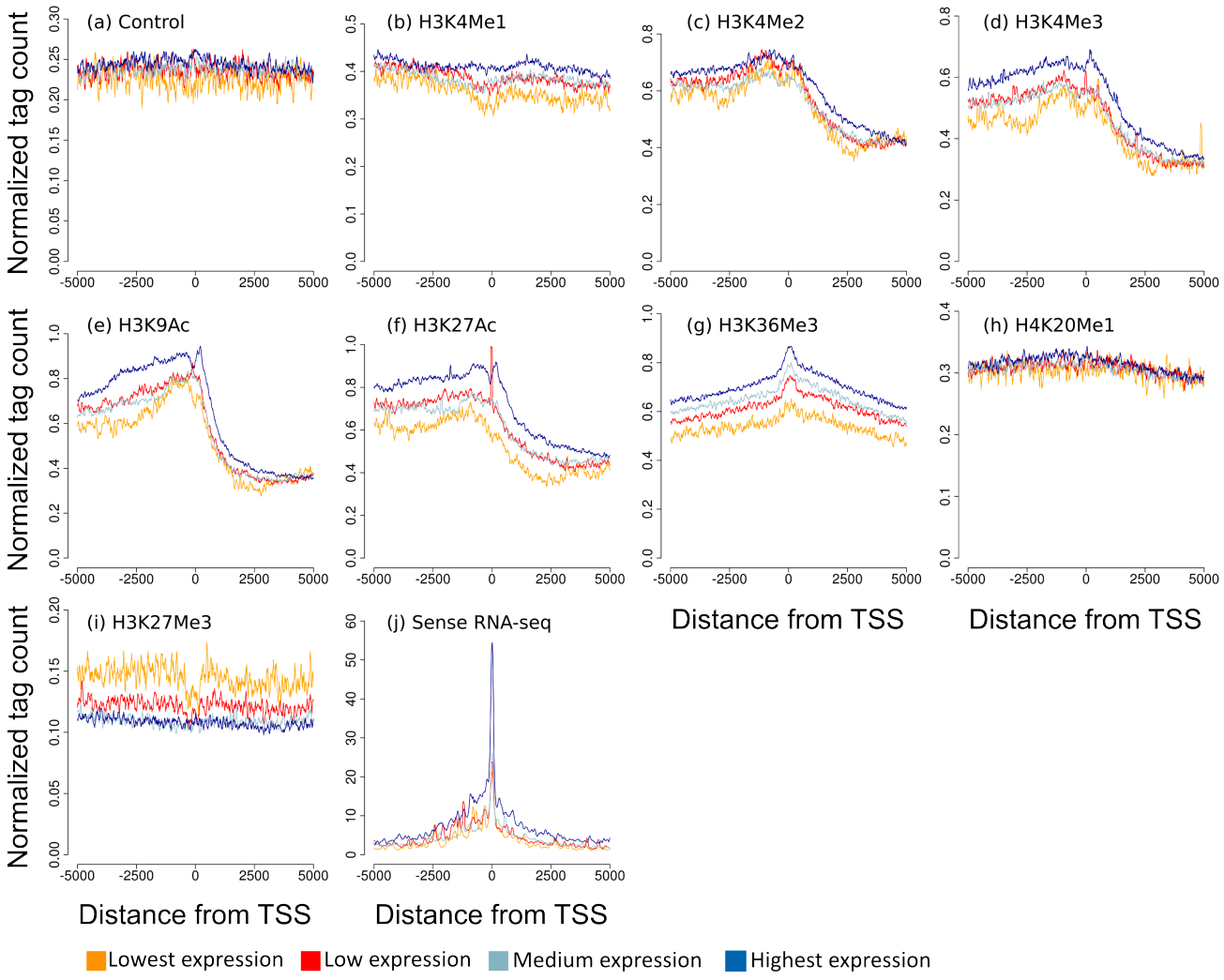
Supplementary Figure S60. Chromatin modification environment around genic promoters in K562. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from polyadenylated RNA from nucleus isolates from K562 cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



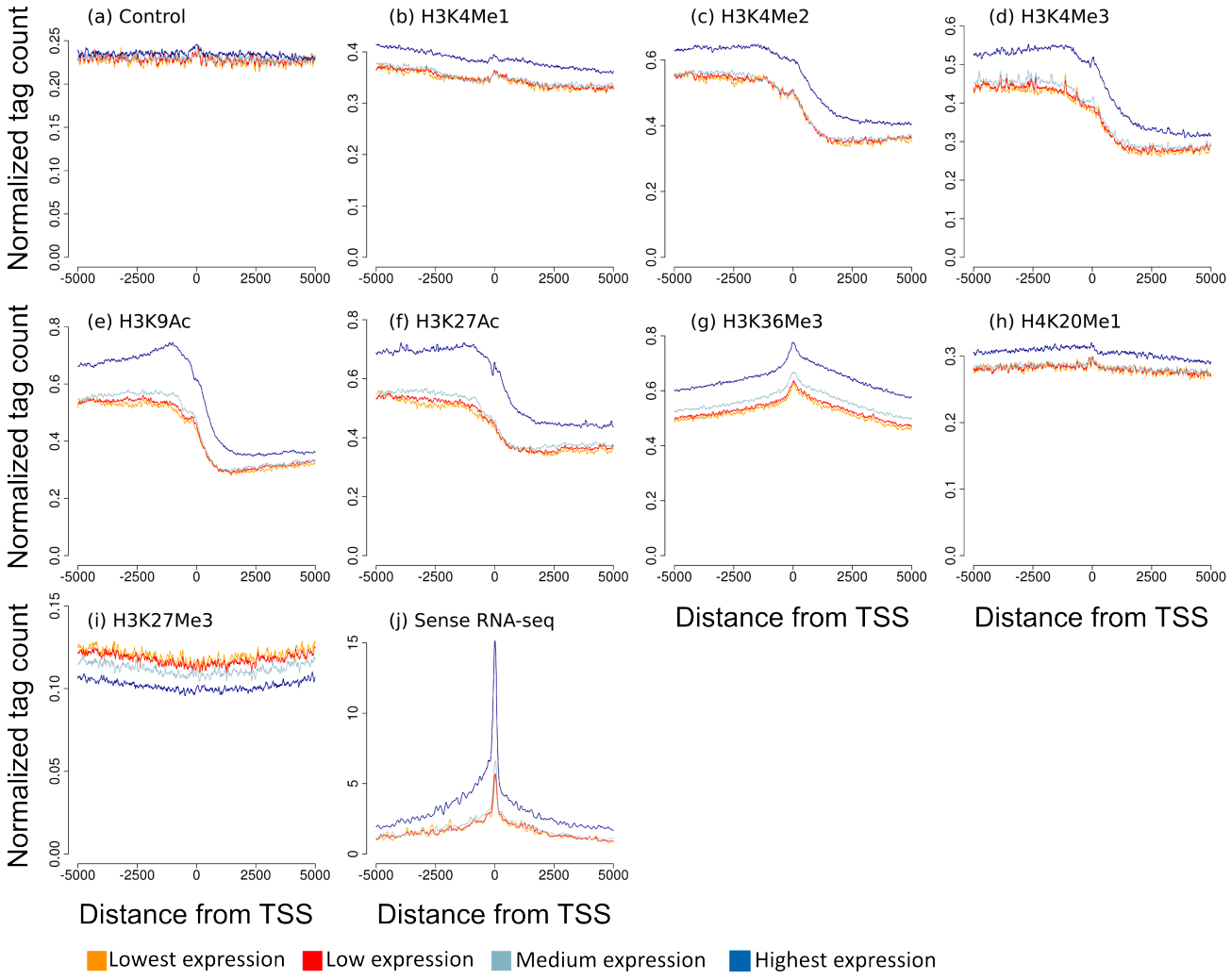
Supplementary Figure S61. Chromatin modification environment around genic promoters in NHEK. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from non-polyadenylated RNA from cytosol isolates from NHEK cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



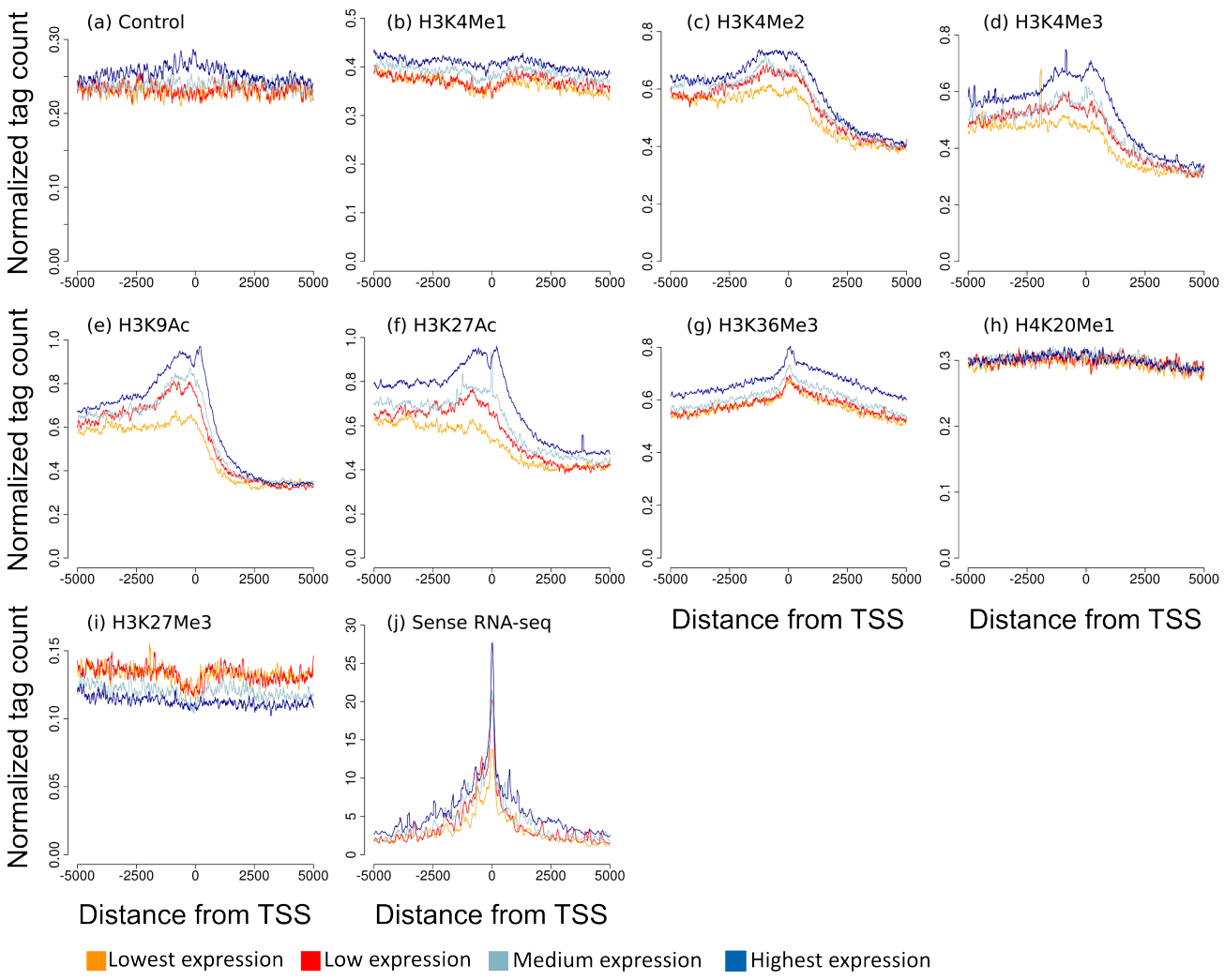
Supplementary Figure S62. Chromatin modification environment around genic promoters in NHEK. Genic promoters were taken from the UCSC genes set, and their activity measured using CAGE data from non-polyadenylated RNA from nucleus isolates from NHEK cells. Genic promoters were divided into the same for bins as cis-NAT promoters for the same data set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the genic TSS (at position 0) were calculated for each bin.



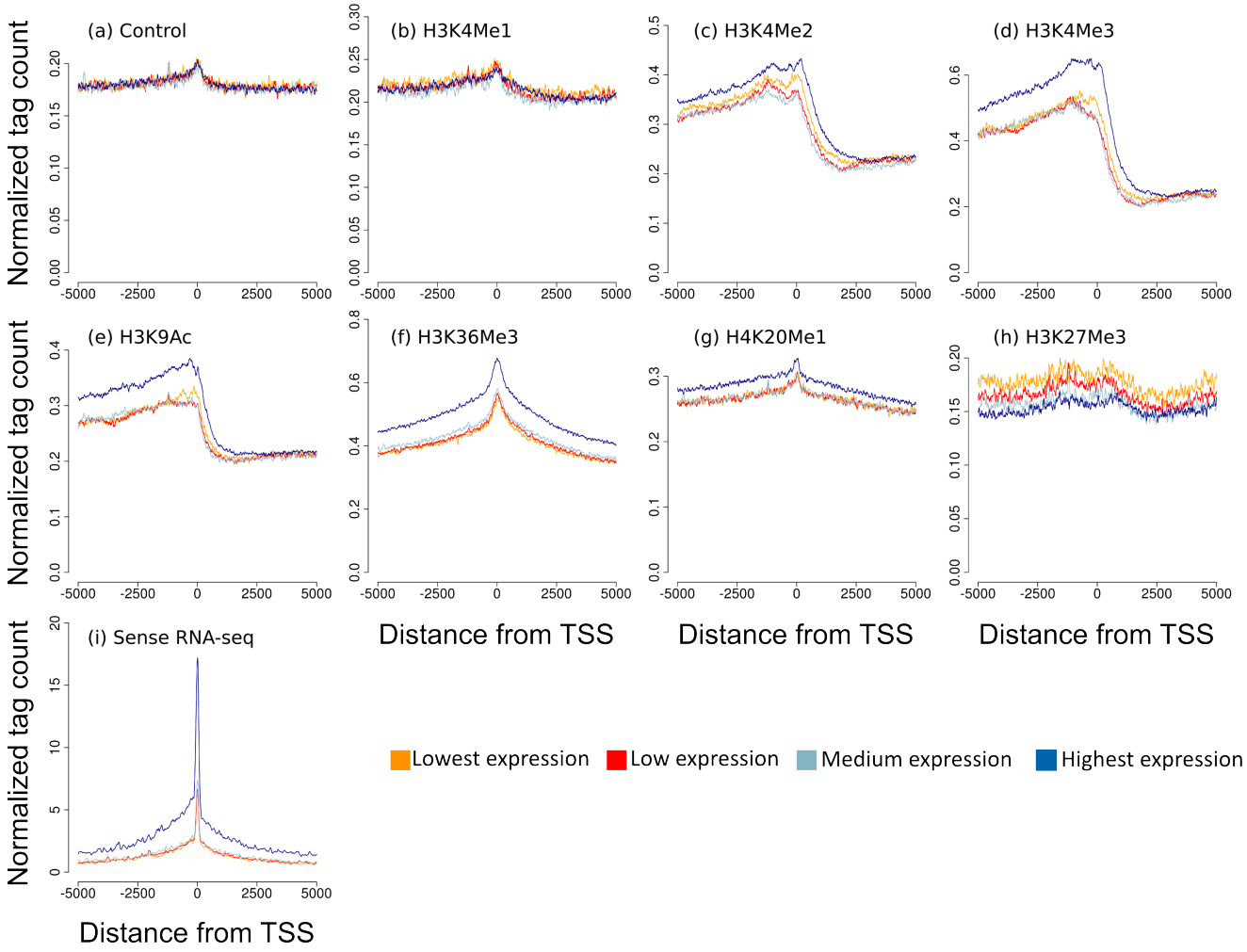
Supplementary Figure S63. Chromatin modification environment around exonic CAGE clusters in GM12878. Exonic CAGE clusters in the GM12878 cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



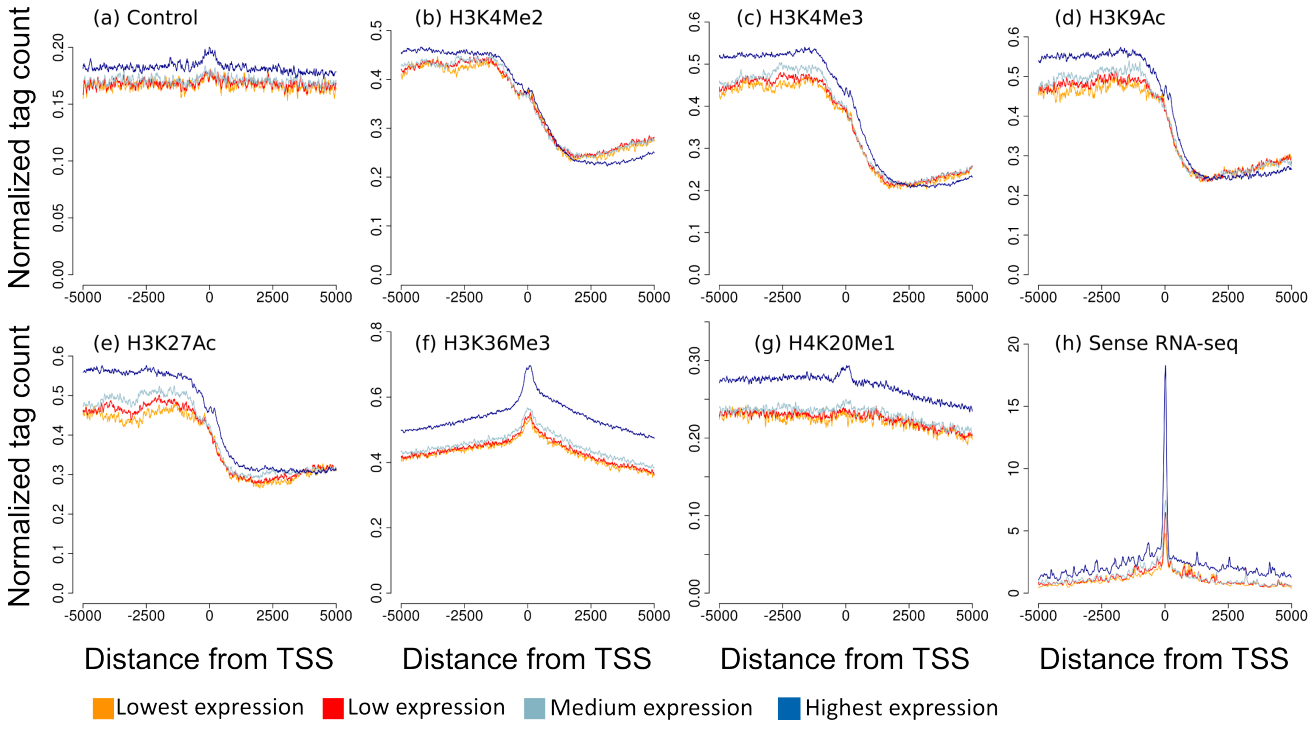
Supplementary Figure S64. Chromatin modification environment around exonic CAGE clusters in GM12878. Exonic CAGE clusters in the GM12878 cell type were identified using CAGE data from total RNA from nucleolus isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



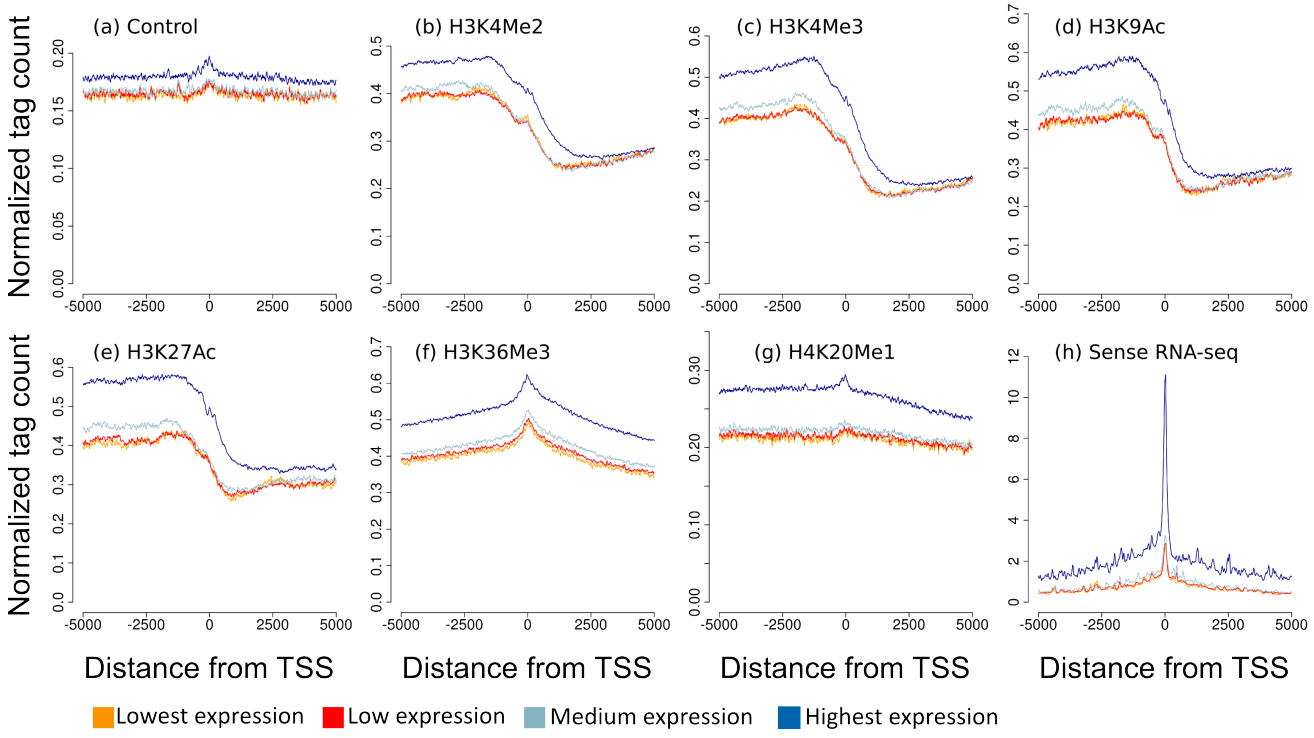
Supplementary Figure S65. Chromatin modification environment around exonic CAGE clusters in GM12878. Exonic CAGE clusters in the GM12878 cell type were identified using CAGE data from non-polyadenylated RNA from nucleus isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



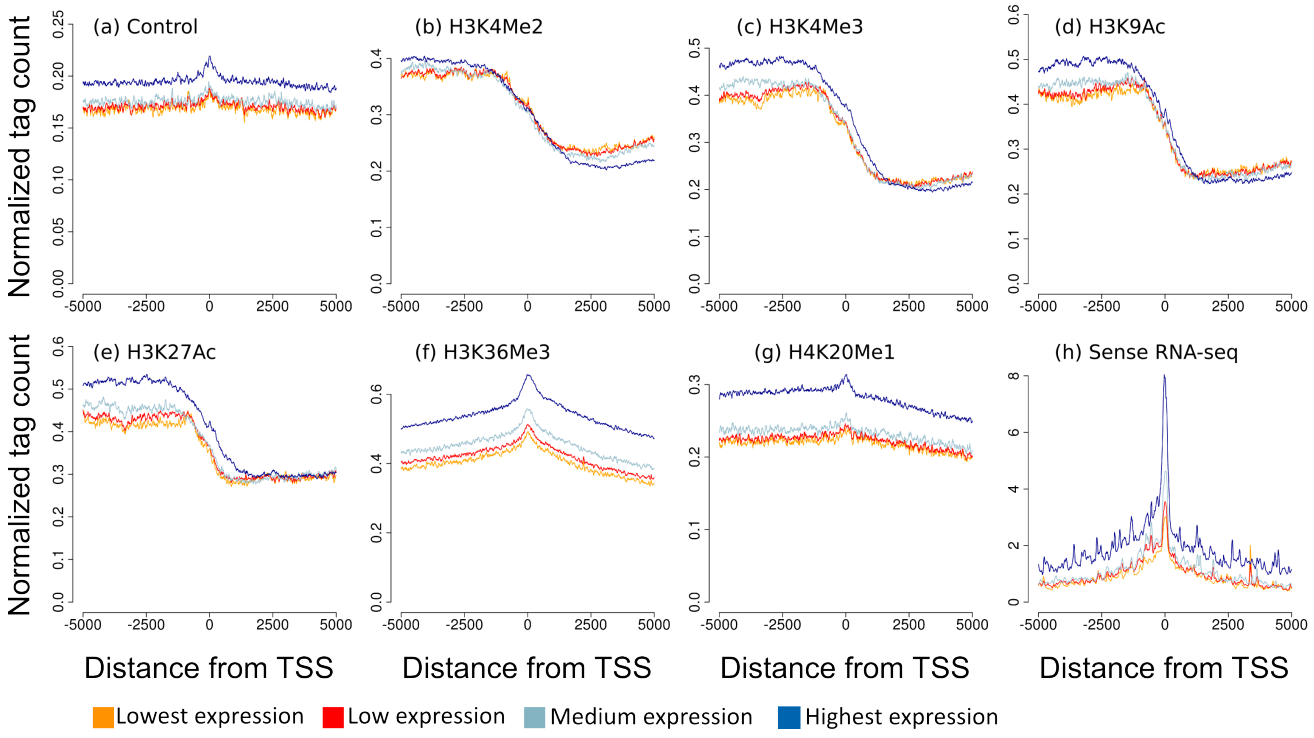
Supplementary Figure S66. Chromatin modification environment around exonic CAGE clusters in H1HESC. Exonic CAGE clusters in the H1HESC cell type were identified using CAGE data from non-polyadenylated RNA from whole-cell isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



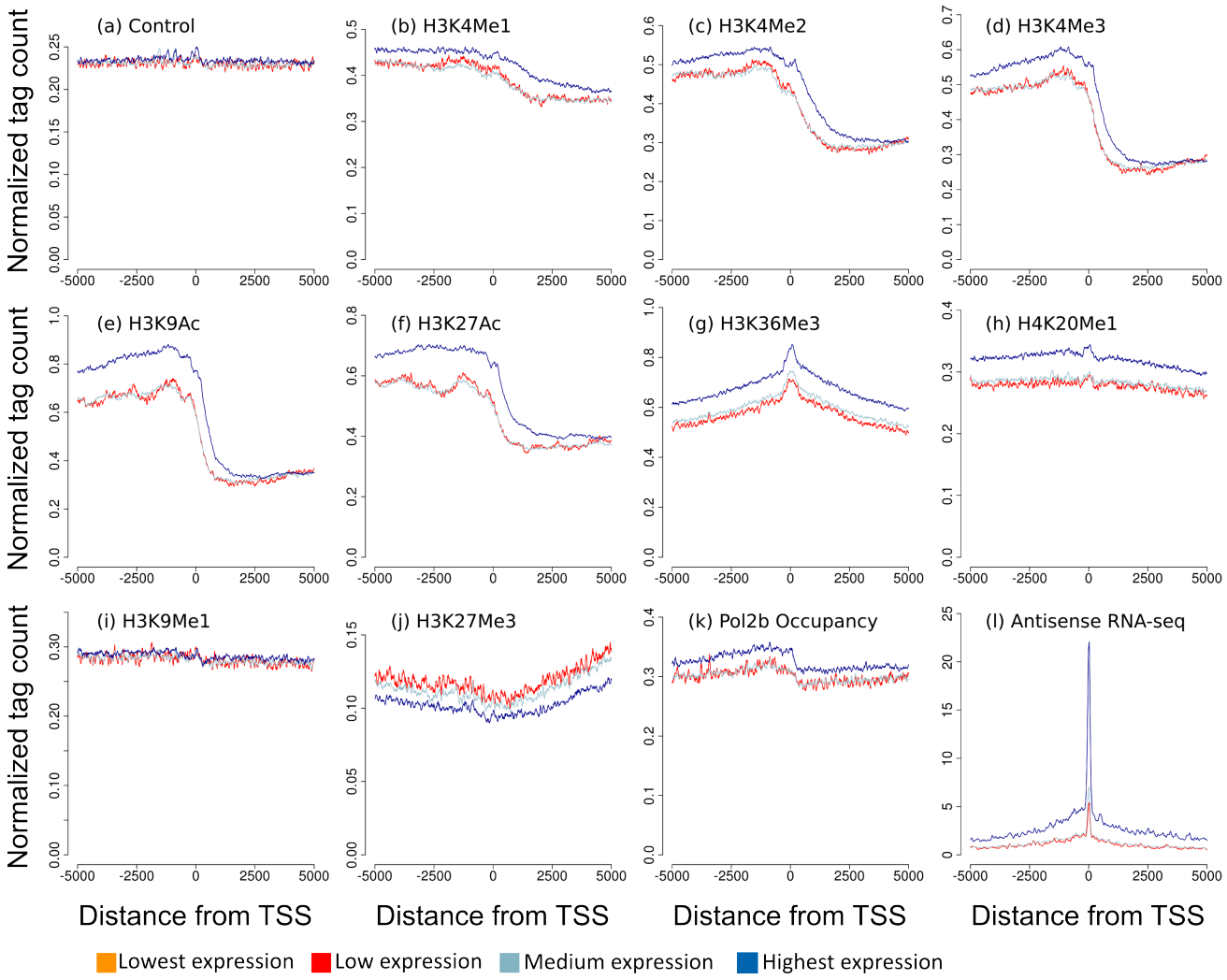
Supplementary Figure S67. Chromatin modification environment around exonic CAGE clusters in HepG2. Exonic CAGE clusters in the HepG2 cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



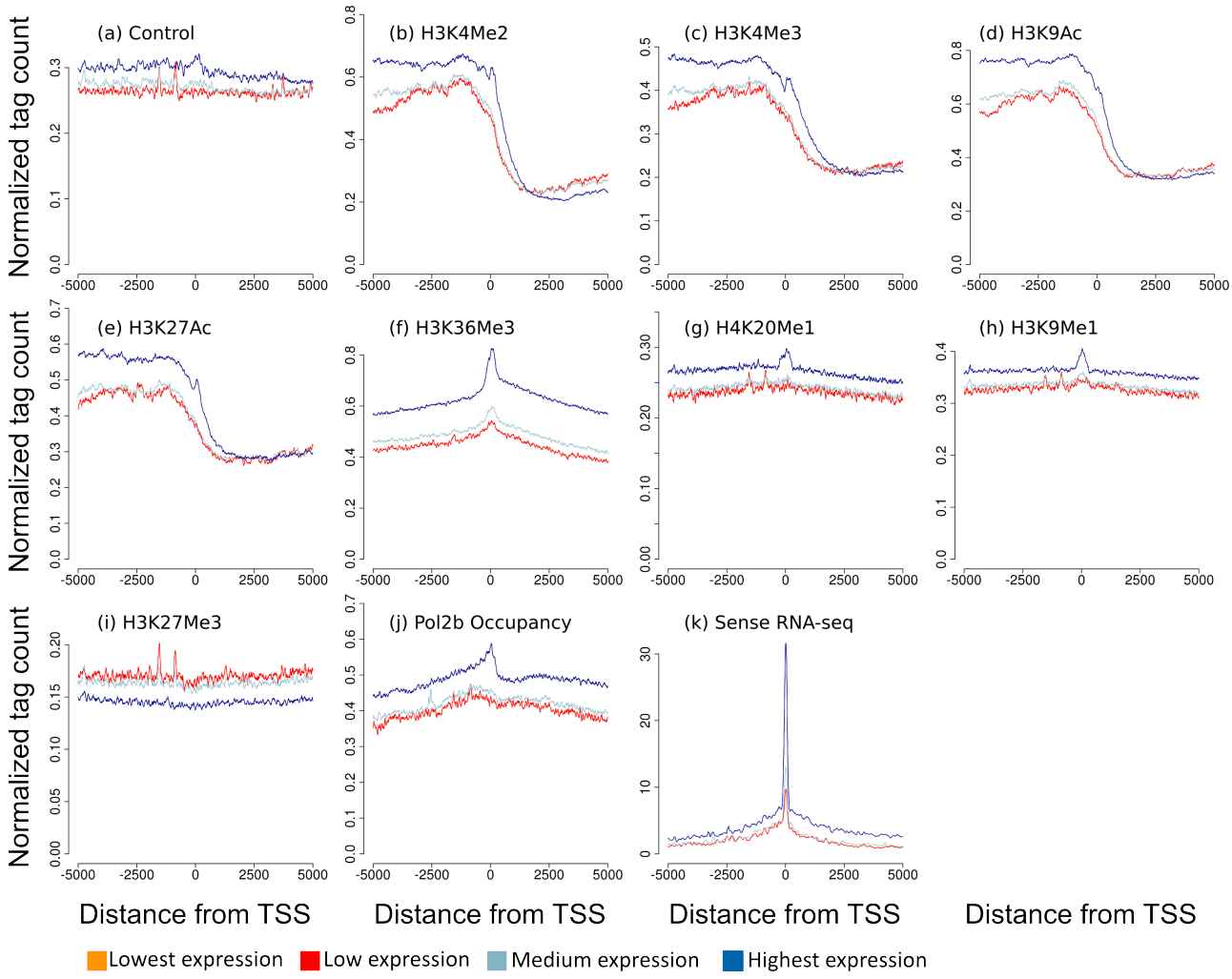
Supplementary Figure S68. Chromatin modification environment around exonic CAGE clusters in HepG2. Exonic CAGE clusters in the HepG2cell type were identified using CAGE data from total RNA from nucleolus isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



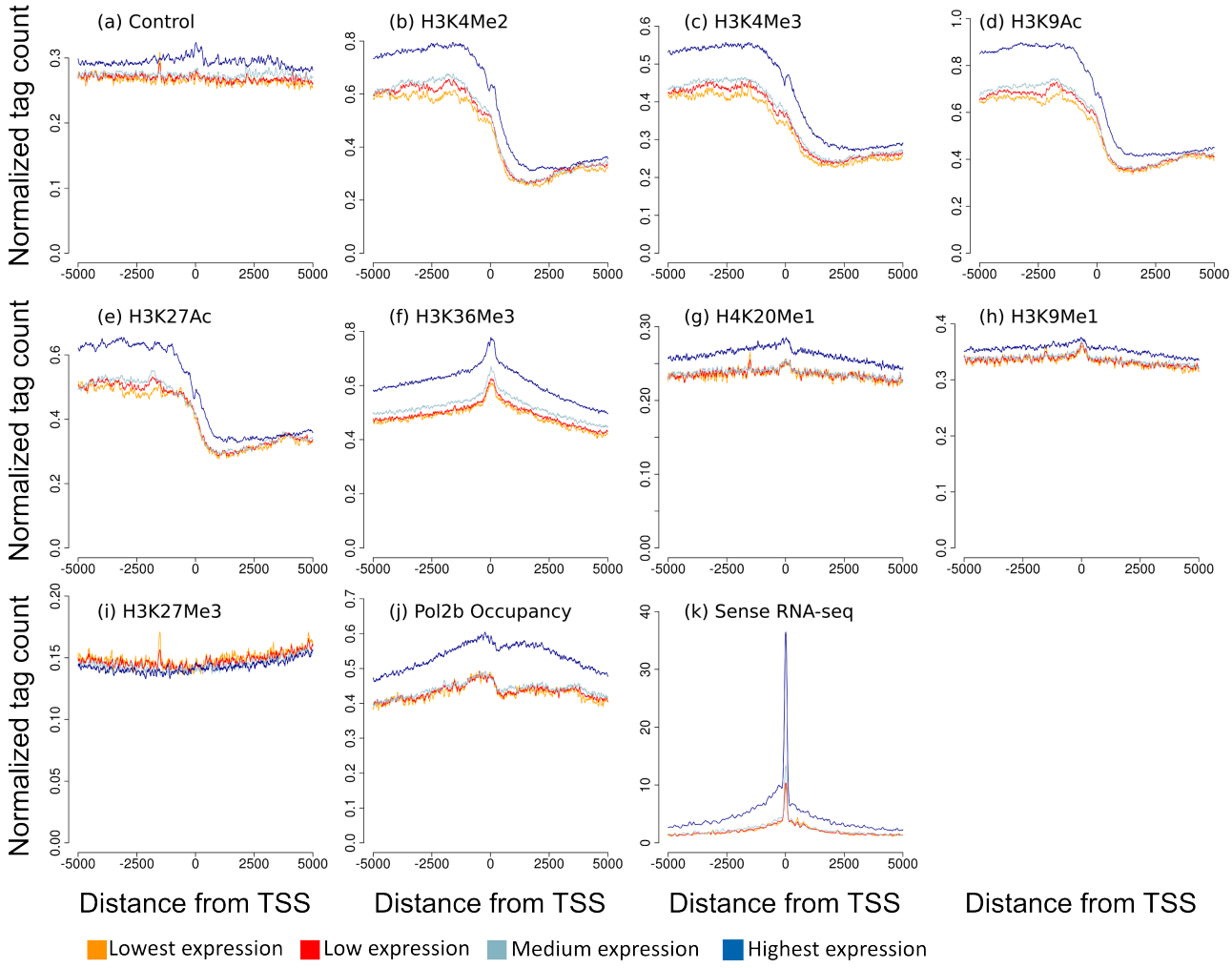
Supplementary Figure S69. Chromatin modification environment around exonic CAGE clusters in HepG2. Exonic CAGE clusters in the HepG2 cell type were identified using CAGE data from non-polyadenylated RNA from nucleus isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



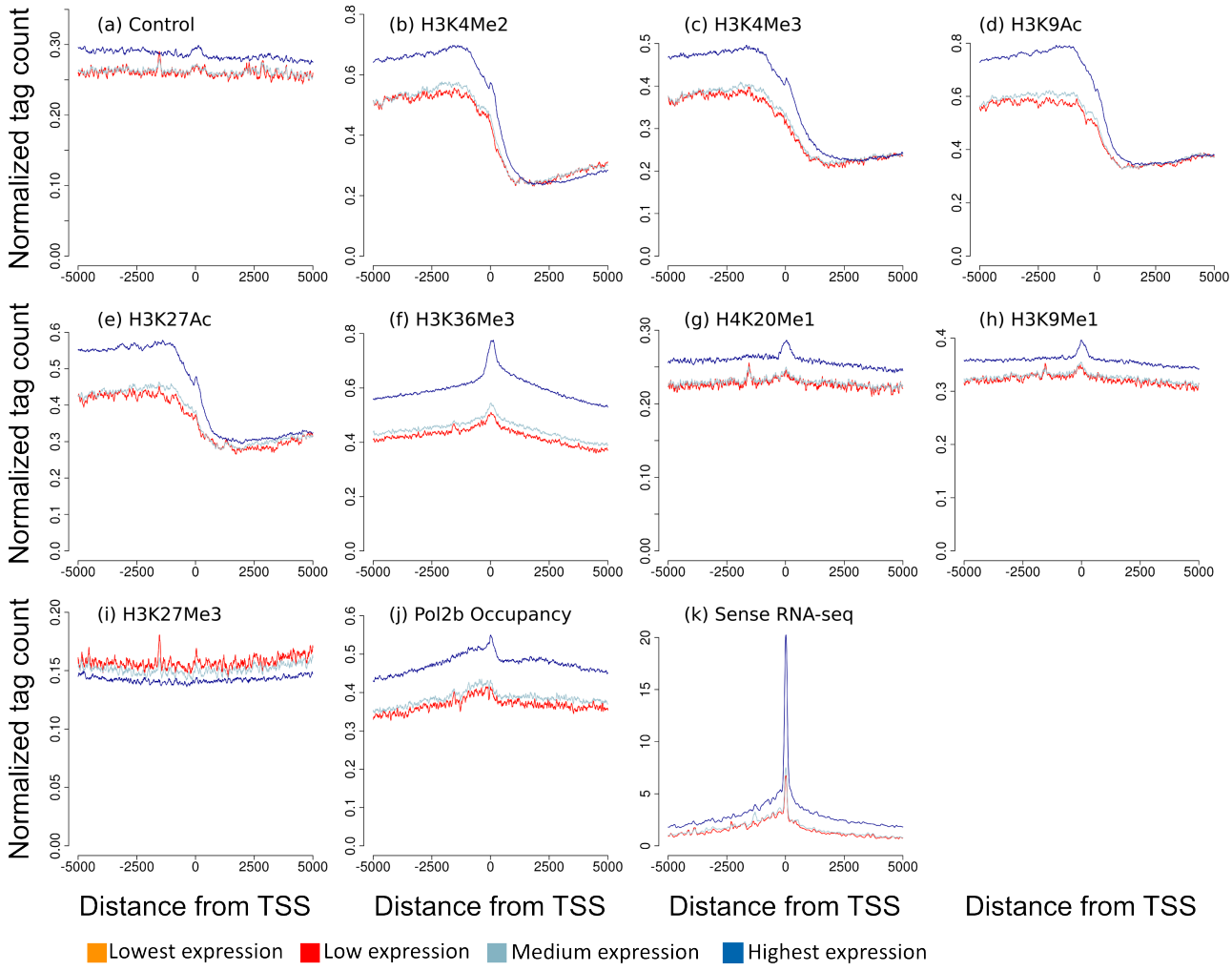
Supplementary Figure S70. Chromatin modification environment around exonic CAGE clusters in HUVEC. Exonic CAGE clusters in the HUVEC cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



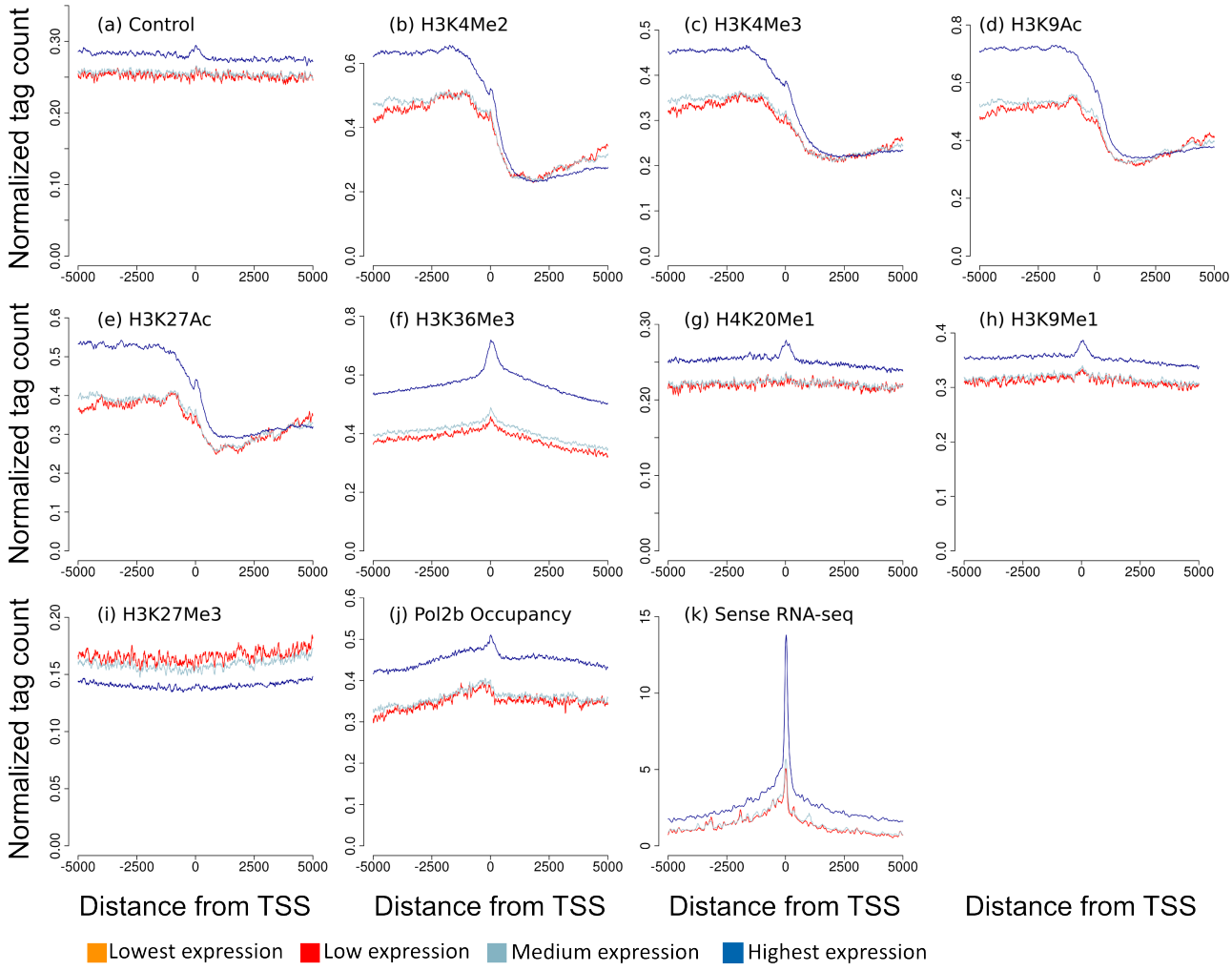
Supplementary Figure S71. Chromatin modification environment around exonic CAGE clusters in K562. Exonic CAGE clusters in the K562cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



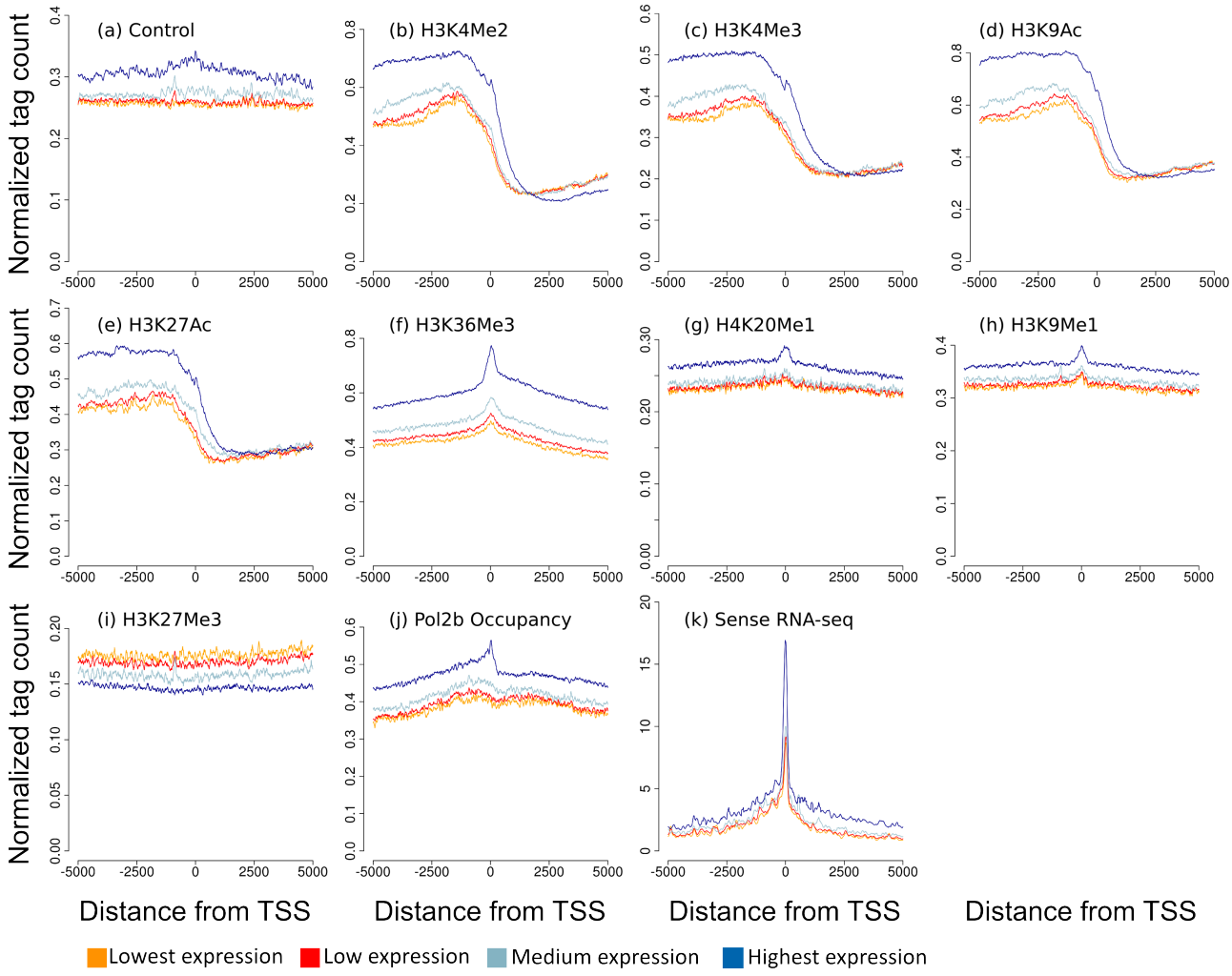
Supplementary Figure S72. Chromatin modification environment around exonic CAGE clusters in K562. Exonic CAGE clusters in the K562cell type were identified using CAGE data from polyadenylated RNA from cytosol isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



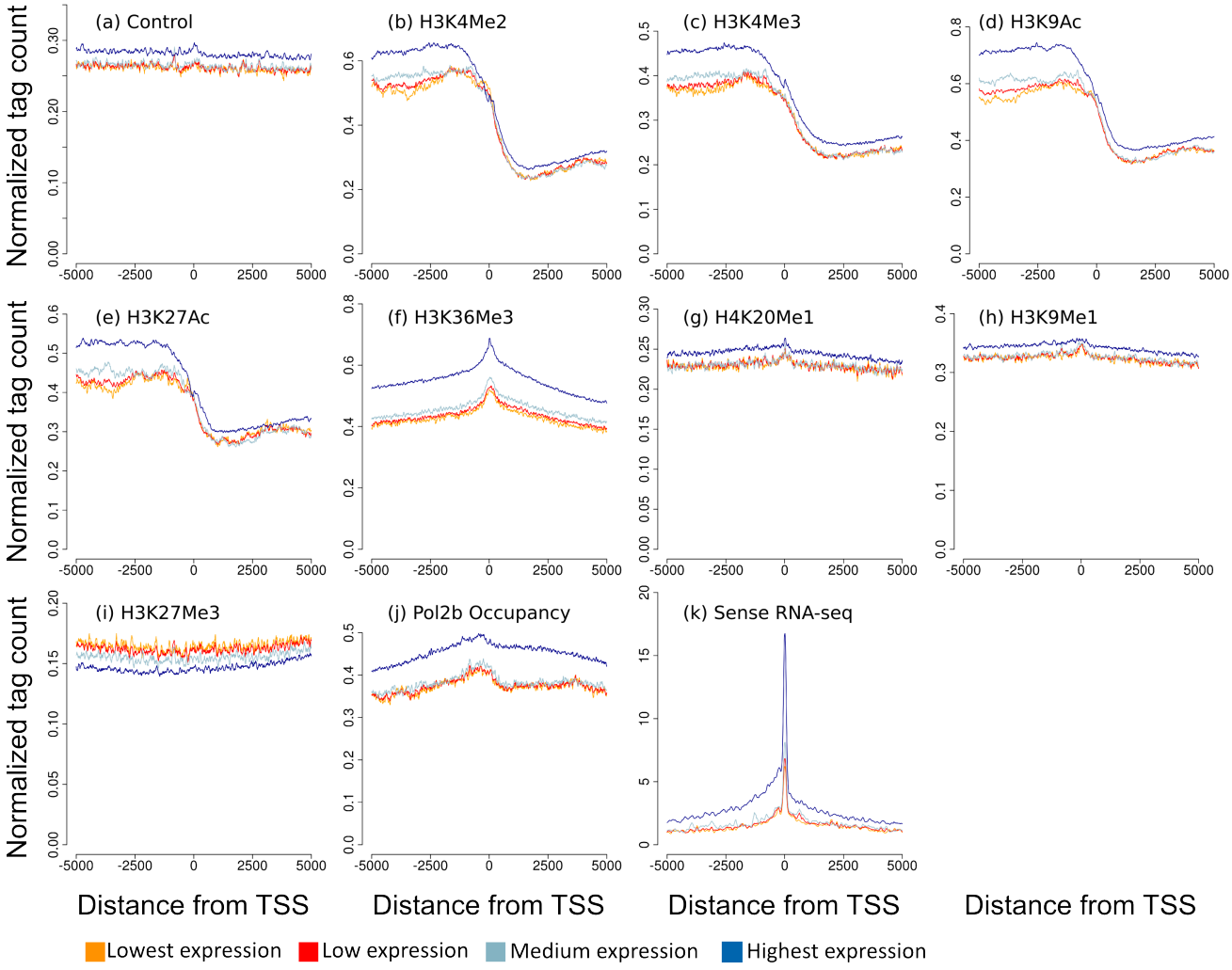
Supplementary Figure S73. Chromatin modification environment around exonic CAGE clusters in K562. Exonic CAGE clusters in the K562 cell type were identified using CAGE data from total RNA from nucleolus isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



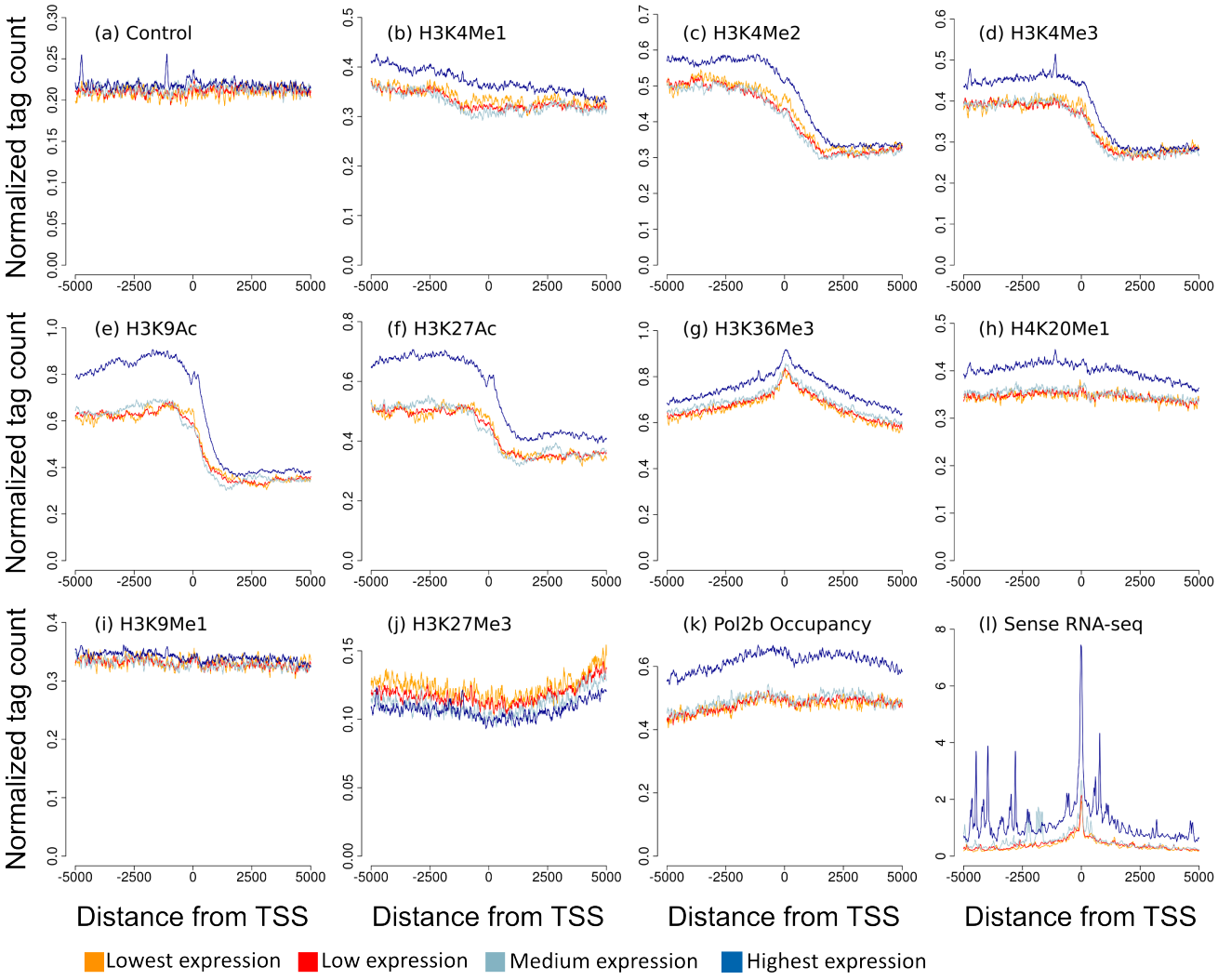
Supplementary Figure S74. Chromatin modification environment around exonic CAGE clusters in K562. Exonic CAGE clusters in the K562 cell type were identified using CAGE data from total RNA from nucleoplasm isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



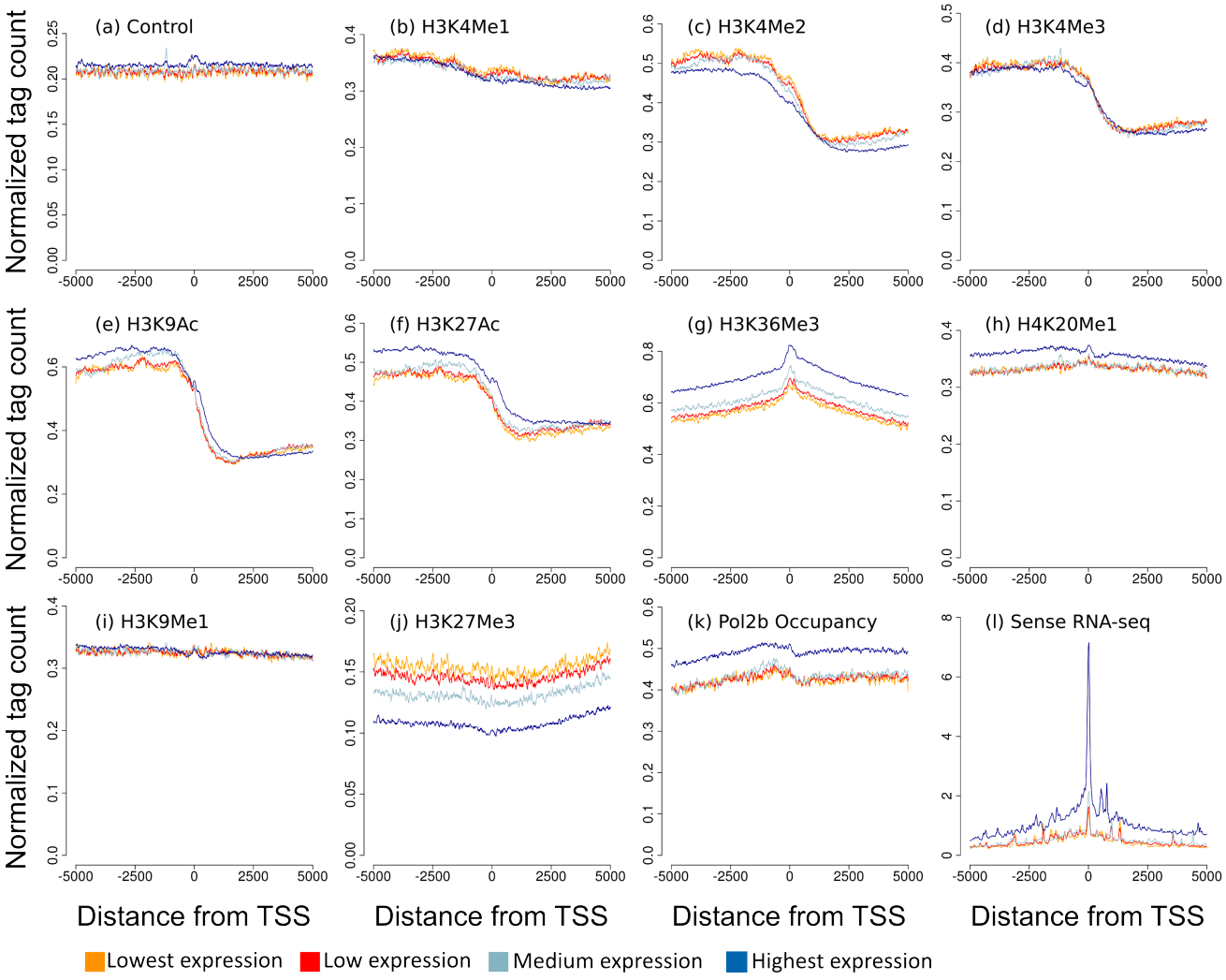
Supplementary Figure S75. Chromatin modification environment around exonic CAGE clusters in K562. Exonic CAGE clusters in the K562 cell type were identified using CAGE data from non-polyadenylated RNA from nucleus isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



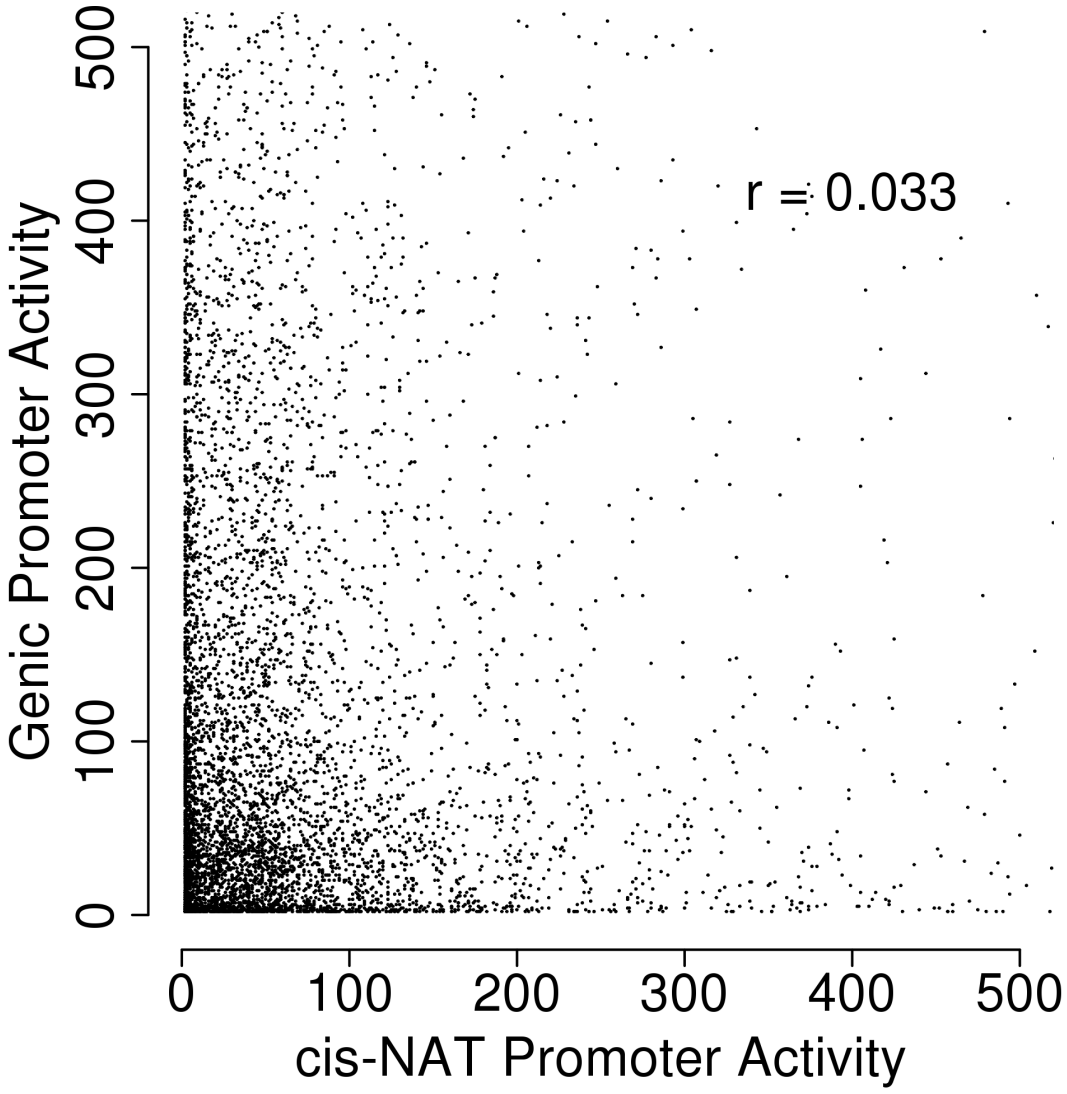
Supplementary Figure S76. Chromatin modification environment around exonic CAGE clusters in K562. Exonic CAGE clusters in the K562 cell type were identified using CAGE data from polyadenylated RNA from nucleus isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



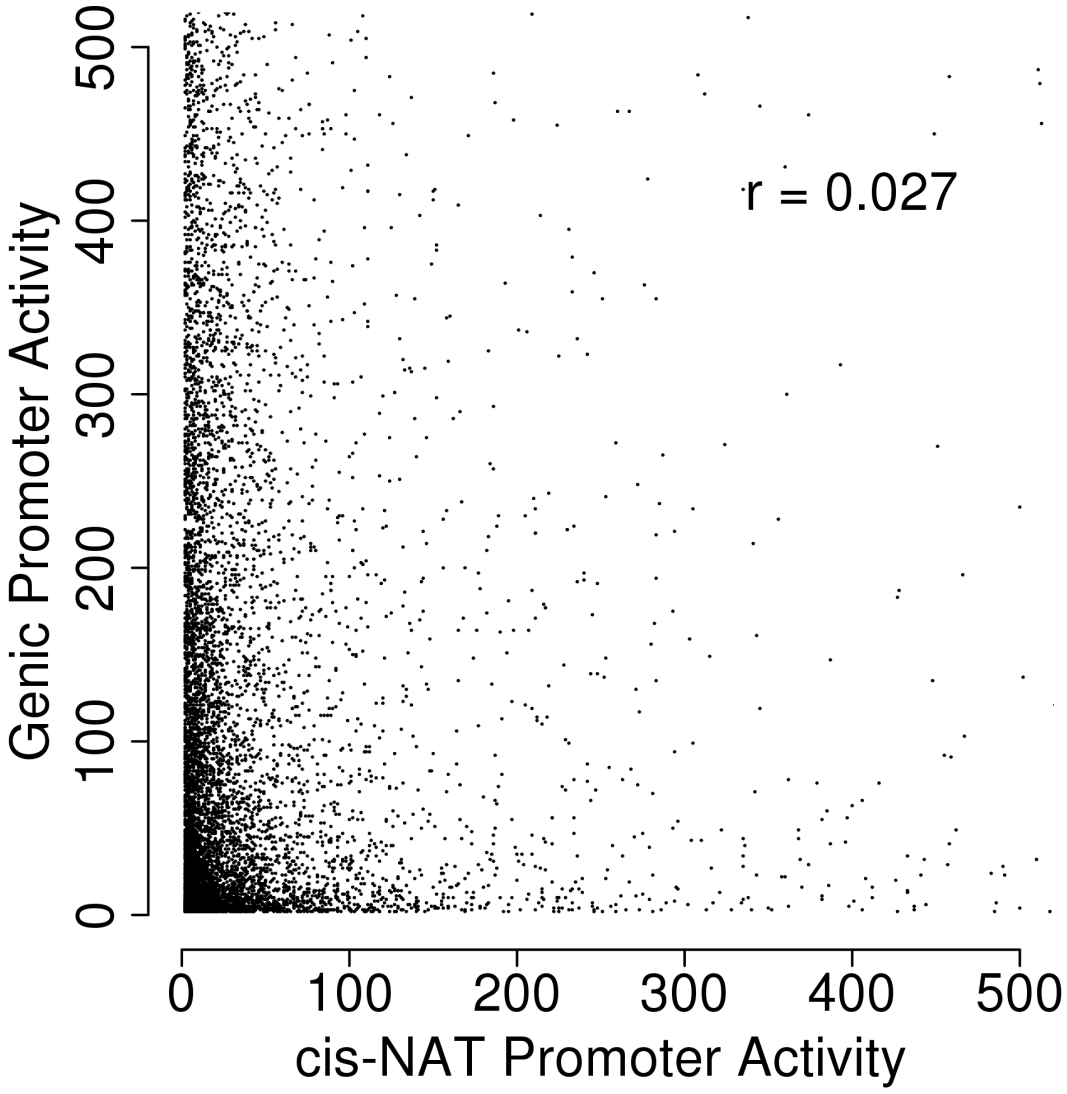
Supplementary Figure S77. Chromatin modification environment around exonic CAGE clusters in NHEK. Exonic CAGE clusters in the NHEK cell type were identified using CAGE data from non-polyadenylated RNA from cytosol isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



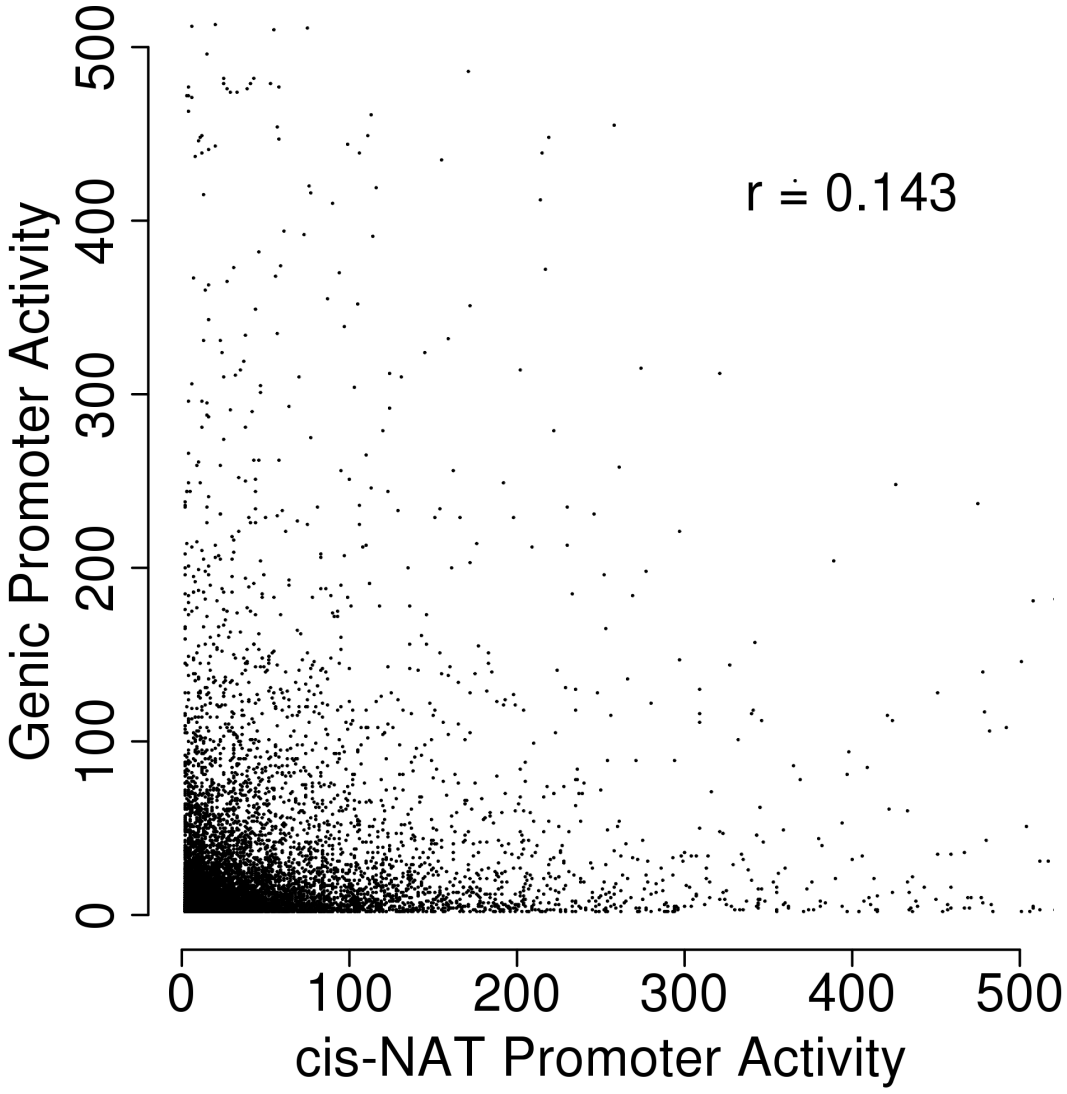
Supplementary Figure S78. Chromatin modification environment around exonic CAGE clusters in NHEK. Exonic CAGE clusters in the NHEK cell type were identified using CAGE data from non-polyadenylated RNA from nucleus isolates. Exonic CAGE clusters were divided into the same four bins as cis-NAT promoters from the CAGE same set, and the normalized average numbers of ChIP-seq reads in 10 base-pair windows +/- 5kb of the exonic CAGE cluster peak (at position 0) were calculated for each bin.



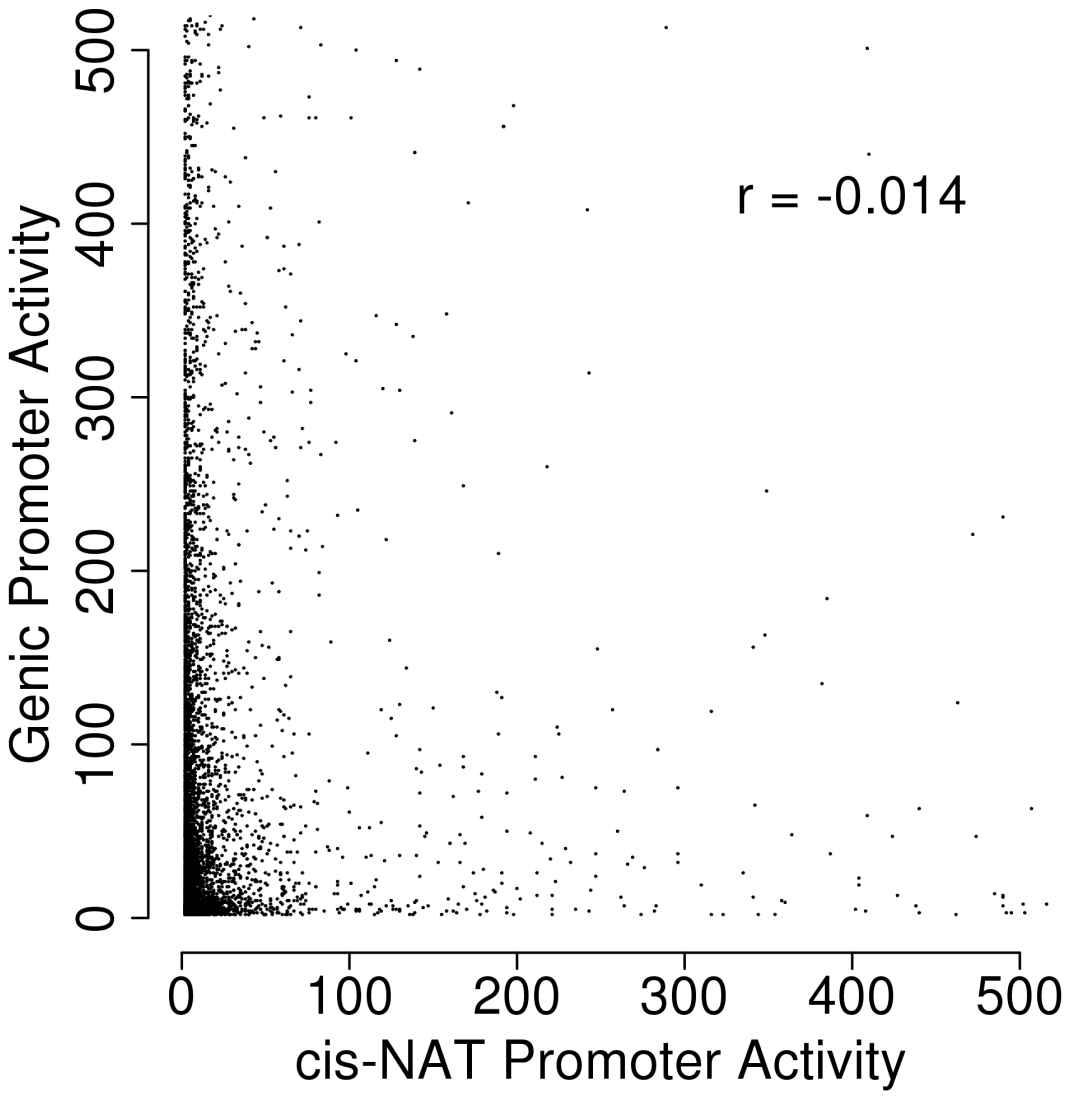
Supplementary Figure S79. Correlation between genic and cis-NAT promoter activity in GM12878. Cis-NAT promoters in the GM12878 cell type were identified using CAGE data from non-polyadenylated RNA form nucleus isolates. Activity of genic promoters and the sum of corresponding cis-NAT promoter activity was measured by CAGE tag counts. A Spearman rank correlation was used to determine the relationship between total cis-NAT promoter activity and genic promoter activity.



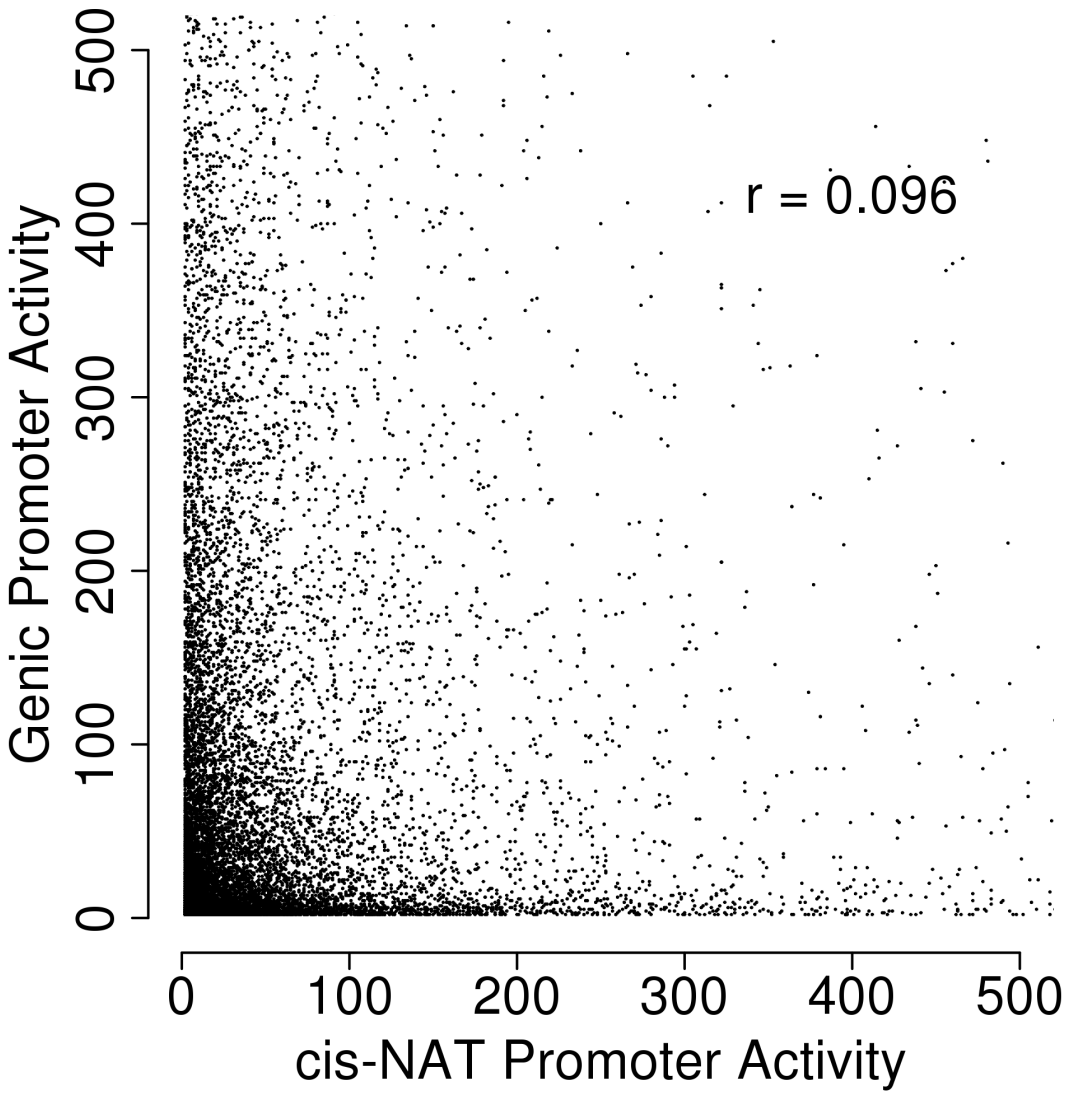
Supplementary Figure S80. Correlation between genic and cis-NAT promoter activity in H1HESC. Cis-NAT promoters in the H1HESC cell type were identified using CAGE data from non-polyadenylated RNA form whole-cell isolates. Activity of genic promoters and the sum of corresponding cis-NAT promoter activity was measured by CAGE tag counts. A Spearman rank correlation was used to determine the relationship between total cis-NAT promoter activity and genic promoter activity.



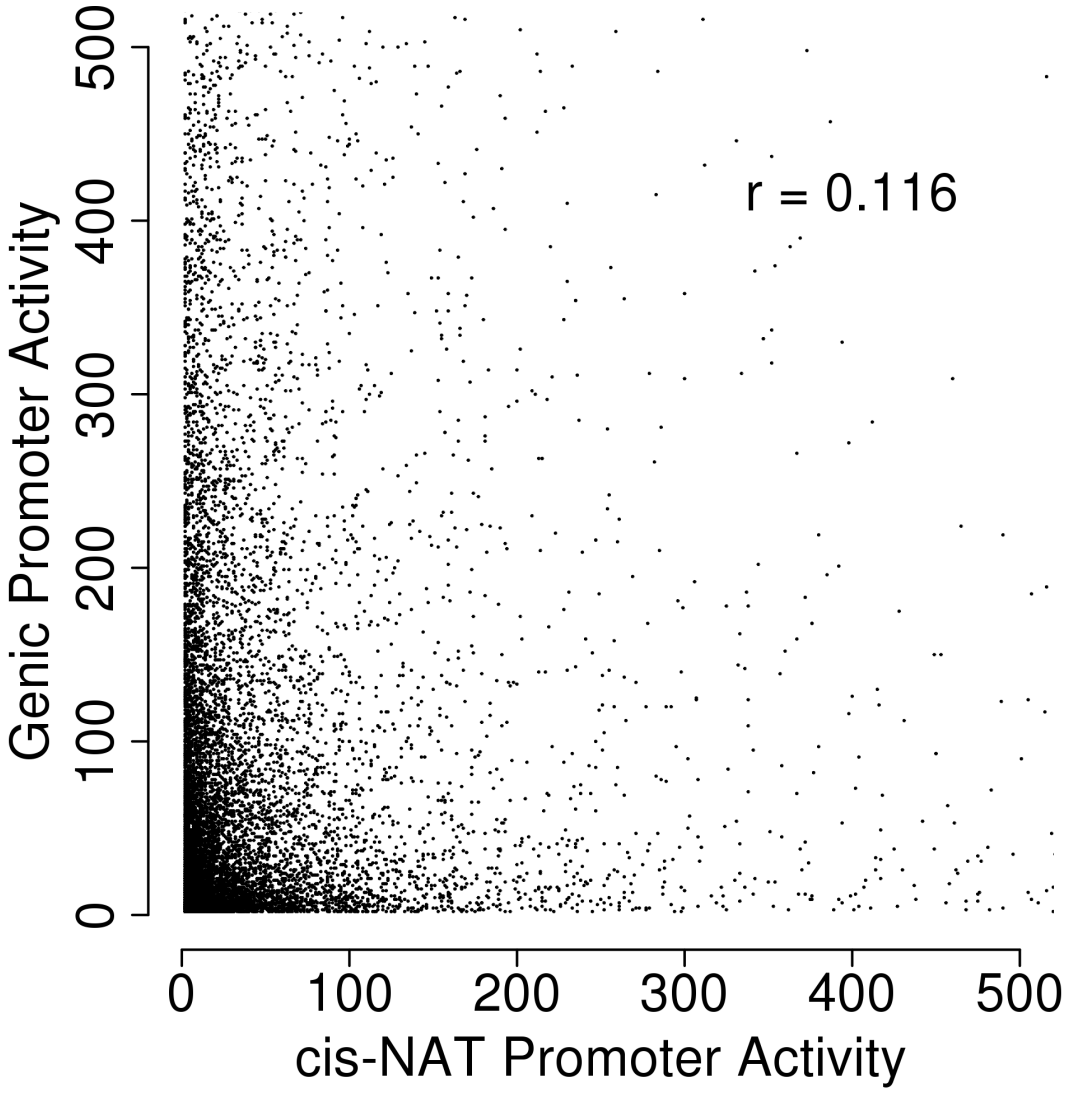
Supplementary Figure S81. Correlation between genic and cis-NAT promoter activity in HepG2. Cis-NAT promoters in the HepG2 cell type were identified using CAGE data from non-polyadenylated RNA form nucleus isolates. Activity of genic promoters and the sum of corresponding cis-NAT promoter activity was measured by CAGE tag counts. A Spearman rank correlation was used to determine the relationship between total cis-NAT promoter activity and genic promoter activity.



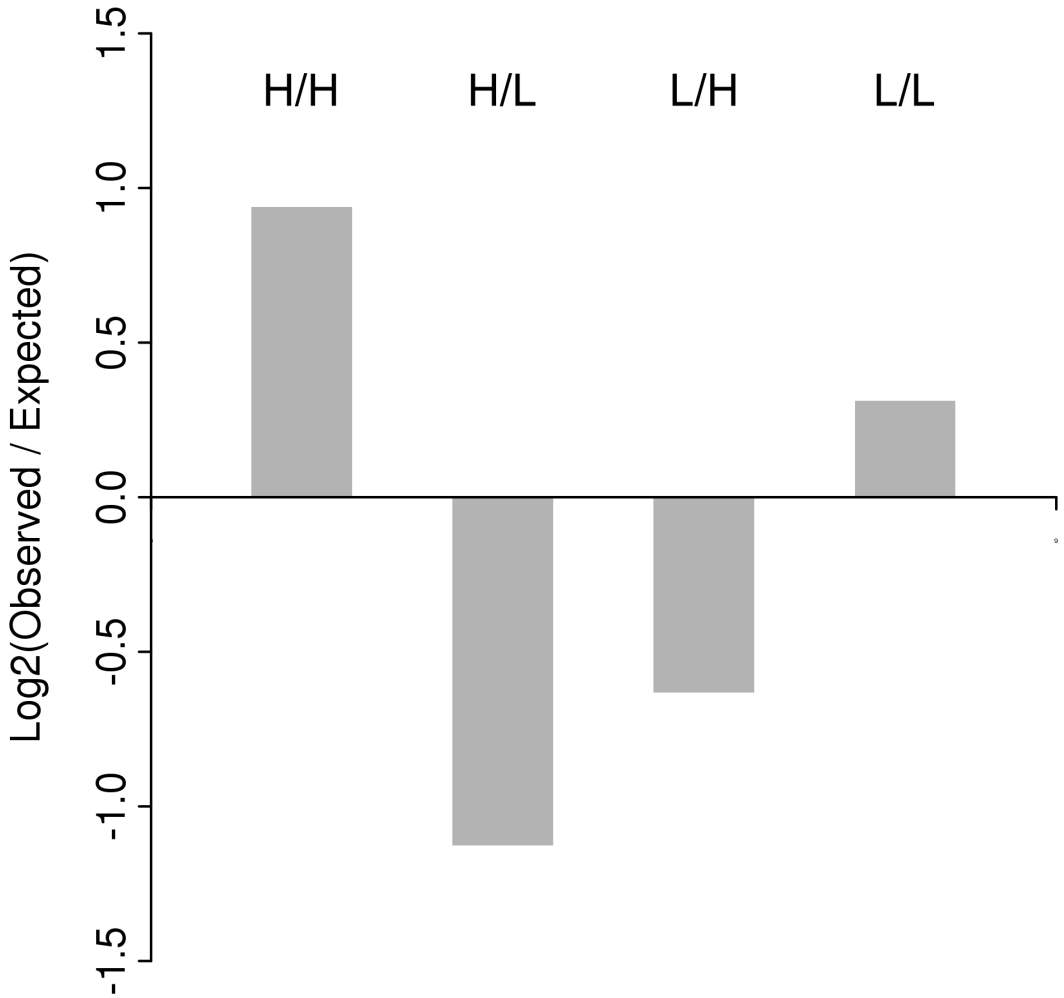
Supplementary Figure S82. Correlation between genic and cis-NAT promoter activity in HUVEC. Cis-NAT promoters in the HUVEC cell type were identified using CAGE data from non-polyadenylated RNA form cytosol isolates. Activity of genic promoters and the sum of corresponding cis-NAT promoter activity was measured by CAGE tag counts. A Spearman rank correlation was used to determine the relationship between total cis-NAT promoter activity and genic promoter activity.



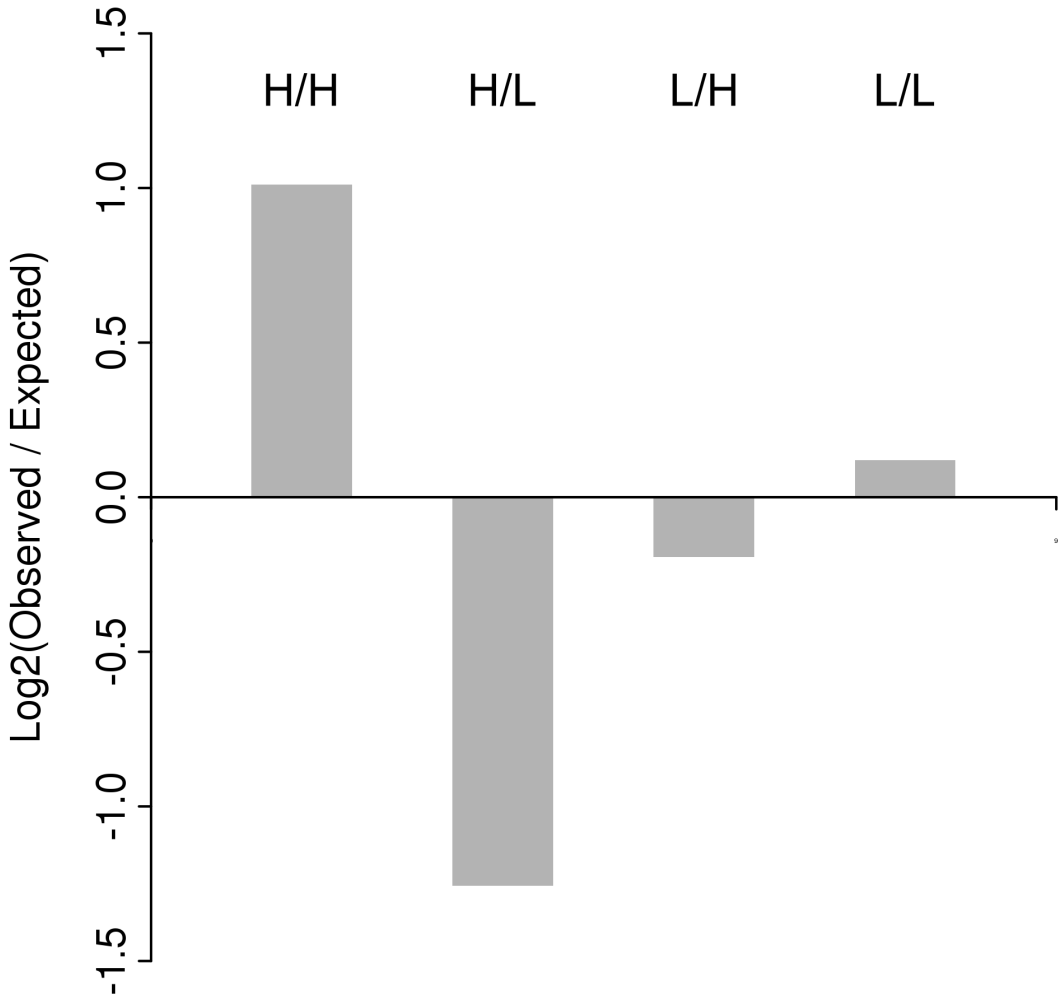
Supplementary Figure S83. Correlation between genic and cis-NAT promoter activity in K562. Cis-NAT promoters in the K562 cell type were identified using CAGE data from non-polyadenylated RNA form nucleus isolates. Activity of genic promoters and the sum of corresponding cis-NAT promoter activity was measured by CAGE tag counts. A Spearman rank correlation was used to determine the relationship between total cis-NAT promoter activity and genic promoter activity.



Supplementary Figure S84. Correlation between genic and cis-NAT promoter activity in NHEK. Cis-NAT promoters in the NHEK cell type were identified using CAGE data from non-polyadenylated RNA form nucleus isolates. Activity of genic promoters and the sum of corresponding cis-NAT promoter activity was measured by CAGE tag counts. A Spearman rank correlation was used to determine the relationship between total cis-NAT promoter activity and genic promoter activity.



Supplementary Figure S85. Association of cis-NAT promoter activity and genic promoter activity for cis-NAT promoters distal to genic promoters. Cis-NAT and genic promoters were classified into high (H) and low (L) categories based on their activity levels across all cell types analyzed here. The observed versus expected levels of association between the resulting four possible category combinations – 1) high cis-NAT & high gene (HH), 2) high cis-NAT & low gene (HL), 3) low cis-NAT & high gene (L/H), 4) low cis-NAT & low gene (L/L) – were then computed using association mining.



Supplementary Figure S86. Association of cis-NAT promoter activity and genic promoter activity for cis-NAT promoters proximal to genic promoters. Cis-NAT and genic promoters were classified into high (H) and low (L) categories based on their activity levels across all cell types analyzed here. The observed versus expected levels of association between the resulting four possible category combinations – 1) high cis-NAT & high gene (HH), 2) high cis-NAT & low gene (HL), 3) low cis-NAT & high gene (L/H), 4) low cis-NAT & low gene (L/L) – were then computed using association mining.