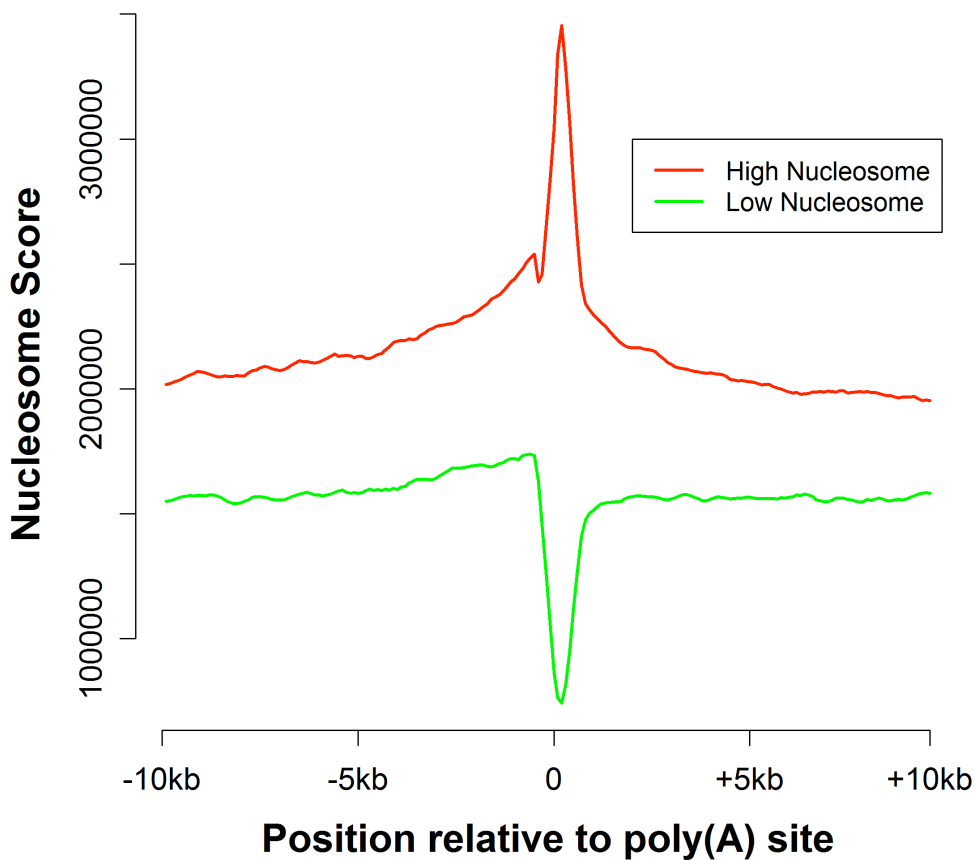
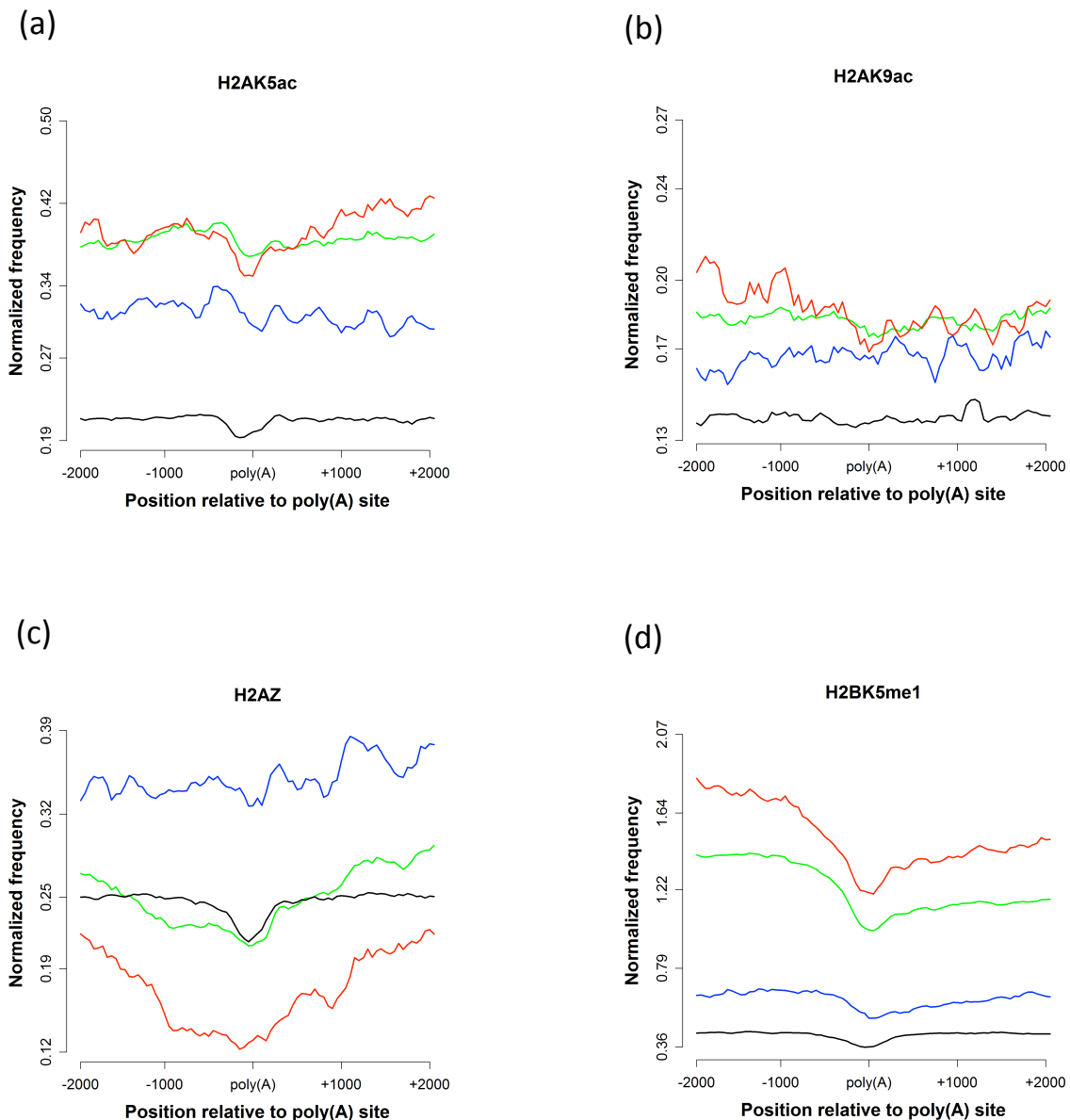


Supplementary Figures

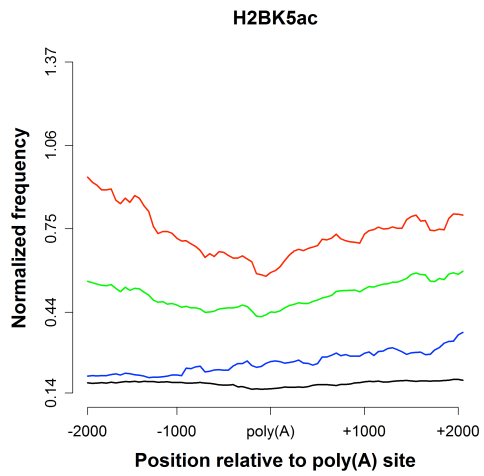
Supplementary Figure 1. Distribution of nucleosome occupancy scores in 20 kb region centered at poly(A) sites for high and low downstream nucleosome occupancy classes. Profile of nucleosome occupancy scores around the poly(A) sites with high (top 25 percentile) (red) and low (bottom 25 percentile) (green) downstream nucleosome affinity in the 600 bp region. The Y-axis shows the total number of pooled chip-seq tags in 500 bps bins and sliding window of 100 bps.



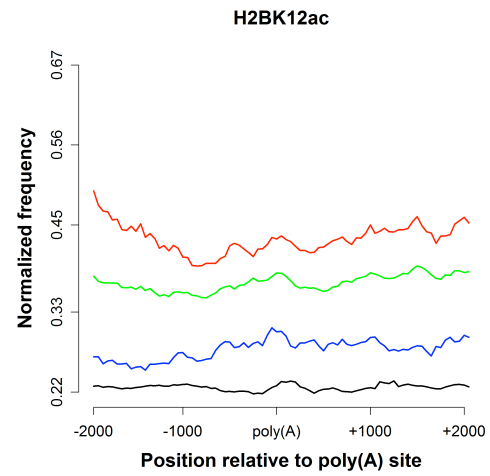
Supplementary Figure 2. Distribution of each histone modification in 4 kb region centered at poly(A) sites. The figures show for each HM the average profiles for all 3' UTR poly(A) sites (green, 31871 sites), only the poly(A) sites belonging to high expression genes (red, 7583 sites) and the poly(A) sites belonging to the low expression genes (blue, 7577 sites), as well as for the 87679 non-poly(A) positions that form the negative set (black).



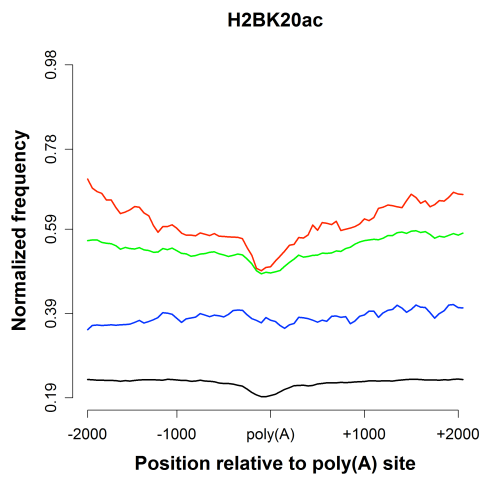
(e)



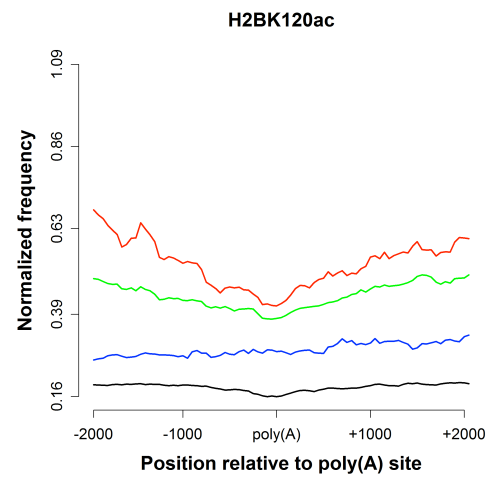
(f)



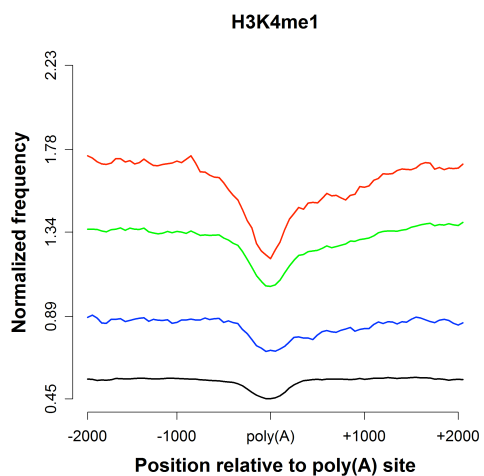
(g)



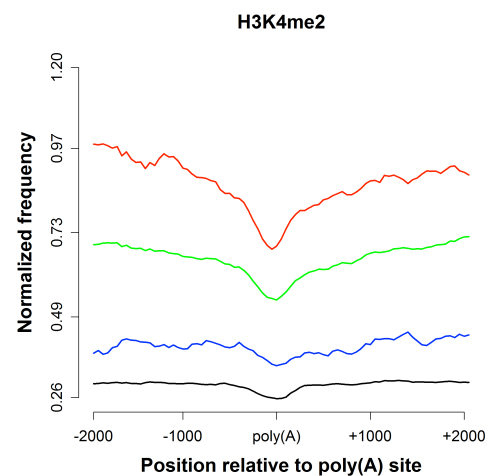
(h)



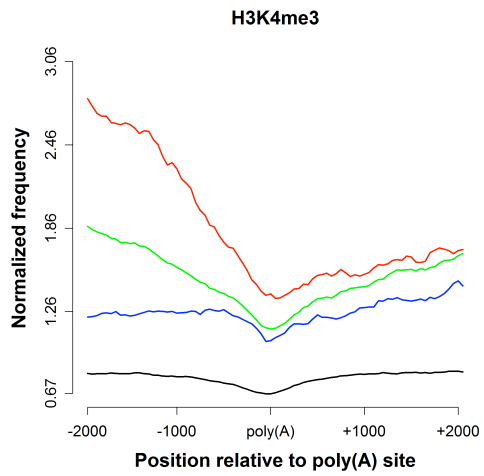
(i)



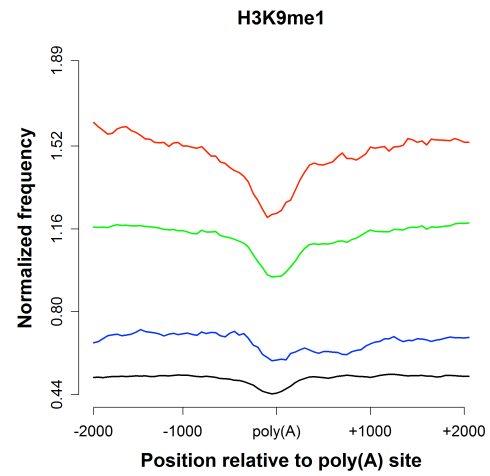
(j)



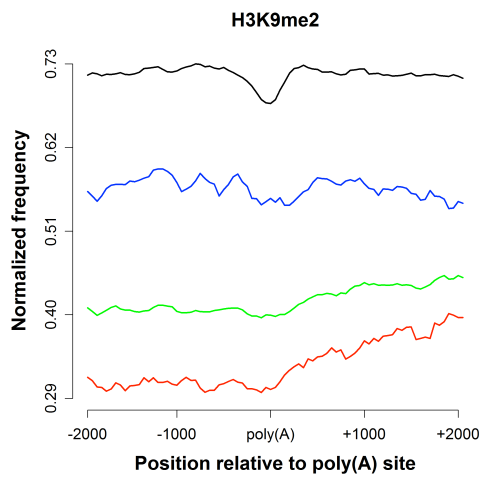
(k)



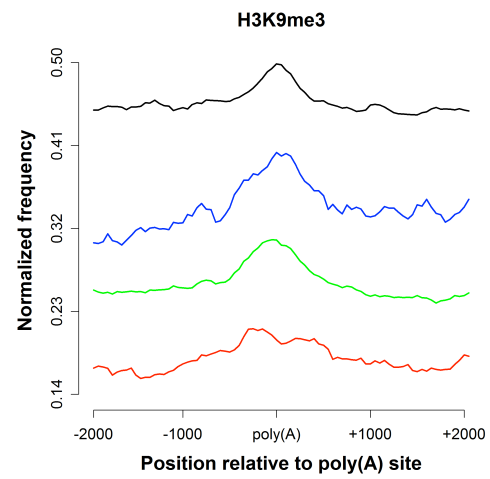
(l)



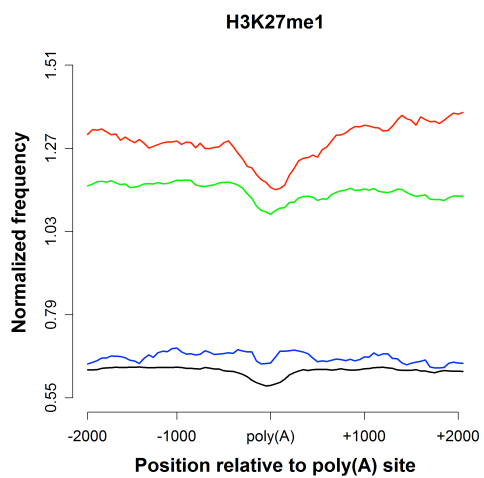
(m)



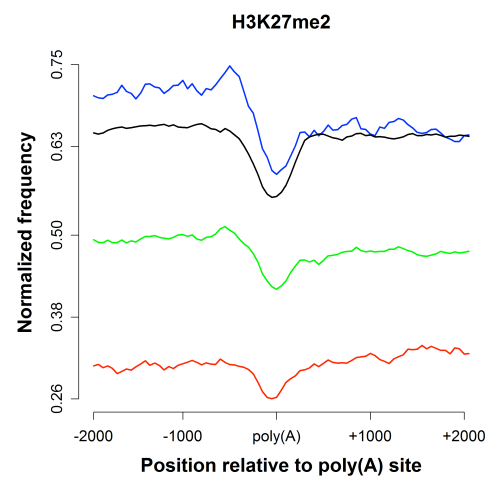
(n)



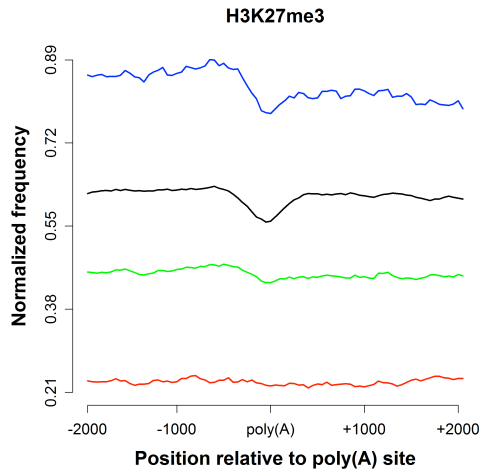
(o)



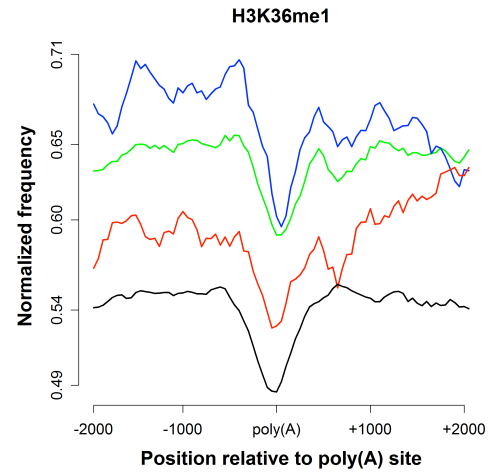
(p)



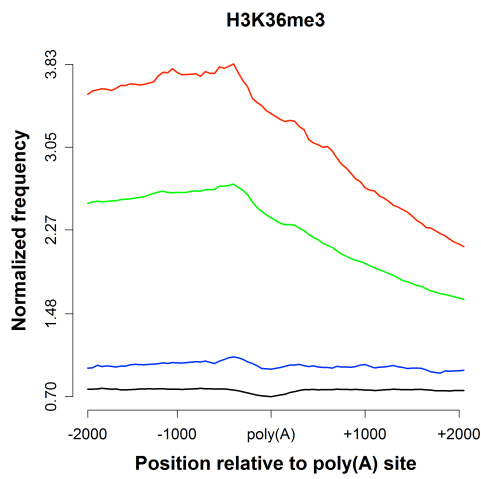
(q)



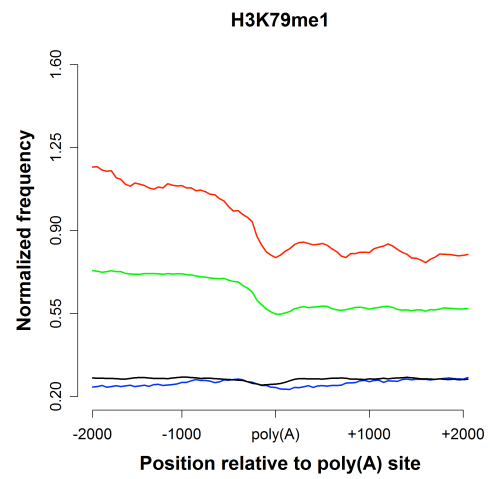
(r)



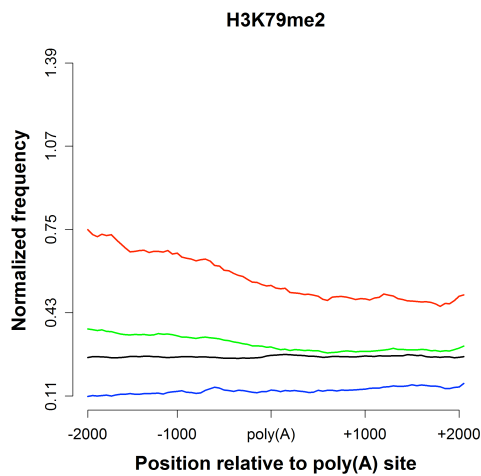
(s)



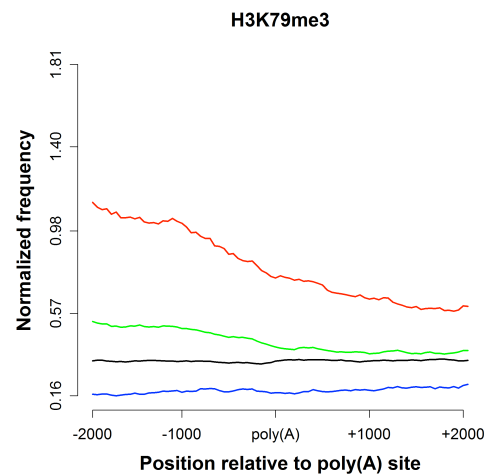
(t)



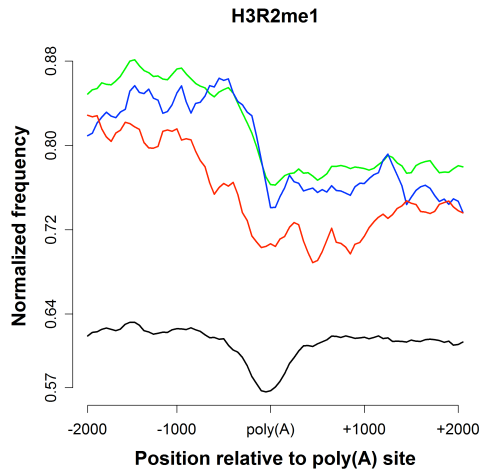
(u)



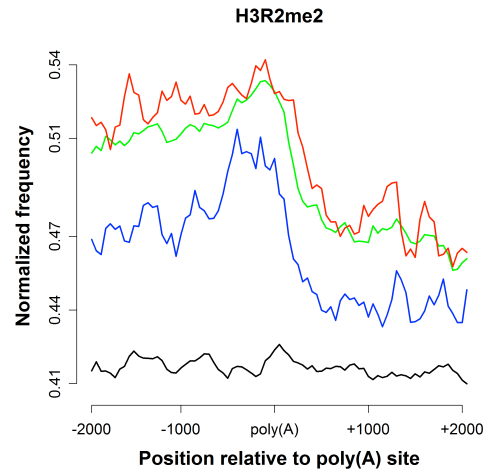
(v)



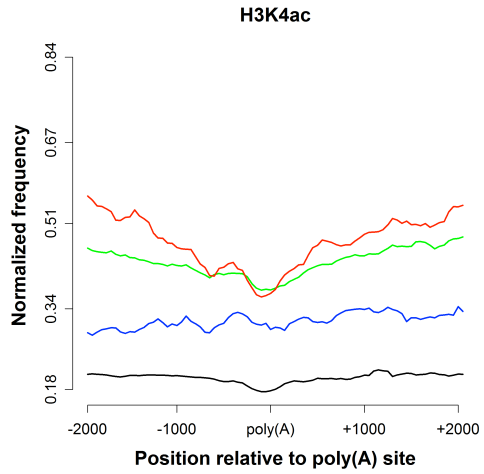
(w)



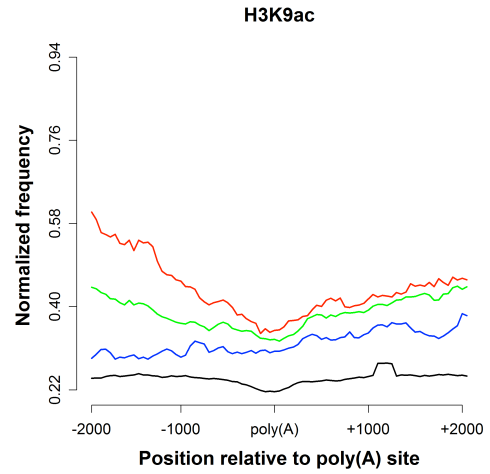
(x)



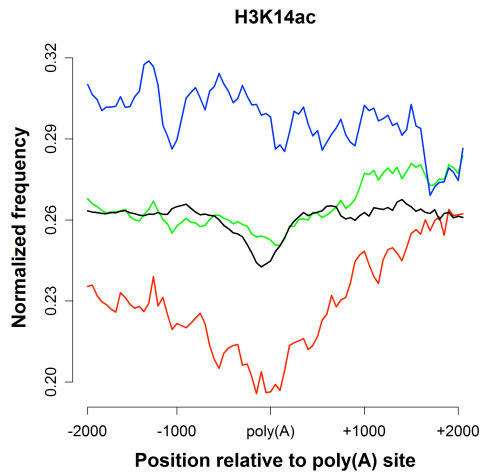
(y)



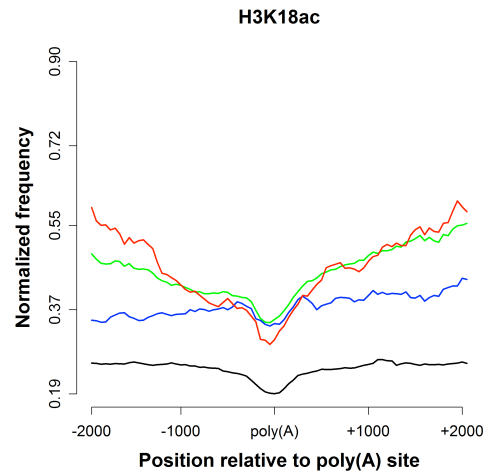
(z)



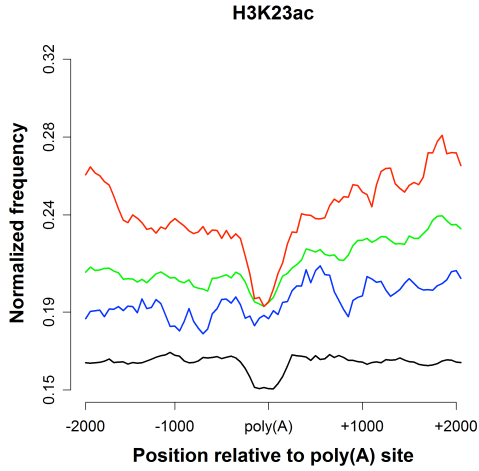
(aa)



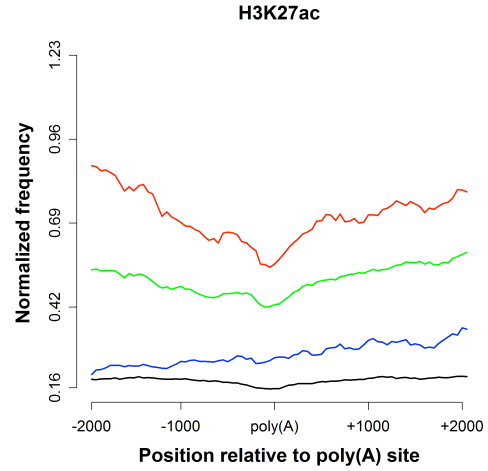
(ab)



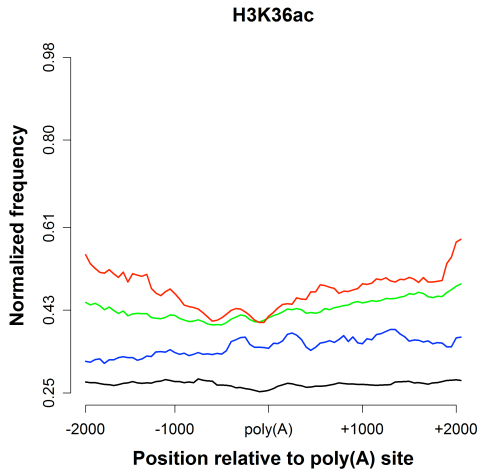
(ac)



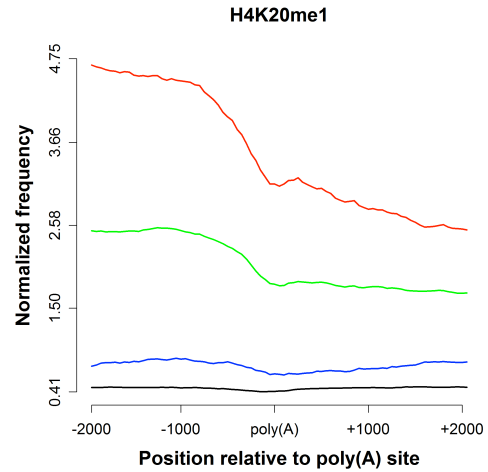
(ad)



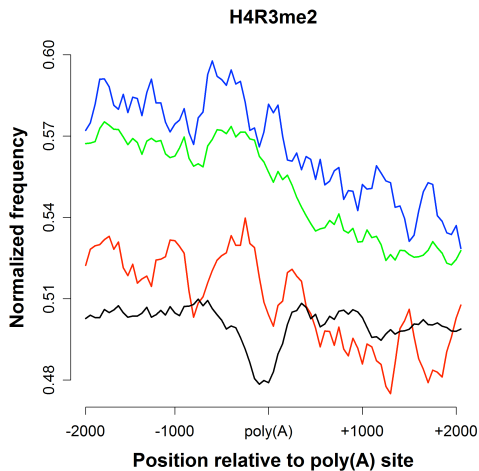
(ae)



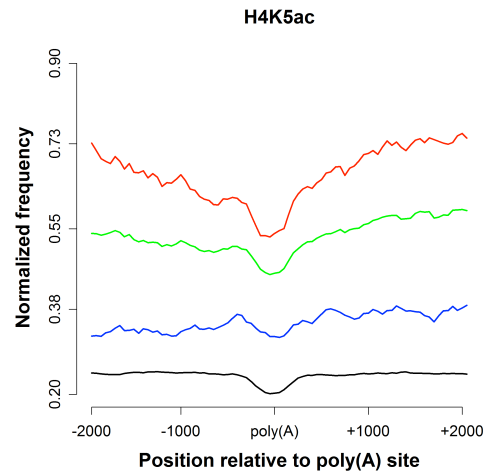
(af)



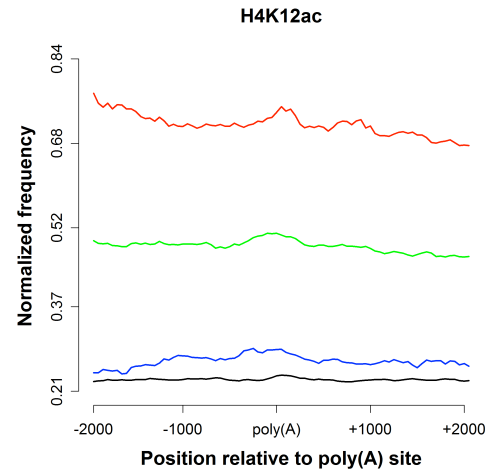
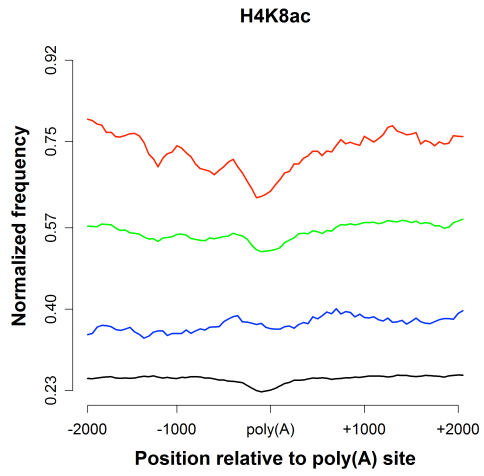
(ag)



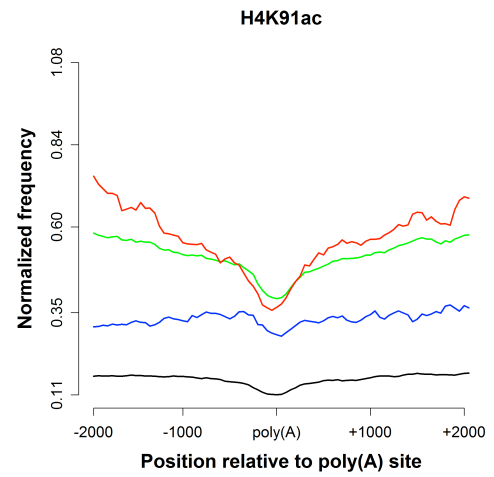
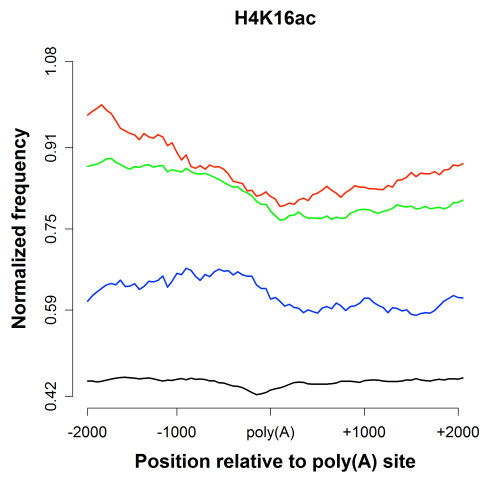
(ah)



(aj)

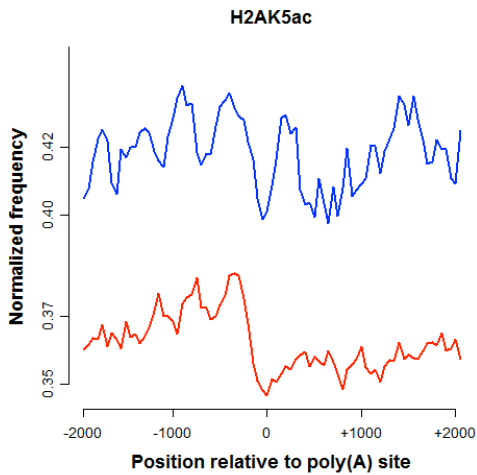


(al)

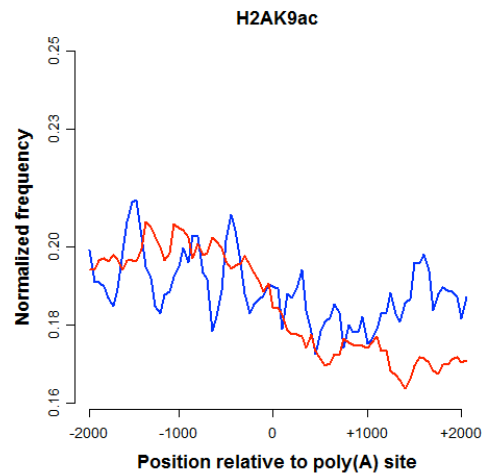


Supplementary Figure 3. Distribution of each histone modification in 4 kb region centered at poly(A) sites for high and low usage poly(A) sites. For each HM the figure shows the average profiles for high usage poly(A) sites (blue, 5139 sites) and low usage poly(A) sites (red, 21204 sites).

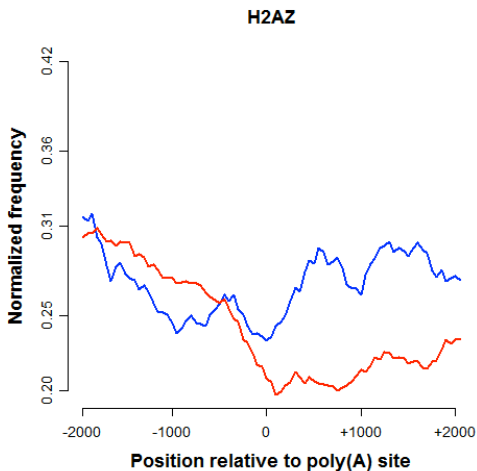
(a)



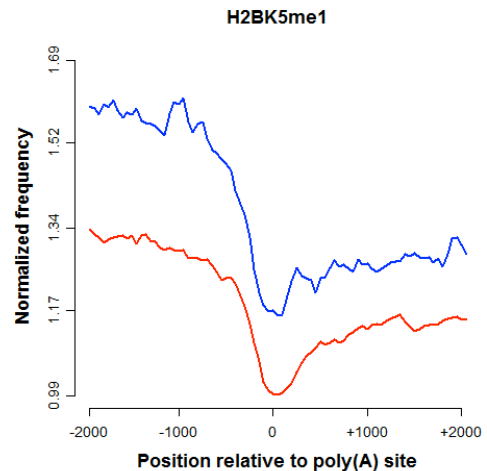
(b)



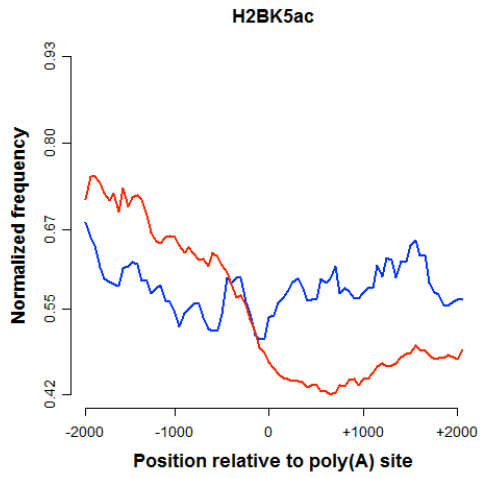
(c)



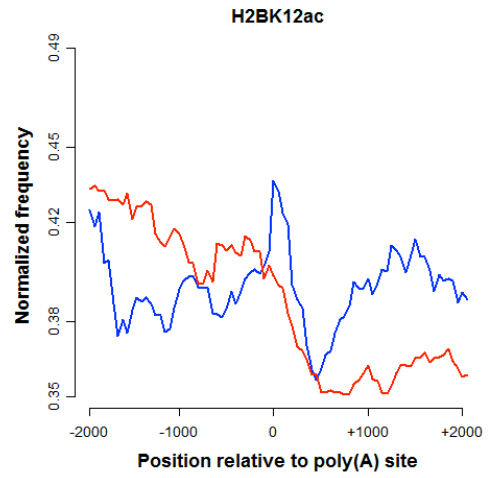
(d)



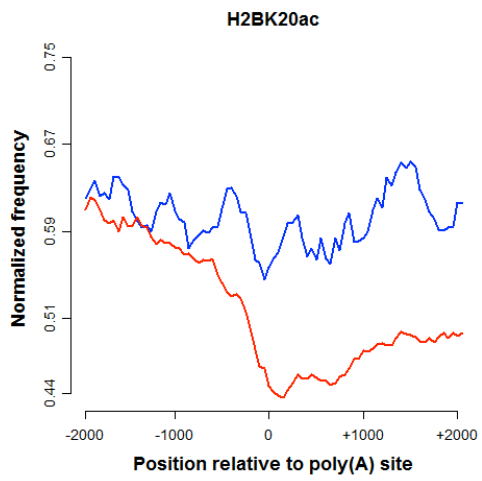
(e)



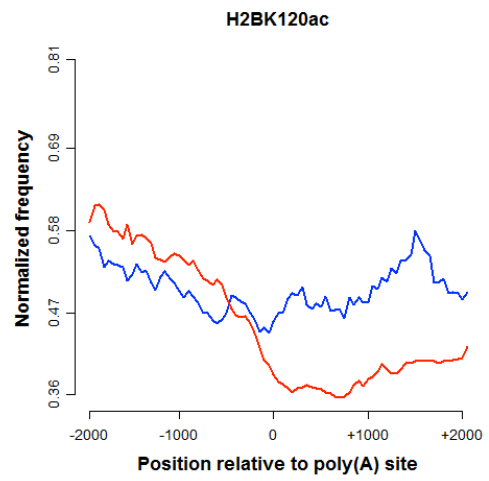
(f)



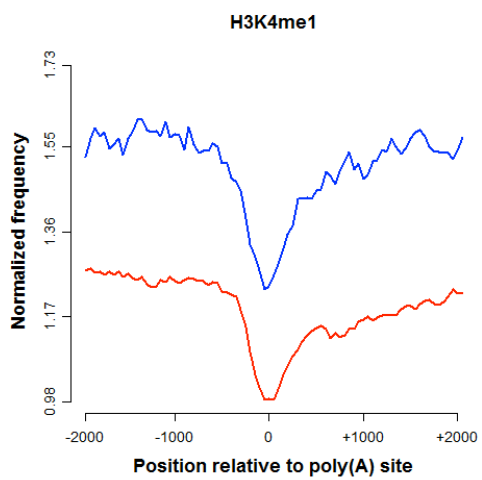
(g)



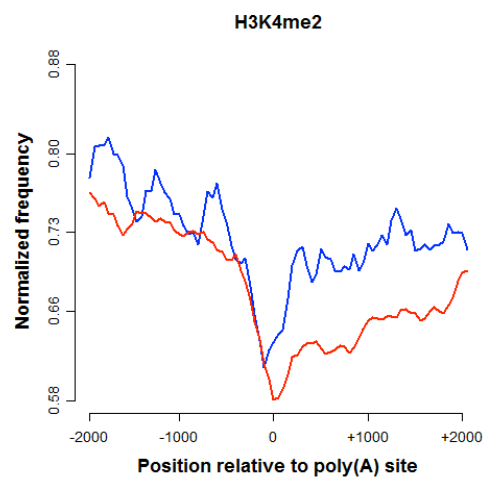
(h)



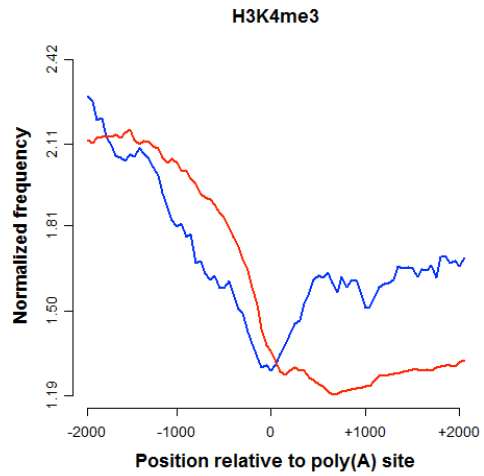
(i)



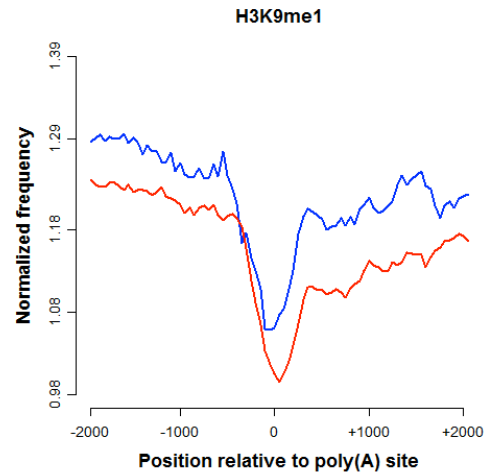
(j)



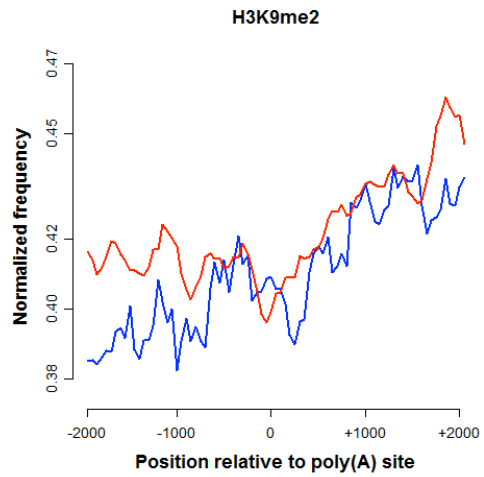
(k)



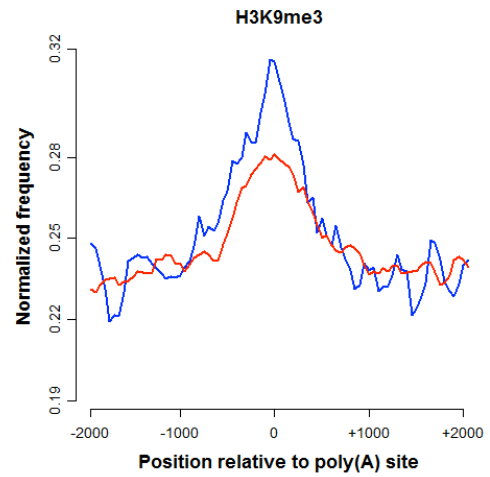
(l)



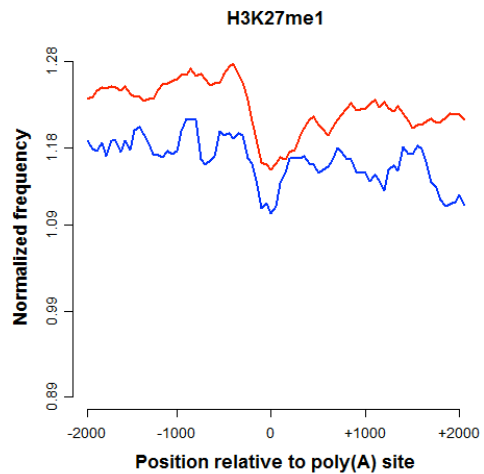
(m)



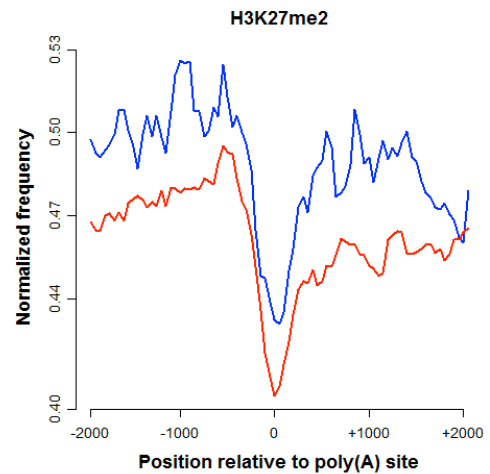
(n)



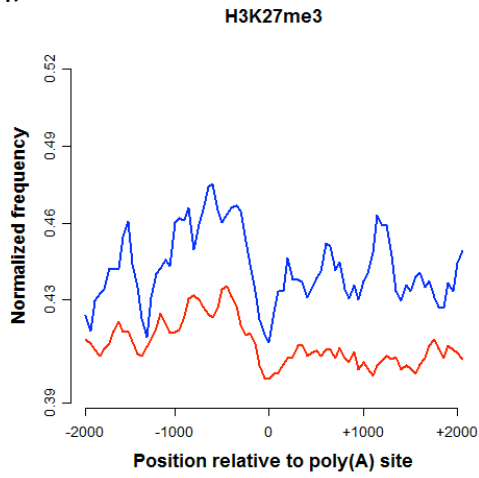
(o)



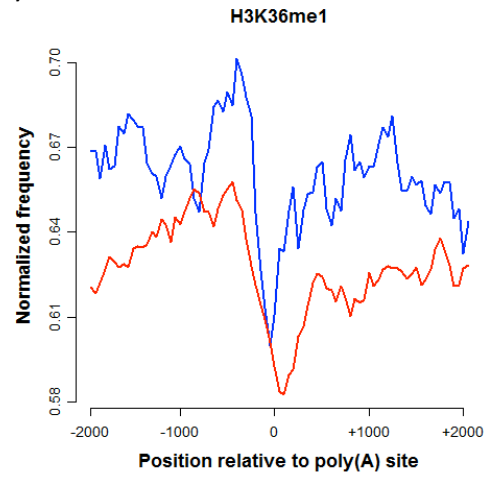
(p)



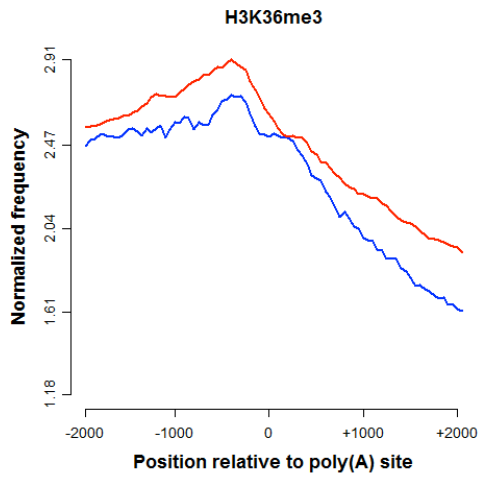
(q)



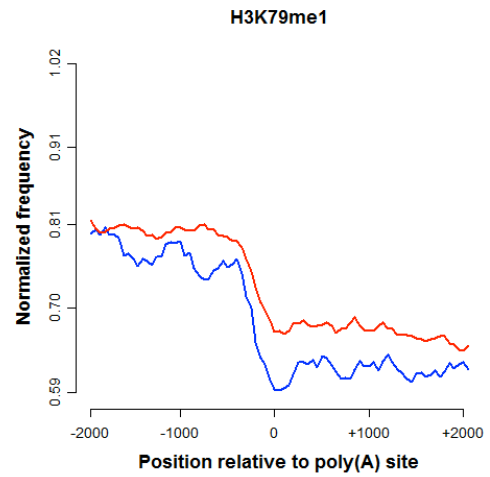
(r)



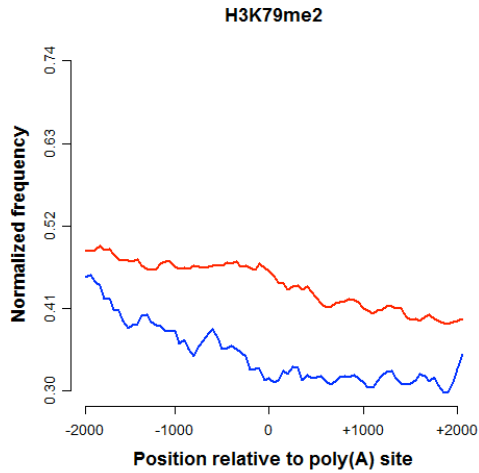
(s)



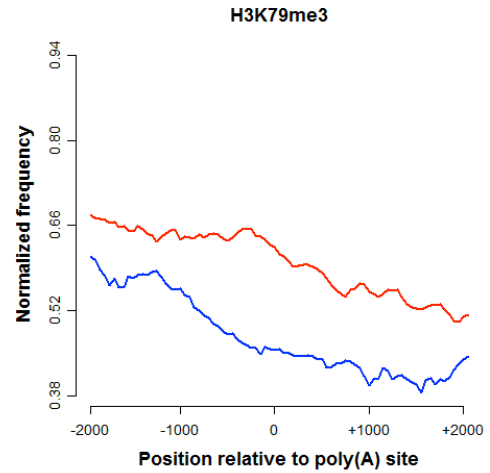
(t)



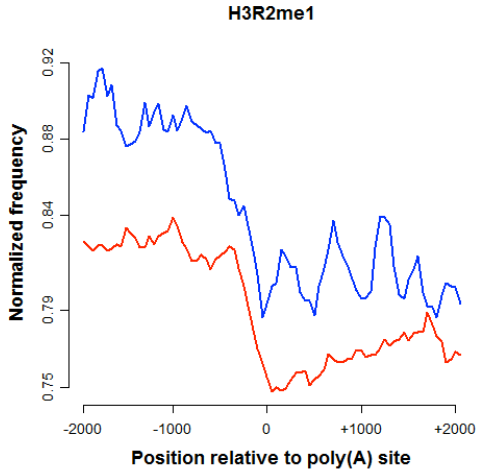
(u)



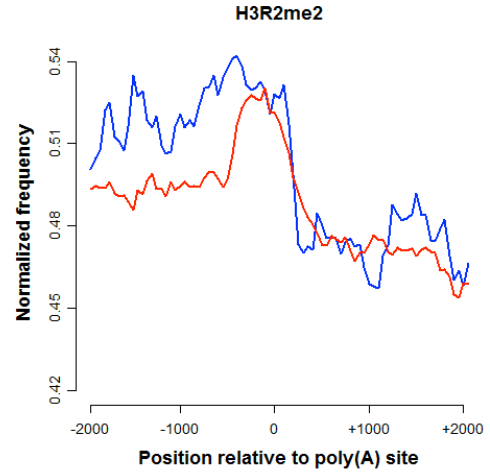
(v)



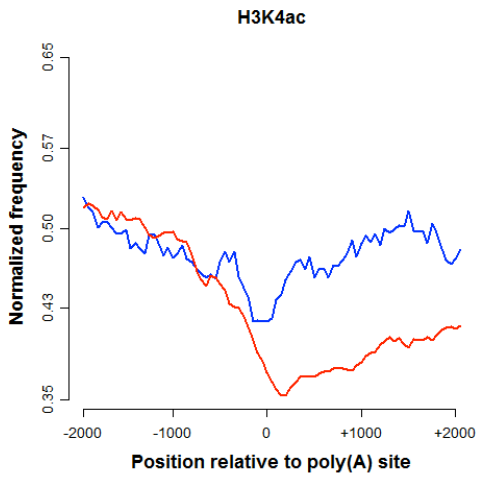
(w)



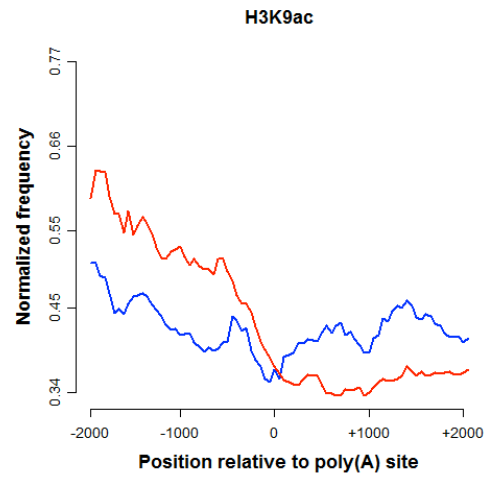
(x)



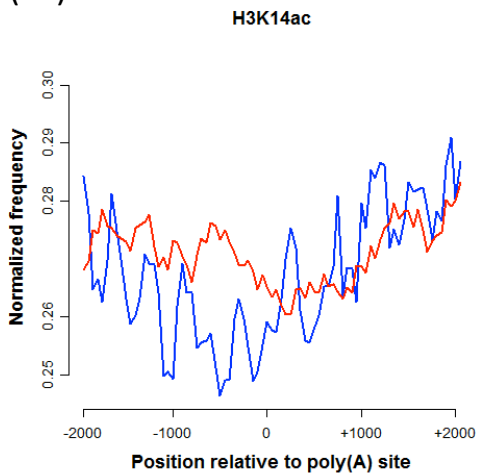
(y)



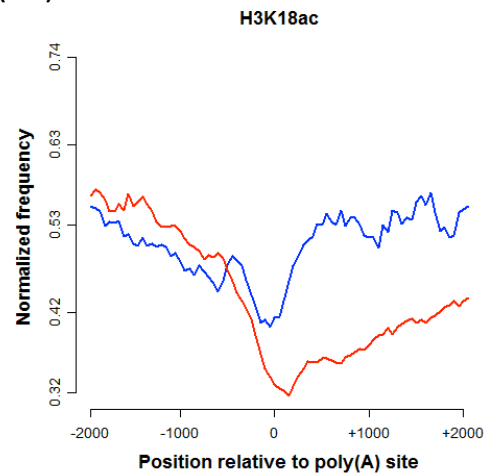
(z)



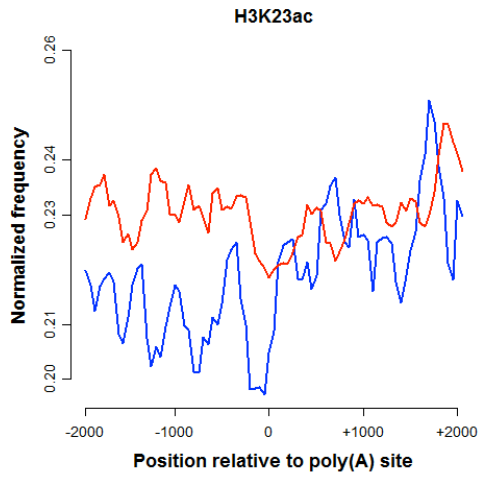
(aa)



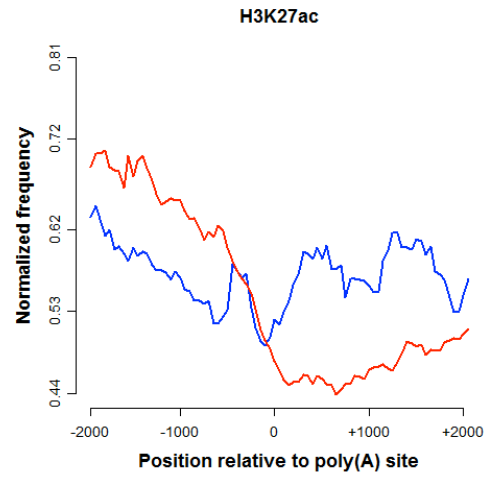
(ab)



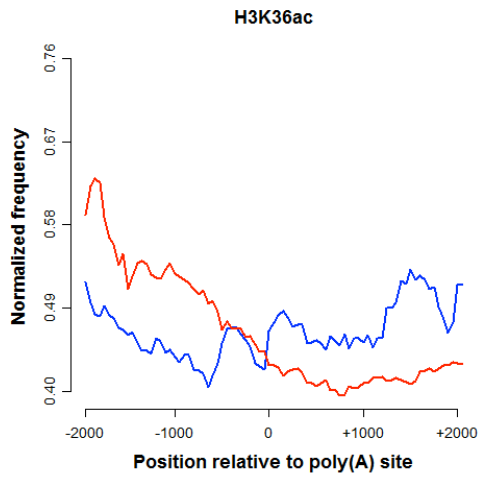
(ac)



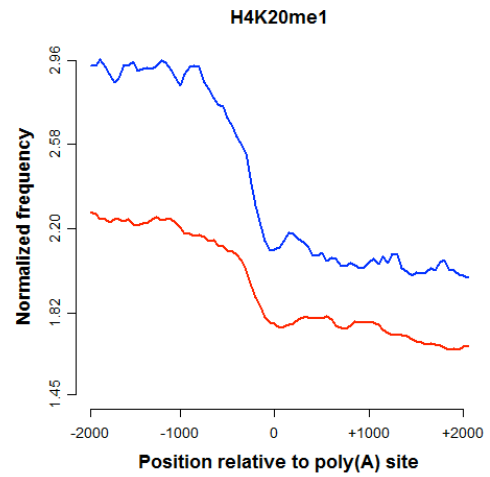
(ad)



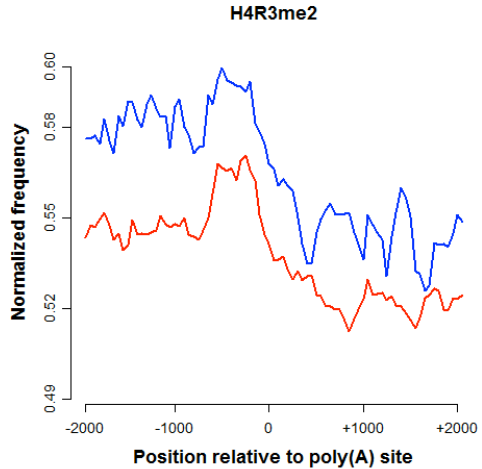
(ae)



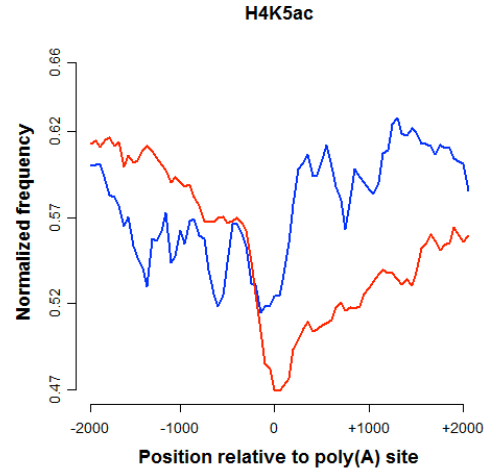
(af)



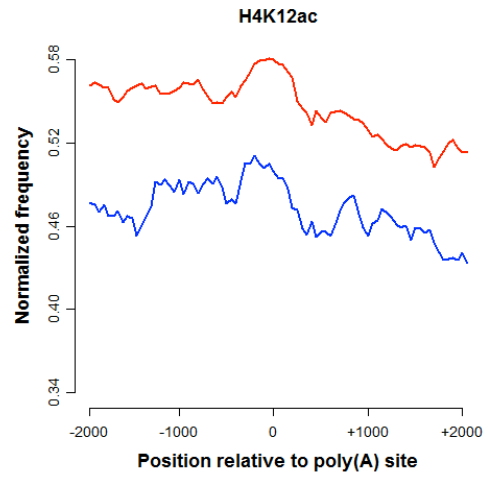
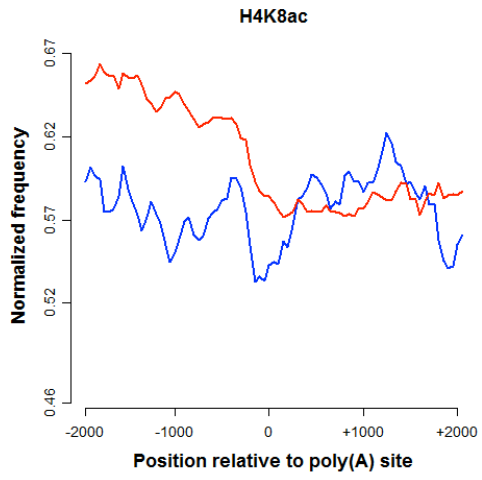
(ag)



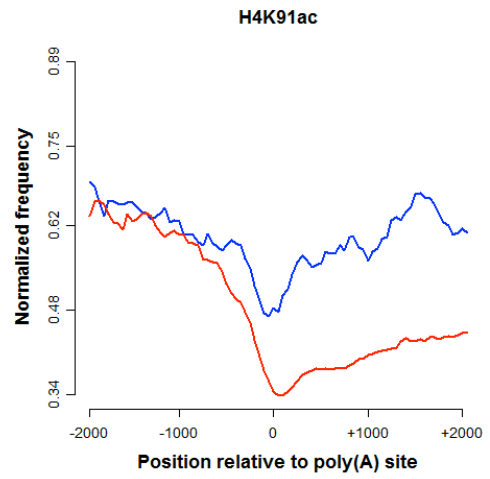
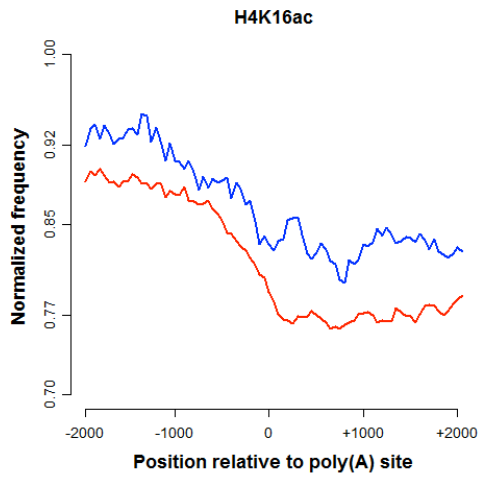
(ah)



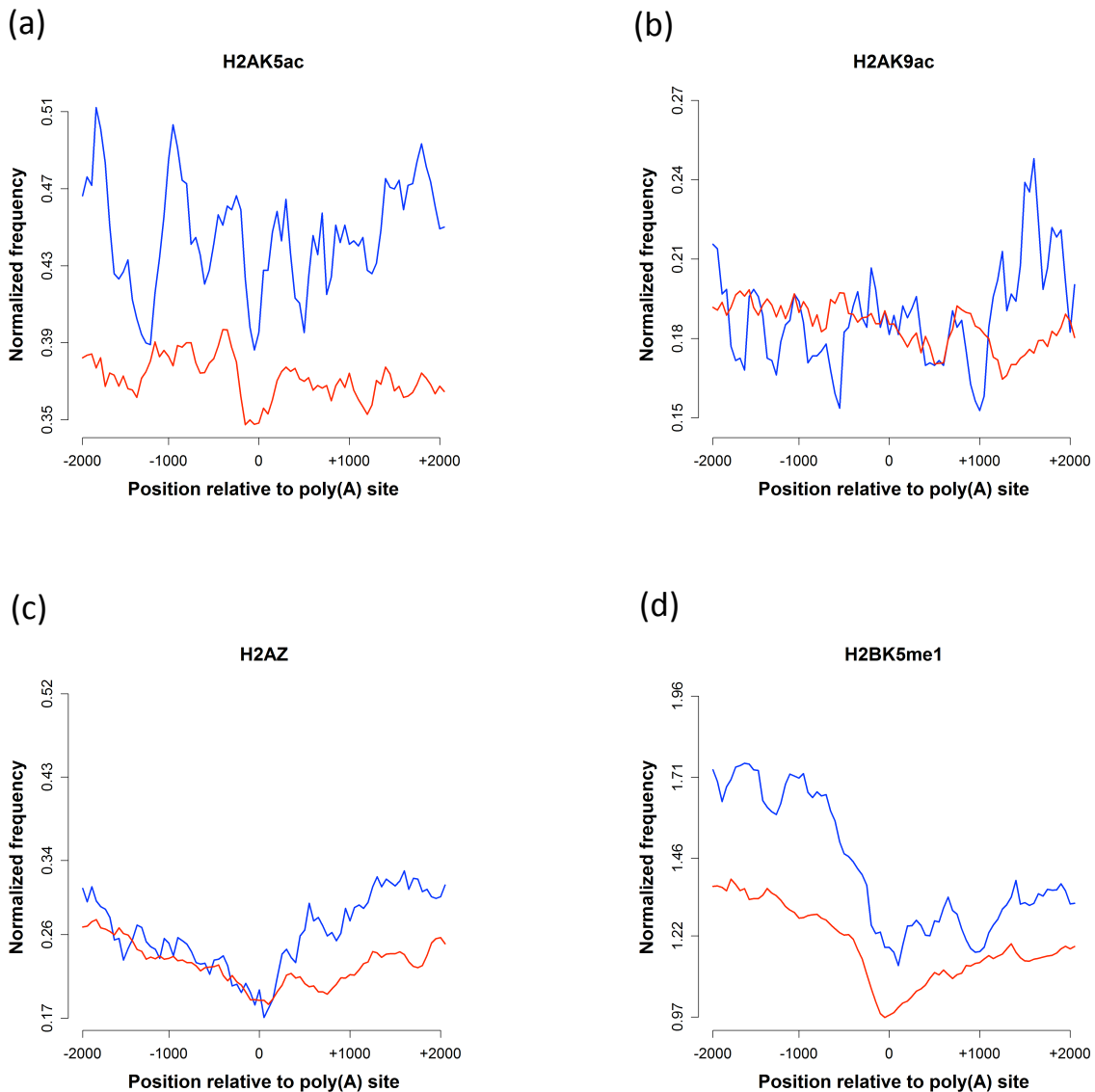
(aj)



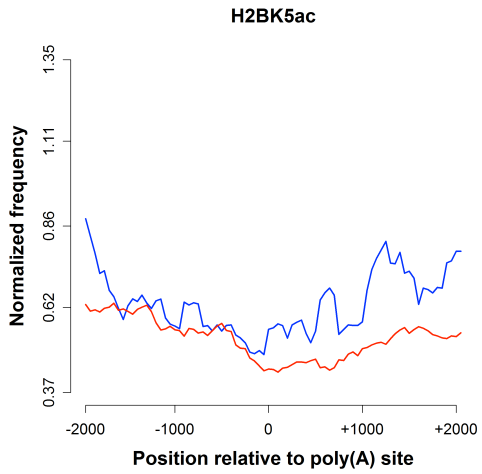
(a)



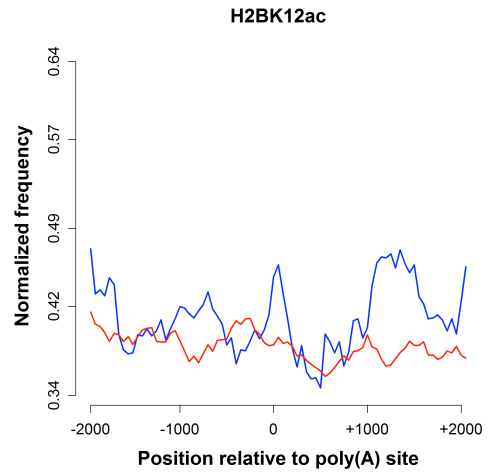
Supplementary Figure 4. Distribution of each histone modification in 4 kb region centered at poly(A) sites for high and low usage poly(A) sites found using EST/cDNAs as well as CD4+ T cell RNA-seq data. For each HM the figure shows the average profiles for high usage poly(A) sites (blue, 1113 sites) and low usage poly(A) sites (red, 7796 sites).



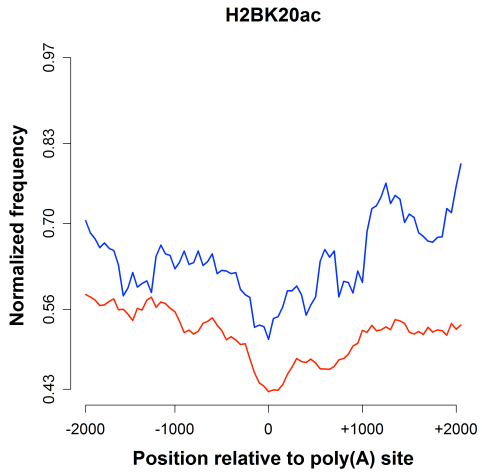
(e)



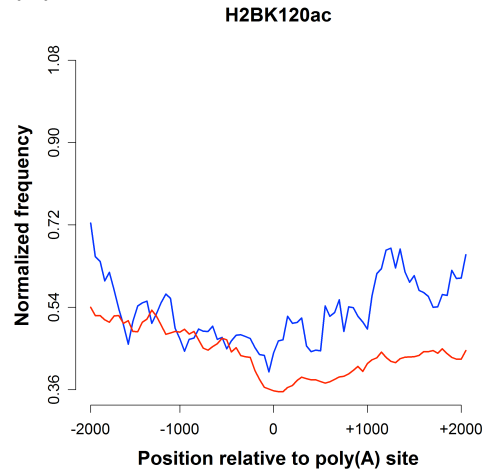
(f)



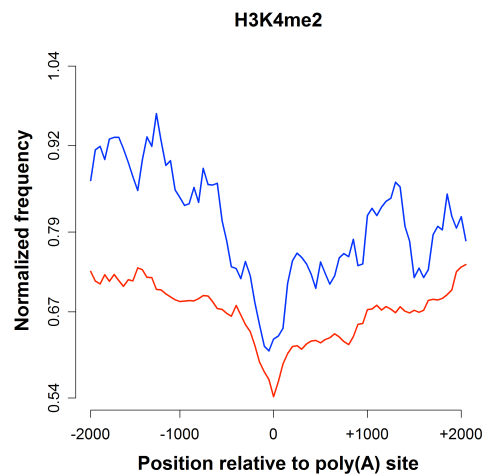
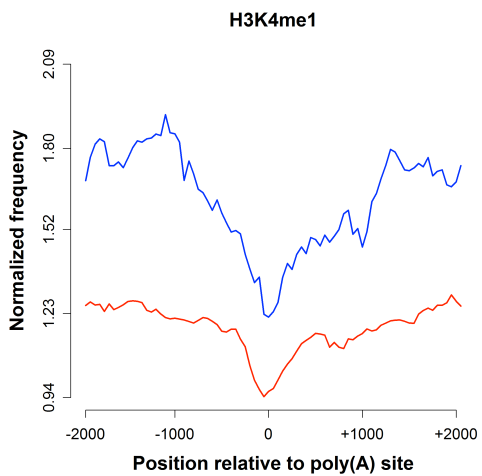
(g)

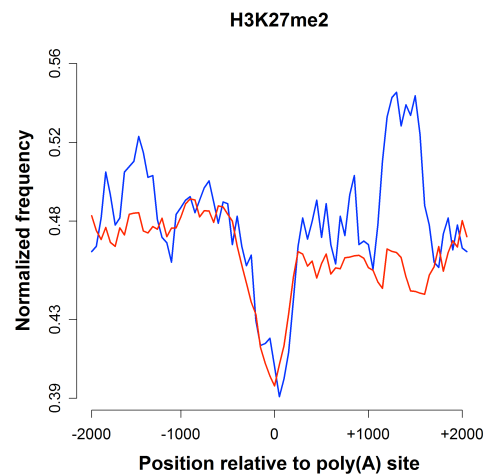
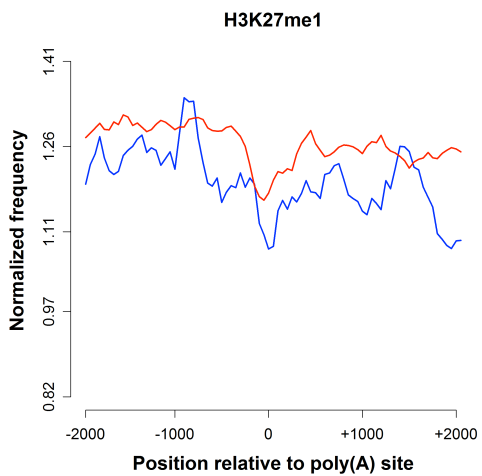
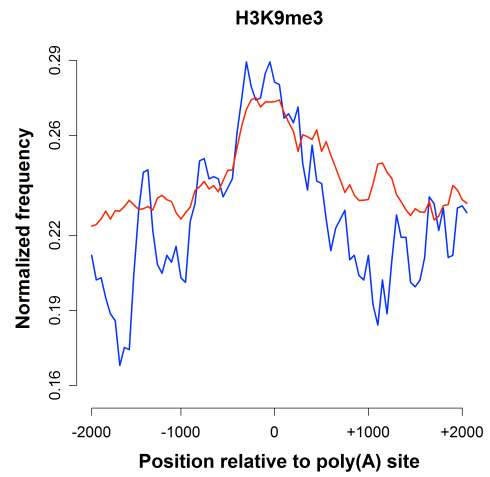
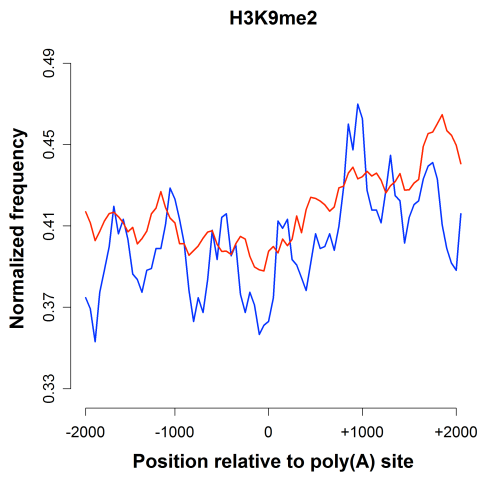
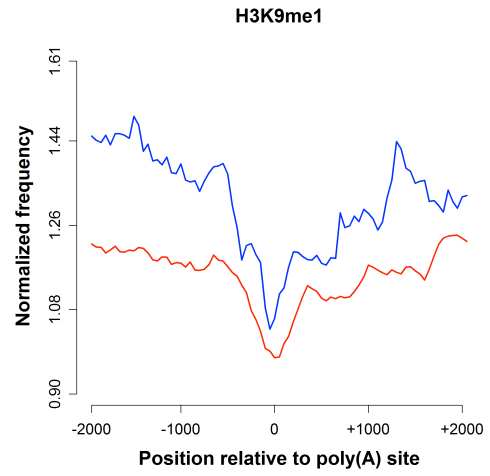
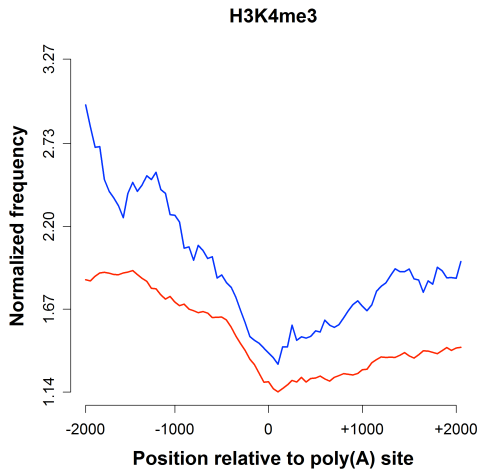


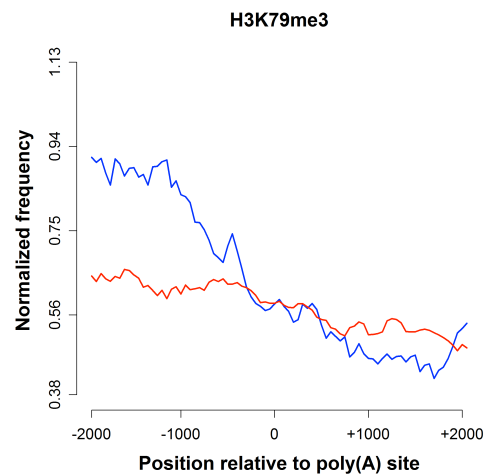
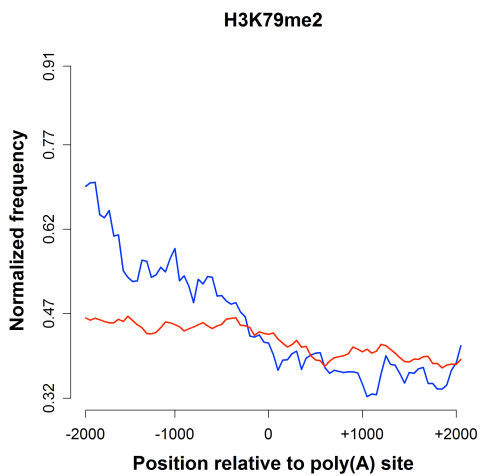
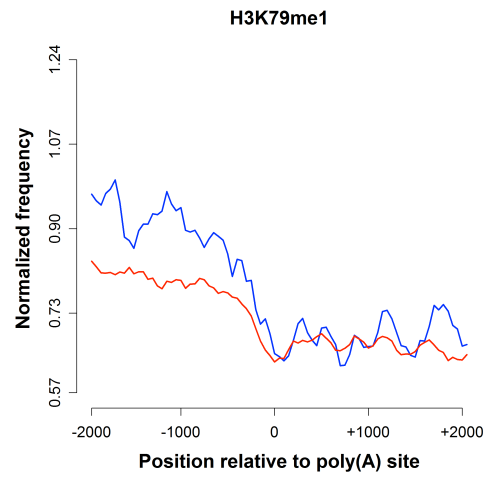
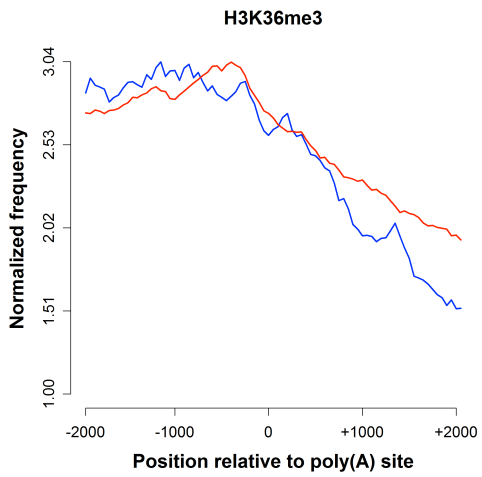
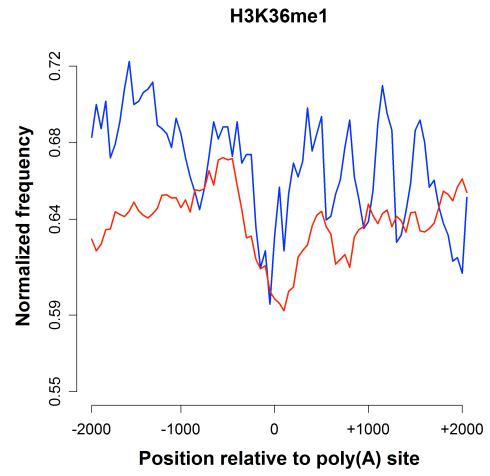
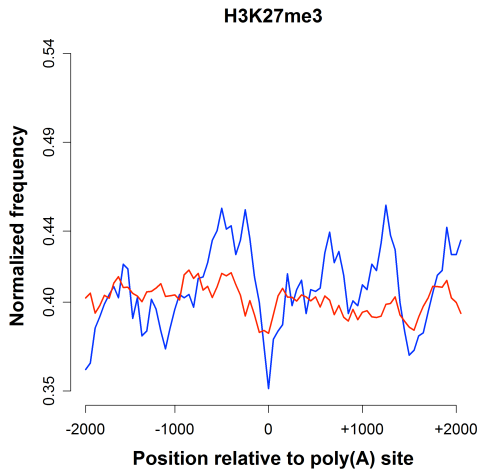
(h)

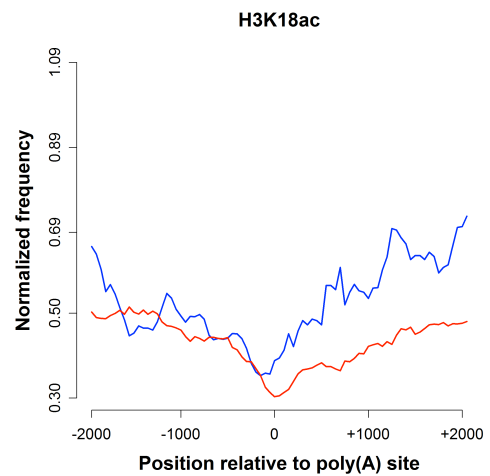
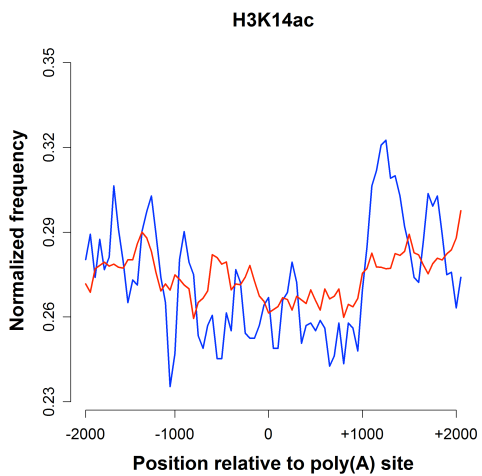
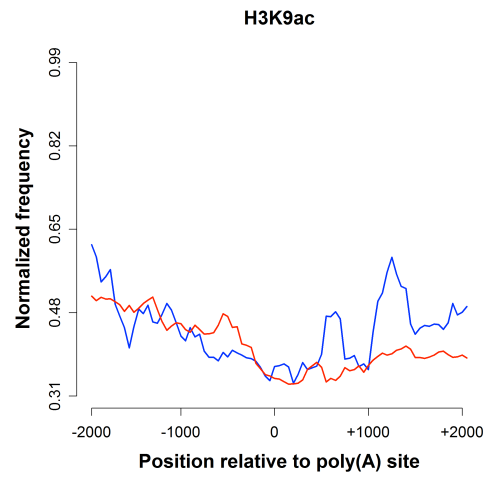
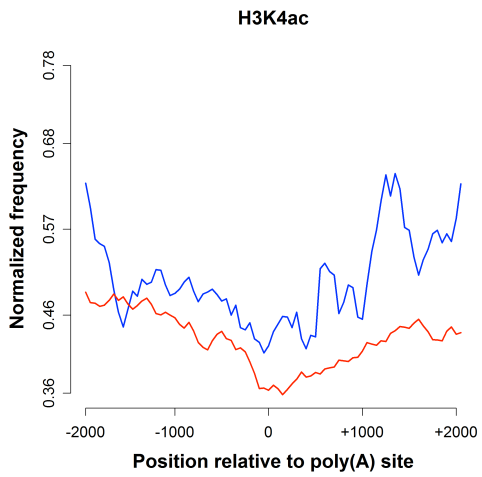
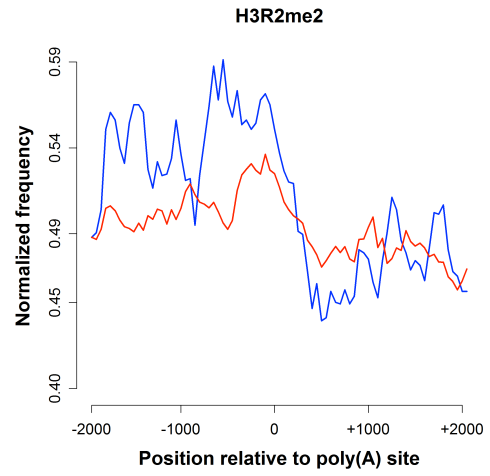
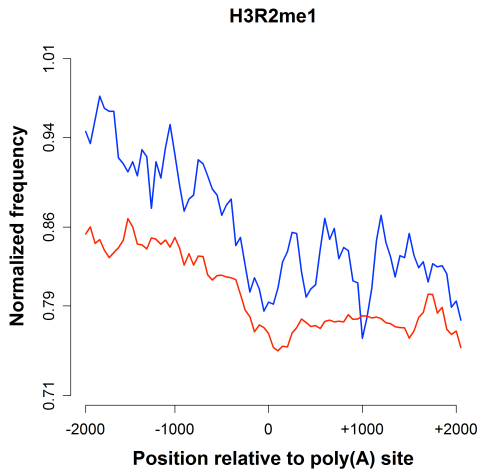


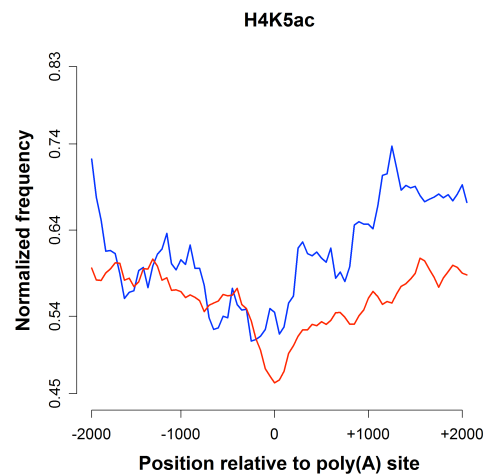
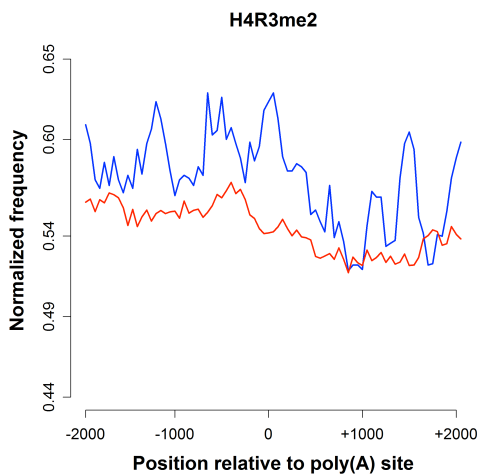
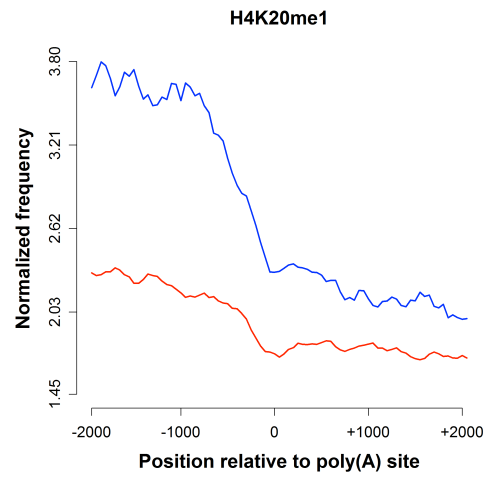
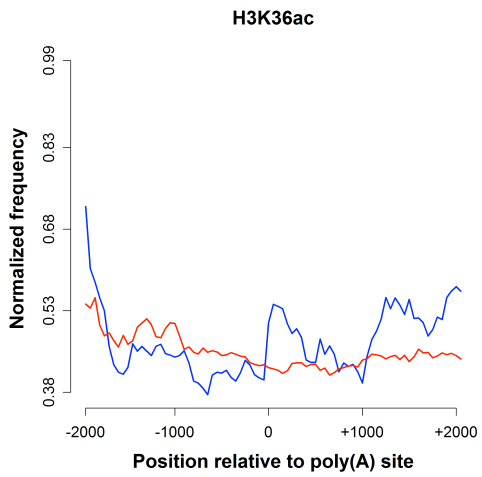
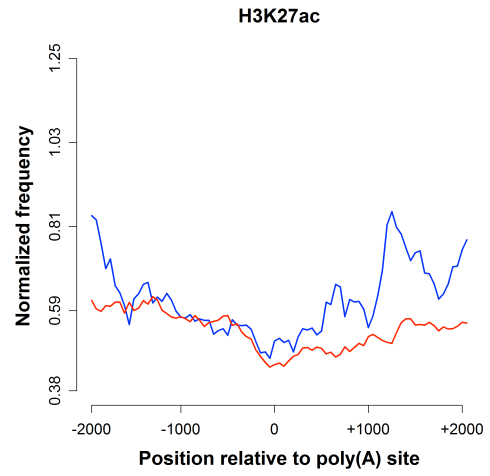
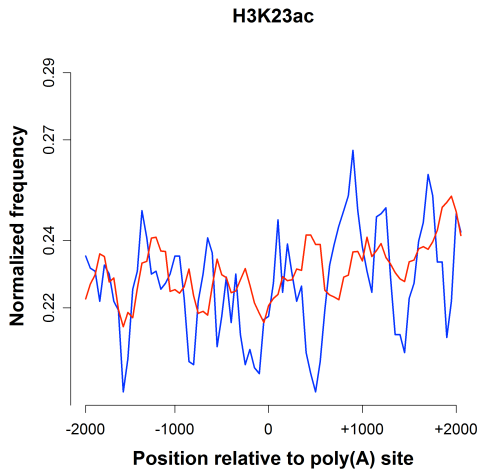
(i)

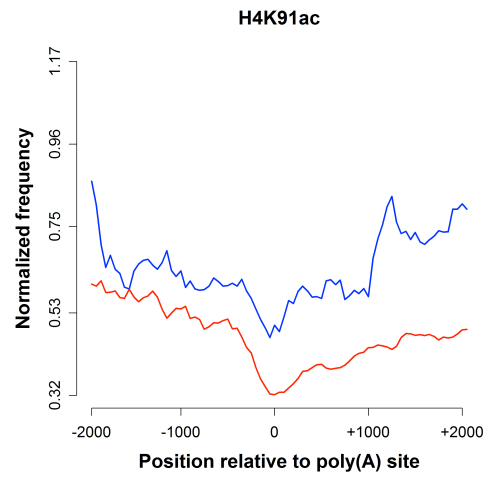
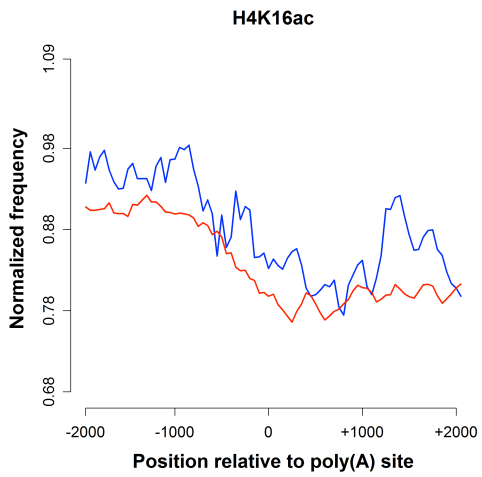
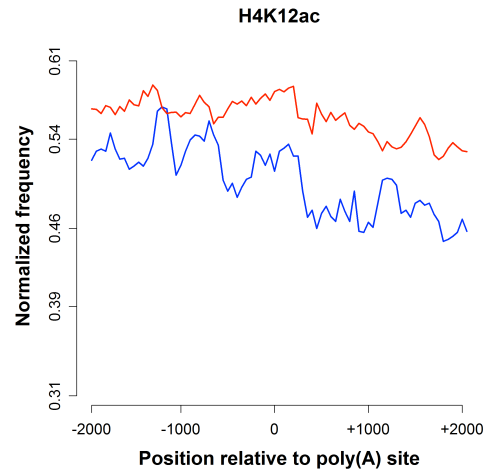
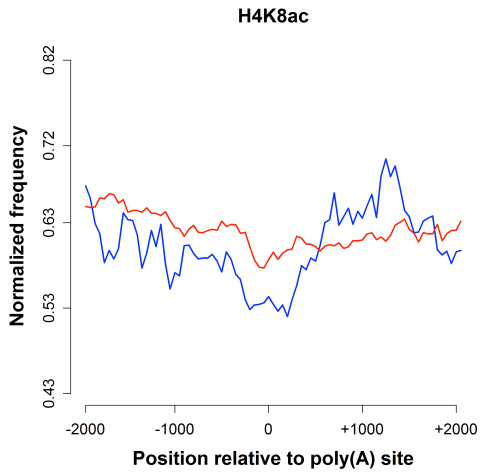












Supplementary Tables

Supplementary Table 1. Support Vector Machine Based Classification of Authentic Poly(A) Sites Using Post-translational Histone Modification Data

Histone Modification	Classification Accuracy*	Sensitivity*	Specificity*
H3K36me3	71.5% - 71.7%	63.8%-63.9%	79.3%-79.4%
H4K20me1	65.6% - 65.8%	52.5%-52.6%	78.5%-78.6%
H2BK5me1	63.4% - 63.6%	50%-50.1%	76.7%-76.8%
H3K79me1	63.2% - 63.4%	50.8%-50.9%	75.6%-75.7%
H4K8ac	63.2% - 63.4%	52.5%-52.6%	74.1%-74.2%
H3K9me1	63.1% - 63.3%	53.1%-53.2%	73.1%-73.2%
H3K27me1	62.7% - 62.9%	50.6%-50.7%	74.8%-74.9%
H4K12ac	62.6% - 62.8%	49%-49.1%	76.2%-76.3%
H3K27ac	61.5% - 61.7%	54.5%-54.6%	68.5%-68.6%
H2BK5ac	61.2% - 61.4%	53.1%-53.2%	69.2%-69.3%
H4K5ac	61.2% - 61.4%	54.1%-54.2%	68.4%-68.5%
H4K91ac	60.7% - 60.9%	49.4%-49.5%	71.9%-72%
H2BK20ac	60.6% - 60.8%	57.4%-57.5%	63.8%-63.9%
H3K4me2	60.5% - 60.7%	48.4%-48.5%	72.6%-72.7%
H4K16ac	60.2% - 60.4%	48.8%-48.9%	71.5%-71.6%
H3K4me1	59.2% - 59.4%	47.5%-47.6%	69.9%-71%
H3K9me2	59% - 59.2%	76.3%-76.4%	41.7%-41.8%
H2BK120ac	58.8% - 59%	52.3%-52.4%	65.3%-65.4%
H3K4ac	58.5% - 58.7%	53%-53.1%	64%-64.1%
H3K9me3	58.1% - 58.3%	78.1%-78.2%	38.1%-38.2%
H2AK5ac	57.3% - 57.5%	48.2%-48.3%	66.4%-66.5%
H3K27me3	57% - 57.2%	73.5%-73.6%	40.4%-40.5%
H3K79me2	56.9% - 57.1%	46%-46.1%	67.7%-67.8%
H3K27me2	56.7% - 56.9%	72.4%-72.5%	40.9%-41%
H2BK12ac	56.6% - 56.8%	47.4%-47.5%	65.9%-66%
H3K36ac	56.5% - 56.7%	47.1%-47.2%	65.8%-65.9%
H3K18ac	56.3% - 56.5%	50.6%-50.7%	61.9%-62%
H2AZ	56% - 56.2%	63.7%-63.8%	48.2%-48.3%
H3K79me3	55.9% - 56.1%	44.6%-44.7%	67.1%-67.2%
H3R2me1	55.2% - 55.4%	48.4%-48.5%	61.9%-62%
H3K4me3	54.7% - 54.9%	49.4%-49.5%	60%-60.1%
H3R2me2	54.3% - 54.5%	45.1%-45.2%	63.5%-63.6%
H3K23ac	54.1% - 54.3%	46.7%-46.8%	61.3%-61.4%
H3K9ac	53.9% - 54.1%	41.9%-42%	65.8%-65.9%
H3K36me1	53.7% - 53.9%	48.1%-48.2%	59.3%-59.4%
H2AK9ac	52.7% - 52.9%	39.1%-39.2%	66.2%-66.3%
H4R3me2	52.6% - 52.8%	47.3%-47.4%	57.8%-57.9%
H3K14ac	50.9% - 51.1%	57.7%-57.8%	44.1%-44.2%

* Controlled for GC content , 95% confidence interval

Supplementary Table 2. Support Vector Machine Based Classification of High Usage and Low Usage Poly(A) Sites (obtained from cDNA/EST Based Approach) Using Post-translational Histone Modification Data

Histone Modification	Classification Accuracy*	Sensitivity*	Specificity*
H3K36me3	66.2% - 66.4%	72.3%-72.4%	60.1%-60.2%
H3K27me1	61% - 61.1%	70.4%-70.5%	51.3%-51.4%
H3K9me1	58.5% - 58.7%	65.7%-65.8%	51.4%-51.5%
H2BK5me1	58.4% - 58.6%	59.4%-59.5%	57.2%-57.3%
H4K20me1	58.2% - 58.4%	64.6%-64.7%	52%-52.1%
H3K4me1	58.1% - 58.3%	51.7%-51.8%	64.3%-64.4%
H3R2me1	57.6% - 57.8%	58.8%-58.9%	56.4%-56.5%
H4K16ac	57.5% - 57.7%	61%-61.1%	54.1%-54.2%
H3K4me3	57.5% - 57.7%	56.3%-56.4%	58.7%-58.8%
H3K79me1	57% - 57.1%	76.1%-76.2%	37.6%-37.7%
H4K12ac	56.4% - 56.6%	72.7%-72.8%	39.9%-40%
H3K36me1	56.3% - 56.5%	52%-52.1%	60.7%-60.8%
H4K8ac	56.2% - 56.4%	70.6%-70.7%	41.7%-41.8%
H4K91ac	56.1% - 56.3%	37.1%-37.2%	75%-75.1%
H3K4me2	55.5% - 55.7%	50.2%-50.3%	61.1%-61.2%
H4R3me2	55.3% - 55.5%	52.5%-52.6%	58.2%-58.3%
H2BK20ac	55.3% - 55.5%	39.5%-39.6%	71%-71.1%
H3K79me3	55.1% - 55.3%	75.2%-75.3%	35.1%-35.2%
H3K27me2	55% - 55.2%	48.4%-48.5%	61.3%-61.4%
H2AK5ac	54.6% - 54.8%	39%-39.1%	70.4%-70.5%
H4K5ac	54.6% - 54.8%	58.7%-58.8%	50.4%-50.5%
H3R2me2	54.3% - 54.5%	56.6%-56.7%	52%-52.1%
H3K9me2	54.2% - 54.4%	54.7%-54.8%	53.5%-53.6%
H3K27me3	54.1% - 54.3%	47.2%-47.3%	60.8%-60.9%
H3K18ac	54% - 54.2%	33.8%-33.9%	74.2%-74.3%
H3K4ac	53.9% - 54.1%	39.5%-39.6%	68.3%-68.4%
H3K79me2	53.7% - 53.8%	79.4%-79.5%	28%-28.1%
H2BK12ac	53.6% - 53.8%	64.3%-64.4%	42.8%-42.9%
H3K36ac	53.6% - 53.8%	59.7%-59.8%	47.4%-47.5%
H2BK120ac	53.5% - 53.7%	32.2%-32.3%	74.7%-74.8%
H2BK5ac	53.4% - 53.6%	55%-55.1%	51.7%-51.8%
H3K27ac	53.4% - 53.6%	60.6%-60.7%	46.2%-46.3%
H3K9me3	53.3% - 53.5%	65.3%-65.4%	41.2%-41.3%
H3K9ac	53.2% - 53.4%	44.6%-44.7%	61.8%-61.9%
H3K14ac	52.7% - 52.9%	67.8%-67.9 %	37.5%-37.6%
H2AZ	52.5% - 52.7%	28.7%-28.8%	76.2%-76.3%
H3K23ac	52.4% - 52.6%	67.6%-67.7%	37.2%-37.3%
H2AK9ac	51.8% - 52%	37.3%-37.4%	66.2%-66.3%

* 95% confidence interval

Supplementary Table 3. Support Vector Machine Based Classification of the High Usage and Low Usage Poly(A) Sites (obtained by CD4+ T cell RNA-seq data as well as cDNA/EST Based Approach) using Post-translational Histone Modification Data

Histone Modification	Classification Accuracy*	Sensitivity*	Specificity*
H3K36me3	68.9%-69.9%	80.3%-80.4%	58.3%-58.4%
H3K27me1	64.9%-65.3%	76.8%-76.9%	53.1%-53.2%
H3K9me1	62.1%-62.5%	71.5%-71.6%	52.8%-52.9%
H3R2me1	61.4%-61.8%	69.6%-69.7%	53.4%-53.5%
H2BK5me1	61.3%-61.7%	68.1%-68.2%	54.8%-54.9%
H4K16ac	61.1%-61.5%	68.7%-68.8%	53.7%-53.8%
H3K36me1	60.5%-60.9%	64.2%-64.3%	57%-57.1%
H4K20me1	60.4%-60.8%	66.2%-66.3%	55%-55.1%
H3K4me1	60.2%-60.6%	60.1%-60.2%	60.4%-60.5%
H3K4me3	60.2%-60.6%	66.3%-66.4%	54.2%-54.3%
H4K12ac	59%-59.4%	75%-75.1%	43.2%-43.3%
H4K8ac	58.7%-59.1%	70.2%-70.3%	47.4%-47.5%
H4R3me2	58.5%-58.9%	60.1%-60.2%	57%-57.1%
H3K79me1	58.3%-58.7%	61.3%-61.4%	55.4%-55.5%
H3K4me2	58.3%-58.6%	57.8%-57.9%	59%-59.1%
H4K91ac	57.9%-58.3%	45.8%-45.9%	70.2%-70.3%
H3R2me2	57.2%-57.6%	60.6%-60.7%	54%-54.1%
H4K5ac	57.1%-57.5%	57%-57.1%	57.6%-57.7%
H2BK20ac	56.6%-56.9%	49.1%-49.2%	64.3%-64.4%
H3K27me2	56.6%-56.9%	56.5%-56.6%	56.8%-56.9%
H2AK5ac	56.2%-56.6%	47.7%-47.8%	64.8%-64.9%
H3K9me2	56.2%-56.6%	68.1%-68.2%	44.5%-44.6%
H3K79me3	56%-56.4%	50.2%-50.3%	62%-62.1%
H3K27me3	56%-56.4%	62.1%-62.2%	50%-50.1%
H3K27ac	55.3%-55.7%	54%-54.1%	56.8%-56.9%
H3K36ac	55.6%-55.9%	61.1%-61.2%	50.2%-50.3%
H3K18ac	55.1%-55.5%	40.1%-40.2%	70.4%-70.5%
H3K4ac	55.1%-55.5%	48.3%-48.4%	62.2%-62.3%
H2BK120ac	55.1%-55.4%	38.2%-38.3%	72.2%-72.3%
H2BK12ac	55%-55.4%	65%-65.1%	45.2%-45.3%
H2BK5ac	54.6%-54.9%	47.3%-47.4%	62.1%-62.2%
H3K9ac	54%-54.4%	61.2%-61.3%	47%-47.1%
H3K79me2	53.8%-54.2%	52.7%-52.8%	55.1%-55.2%
H3K9me3	53.6%-54%	68.4%-68.5%	39%-39.1%
H3K14ac	53.5%-53.9%	65%-65.1%	42.1%-42.2%
H3K23ac	53.4%-53.8%	65.8%-65.9%	41.2%-41.3%
H2AZ	52.9%-53.3%	40%-40.1%	66%-66.1%
H2AK9ac	52.9%-53.3%	54%-54.1%	52%-52.1%

* 95% confidence interval