Supplementary Tables

Supplementary Table 1: PCR primers

Gene	Region	Forward Sequence	Reverse Sequence	Chr	Position
Rtn4	Promoter	CCATAGCACTGGAACTGCCTA	TACCAGCACCAAACAAAGCA	11	29590649
Rtn4	TSS	TTCATTGTTTGTCGGGGTTT	GCTGATAGGCGAGAACTTCG	11	29593867
Rtn4	Body	TGATGGGAAACTTGGGATGT	CCCACGTTGGCAGACTATTT	11	29634864
Fnip1	Promoter	CATGGACAGCCTCTTGTCCT	GTGTGTGAGCTGCTGACGTT	11	54249144
Fnip1	TSS	GGTCTGTCAGGCTTCTTGGA	TTTCGCATTTCTCCTTCGAT	11	54252199
Fnip1	Body	GGGAGGATCAGAAAGGAGGA	CCCTTGGCTCAACACTGATT	11	54269013
Dnmt3a	Promoter	CAAAGACGCAGTGTTGGTGT	ACAGCAACTGTCAGGGGACT	12	3804752
Dnmt3a	TSS	TCCCCTTGCATACACTCACA	TCGTCCTTCACTGTGCAGAT	12	3808067
Dnmt3a	Body	CTTGCTGGTCCTGGAGAAAG	CCCAAAGGAACTGGAAACAA	12	3874444
Scube2	TSS	GCCAAAGCATCTCCAGAGTC	GTGGGCACGGAACTGTTACT	7	117008446
Scube2	Body	CTCGAACCTCCTCCTGTGAG	TCAATTCCATGGTGTGCTGT	7	116967705
Pde8a	TSS	CATCCACACTTCCGAAAACC	GGTCGACCATCGCTGAATAG	7	88358817
Pde8a	Body	GTGGCCGATCTTGTCTCATT	TGAGGAAAGCTTGTGTGTGG	7	88445832
HOXA11		AGGAGAAGGGGTTCCTTCAA	CTCCGCGGTTTGTCAATAAT	6	52196629
B-Actin		CCCCAACACACCTAGCAAAT	ACTGCCCCATTCAATGTCTC	5	143668974
Ink4A	Promoter	GATGGAGCCCGGACTACAGAAG	CTGTTTCAACGCCCAGCTCTC	4	88928272

The sequence and location of PCR primers used for qPCR validation experiments.

Supplementary Table 2: Sequencing yield

Cell Type	ChIP Type	Data Source	# Raw reads	# Uniquely mapped reads (% raw)	# Reads minus duplicates (% raw)
ES	H3K27me3	WEHI	8,791,361	8,214,590 (93%)	7,676,635 (87%)
	H3K27me3	Mikkelsen	17,748,450	10,632,477 (60%)	9,986,124 (56%)
	H3K36me3	Mikkelsen	19,815,298	10,658,308 (54%)	10,026,039 (51%)
	H3K4me3	Mikkelsen	26,362,119	16,929,883 (64%)	15,394,286 (58%)
	RNA-pol II	WEHI	8,900,733	8,218,207 (92%)	7,979,826 (90%)
	Input (WCE)	WEHI	9,215,520	8,761,014 (95%)	7,422,222 (81%)
	Input (WCE)	Mikkelsen	2,760,270	1,171,324 (42%)	1,092,475 (40%)
G1ME	H3K27me3	WEHI	8,869,320	7,900,873 (89%)	7,529,200 (85%)
	RNA-pol II	WEHI	9,505,838	6,748,122 (71%)	6,655,513 (70%)
	Input (WCE)	WEHI	9,505,094	8,927,160 (94%)	7,511,561 (79%)
MEF	H3K27me3	Mikkelsen	29,508,457	17,554,332 (59%)	16,190,897 (55%)
	H3K36me3	Mikkelsen	23,643,415	12,603,553 (53%)	11,754,386 (50%)
	H3K4me3	Mikkelsen	34,040,432	17,277,190 (51%)	14,829,144 (44%)
	Input (WCE)	Mikkelsen	17,411,296	7,743,641 (44%)	6,805,387 (39%)
NP	H3K27me3	Mikkelsen	23,603,819	12,167,746 (52%)	11,348,078 (48%)
	H3K36me3	Mikkelsen	23,580,211	9,881,284 (42%)	9,147,401 (39%)
	H3K4me3	Mikkelsen	23,779,905	9,764,415 (41%)	8,629,209 (36%)
	Input (WCE)	Mikkelsen	10,256,754	6,898,325 (67%)	6,371,653 (62%)

The number of reads produced by each lane of data and the corresponding number of reads available after mapping and removing duplicate reads.

Supplementary Figures

Supplementary Figure 1: Classification criteria

(A)

3000bp   1100bp

| Promoter | TSS | Broad | Stop |

(B)

Sum above average ~10%

Sum above average ~50%
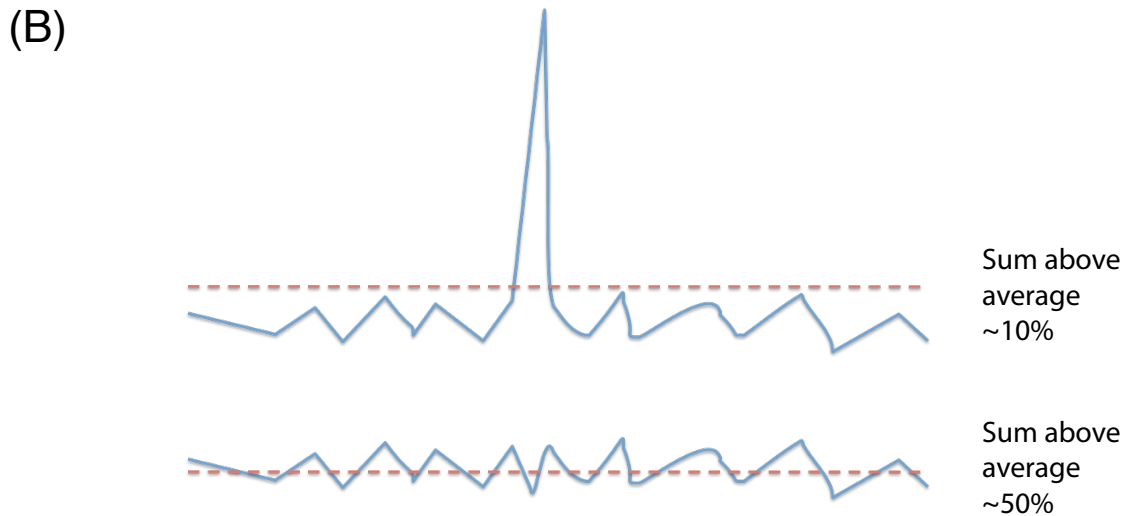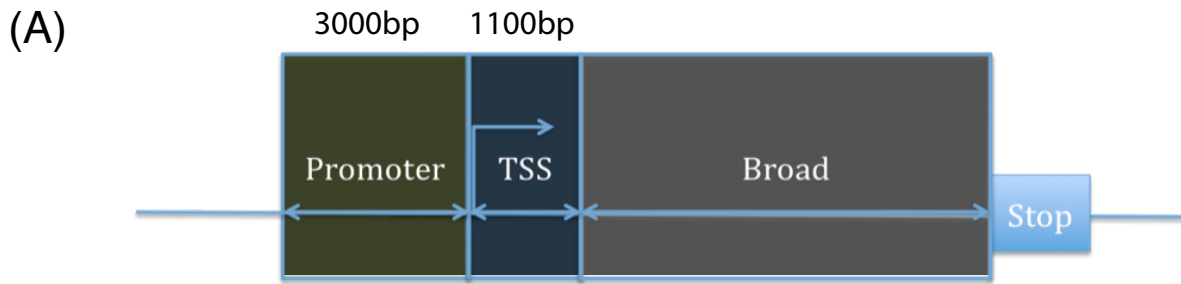
(A) A schematic illustration of how each gene is broken up when attempting to classify genes. The TSS region is anchored 100bp upstream of the TSS and is 1100bp in length. (B) A schematic illustrating the requirement that genes must have at least 35% of their ChIP-seq signal above the average to be classified as broadly marked. If a large peak exists somewhere in the signal, this will pull the average up above the surrounding signal, causing most of the signal to fall below the average. It is this effect that is exploited to avoid selecting any genes with a peak in the gene as being broadly bound.

Supplementary Figure 2: QC metrics of the Mikkelsen vs WEHI data in ES cells

Mikkelsen
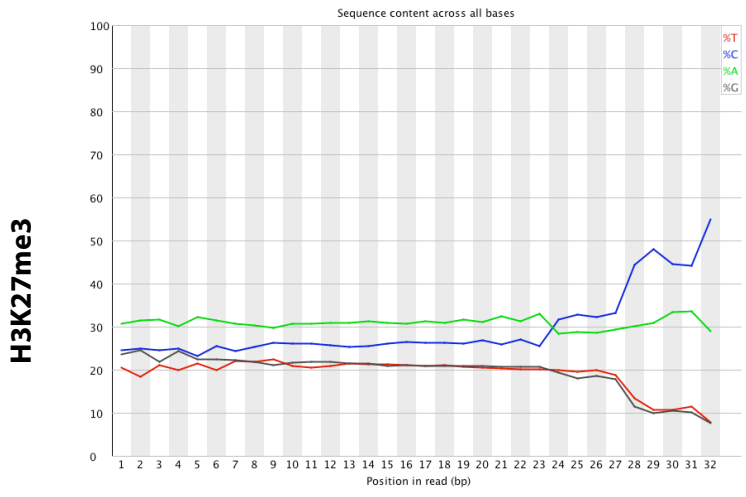
## WEHI
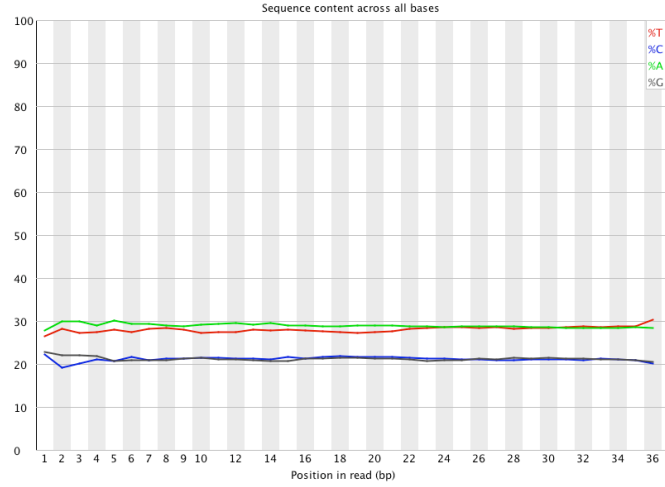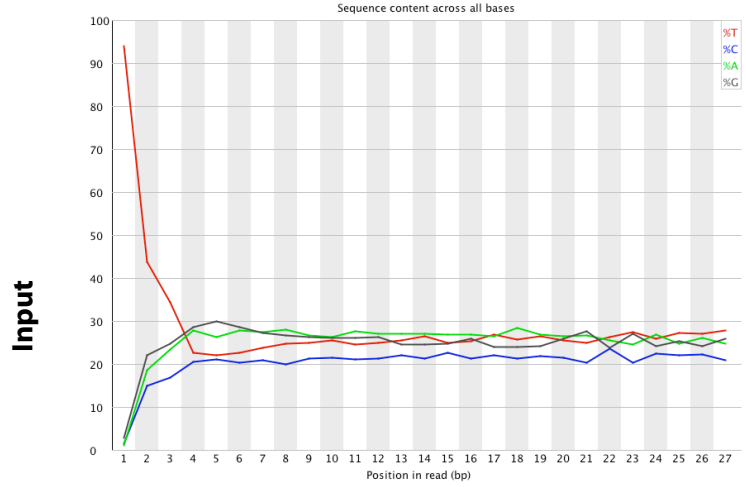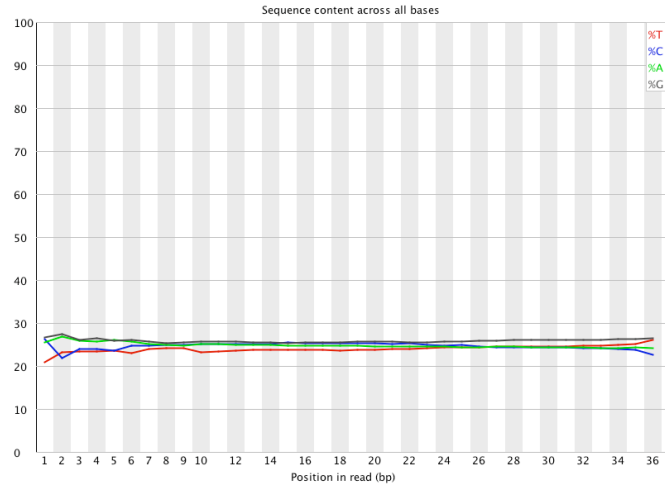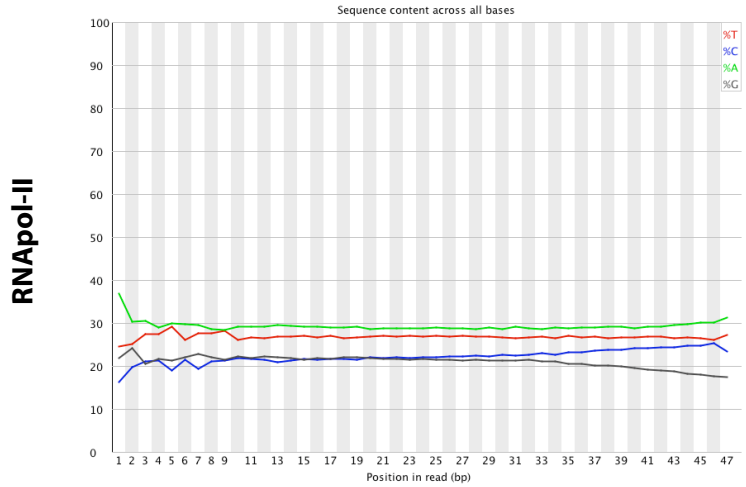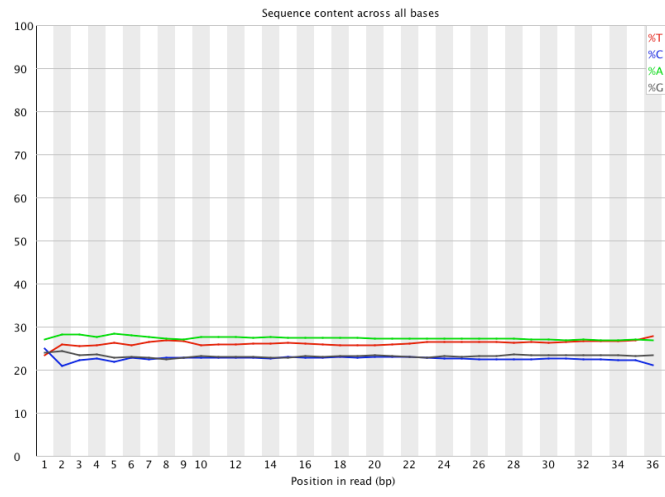
Quality Control comparison of the in house (WEHI) and Mikkelsen et al. (2007) data for Input, H3K27me3 and RNApol-II in ES cells. Each plot shows the fraction of reads that were A,C,G and T at each position along the read. If there was no bias in the experiment, the base composition would be the same across the entire length of the read.

Supplementary Figure 3: MACS selects sharp profiles while the Poisson test selects broad profiles

(A)

ES H3K27me3

(B)

G1ME H3K27me3
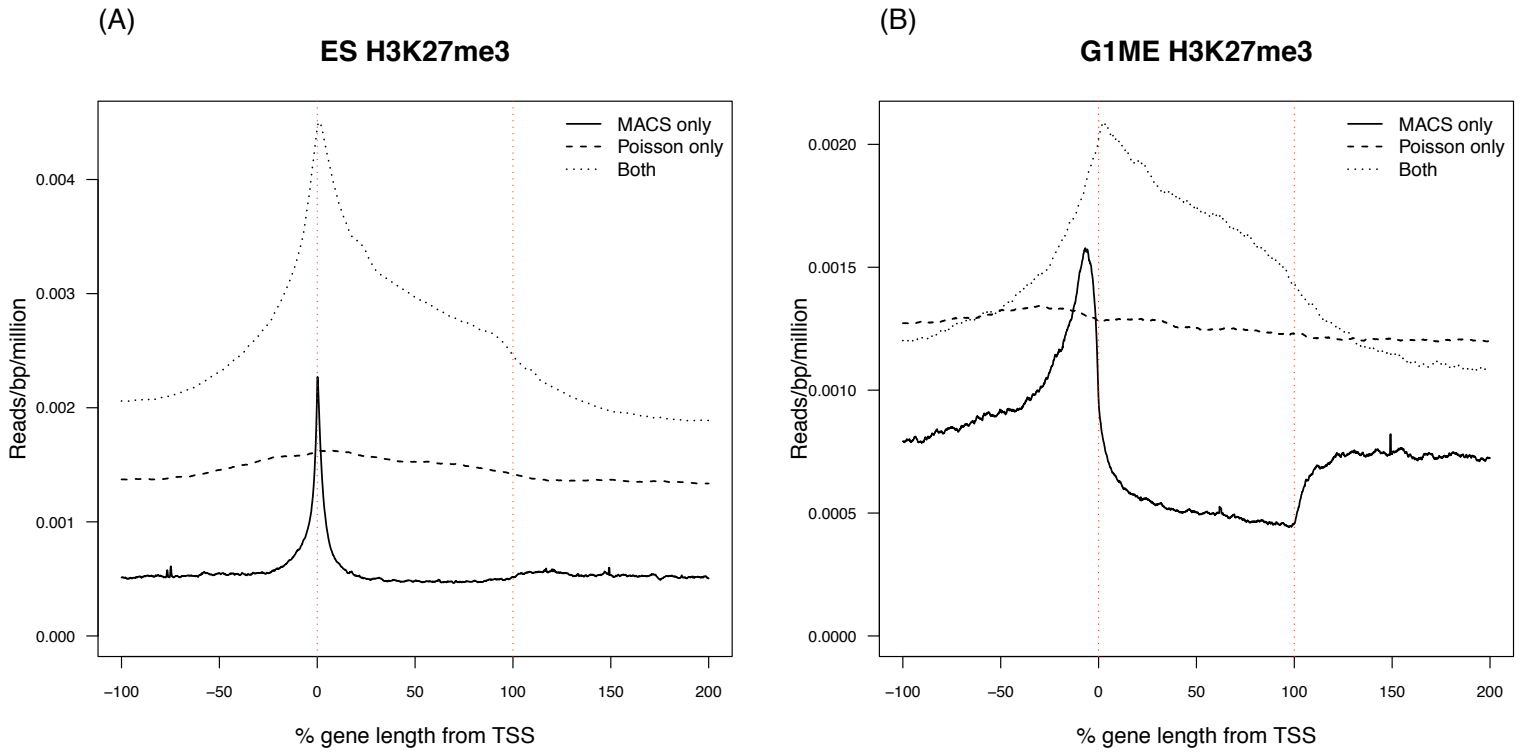
The ASE profiles of those genes only called marked by the MACS algorithm, the Poisson test or both methods in ES cells (A) & G1ME cells (B). The MACS exclusive genes tend to have a very sharp peak, whereas the Poisson test tends to find genes with broader marking domains across the locus.

Supplementary Figure 4: TSS and Scaled plots of each set of bound genes
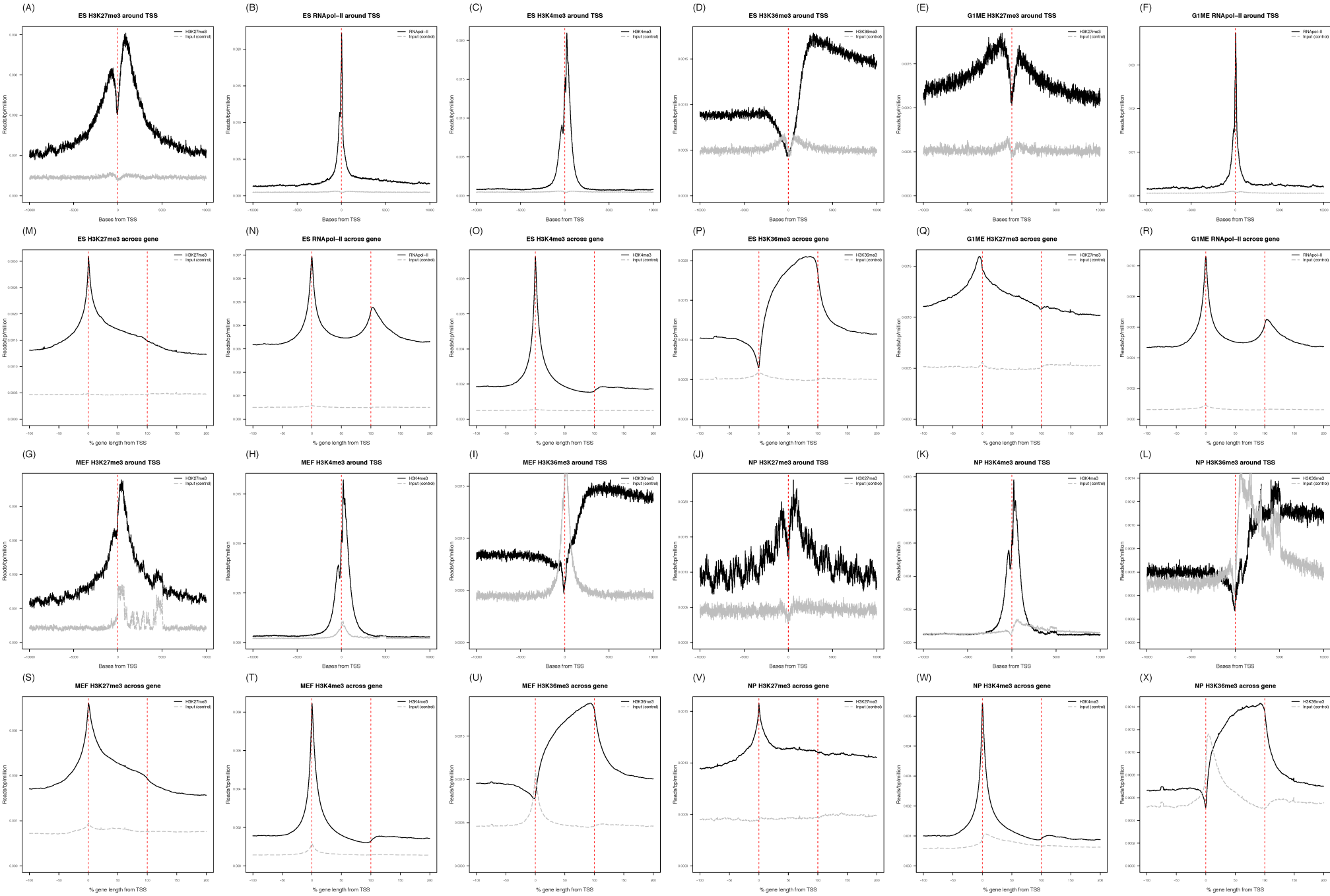
ASE and TSS plots for all of the cell types/ChIP assays for which we have data. In each case the Input control is shown as a separate track.

Supplementary Figure 5: PCR validation
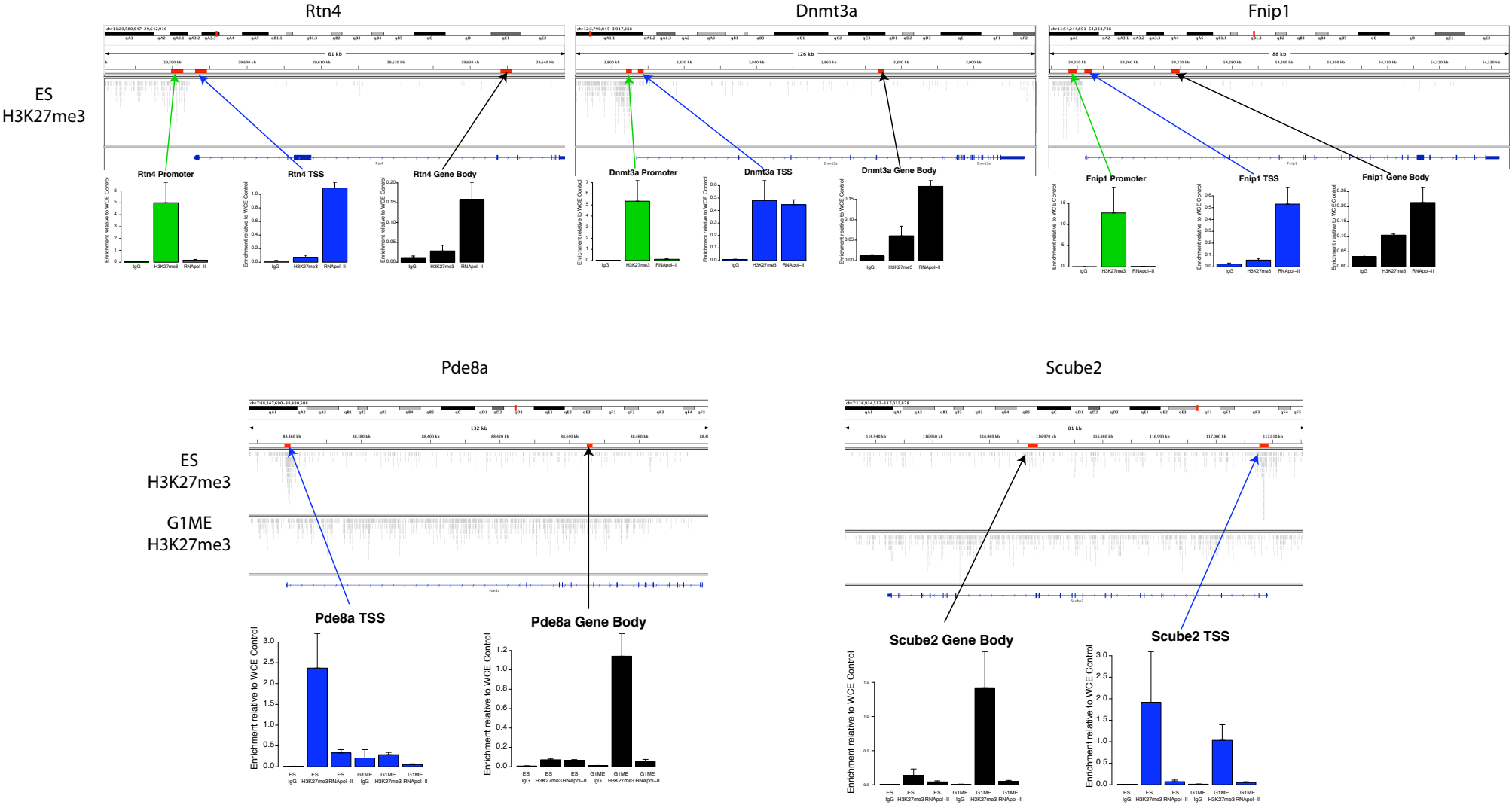
Each panel shows the ChIP-seq signal of H3K27me3 around a gene. For each gene PCR primers at the promoter region, TSS or the gene body are marked in red. PCR results are used to confirm the ChIP signal at each location, as well as measure the level of RNApol-II binding. The barplot for each primer shows the mean PCR signal relative to the whole cell extract, with error bars indicating standard error, appearing in the order IgG, H3K27me3 then RNApol-II, reading from left to right. The top row shows three genes classified as promoter in G1ME, while the bottom row shows two genes which are classified as TSS in ES cells and broad in G1MEs. Each PCR experiment was replicated (n=3 for Rtn4, Dnmt3a & Fnip1, n=4 for Pde8a and Scube2) and confirms the results of the ChIP-seq data.

Supplementary Figure 6: The Mikkelsen ES, MEF and NP data split into classes with corresponding expression data
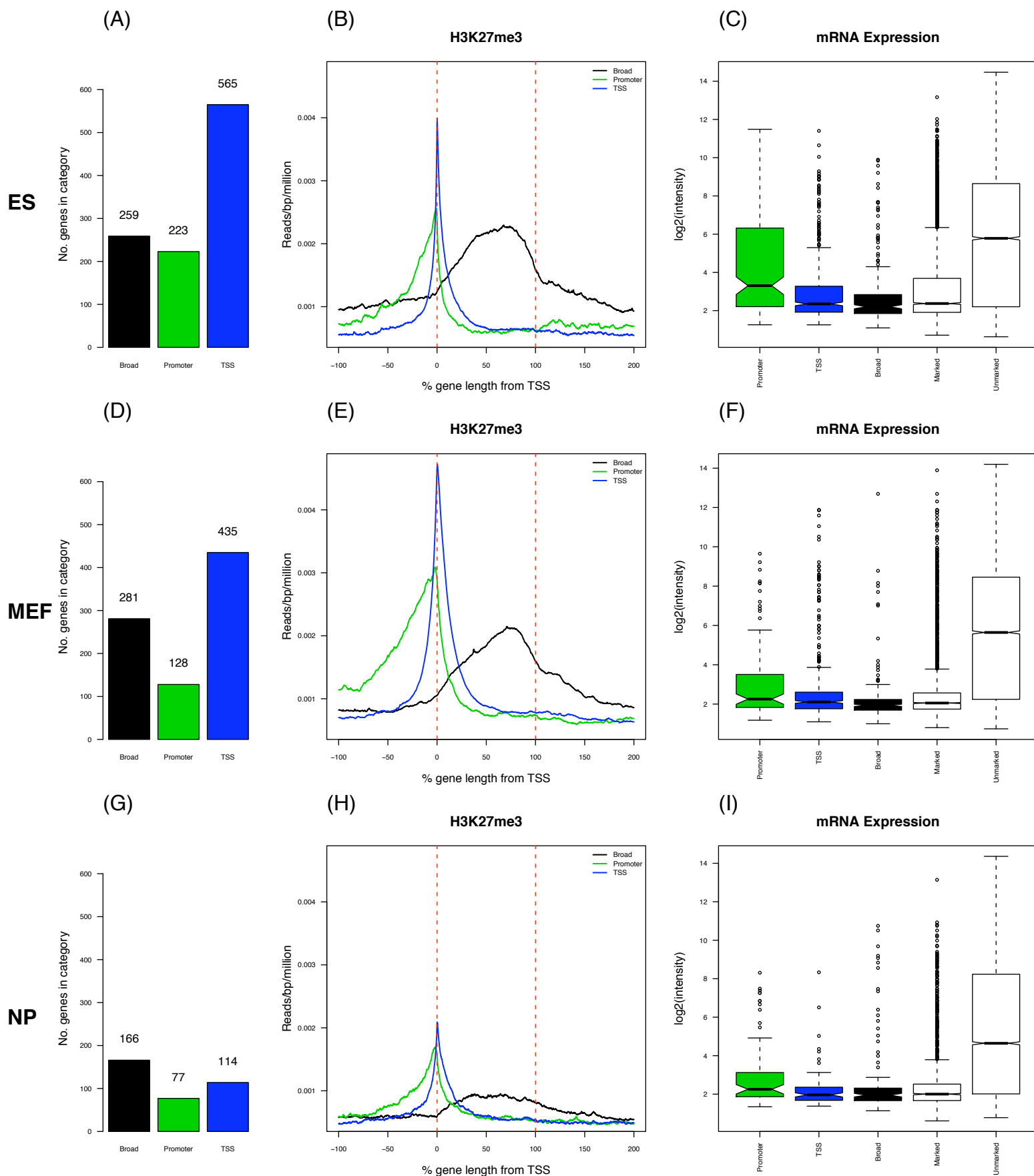
Barplots that show the number of genes called in each class in each cell type - (A) ES (D) MEF (G) NP cells. The averaged scaled signal plots for those genes called as promoter, TSS or broad in the Mikkelsen data for (B) ES, (E) MEF or (H) NP cells. Boxplots show the corresponding expression levels of genes in (C) ES, (F) MEF (I) NP cells. The three rows show data from ES, MEF and NP cells respectively.

Supplementary Figure 7: The genes with top/bottom 5% of expression show promoter/broad signal
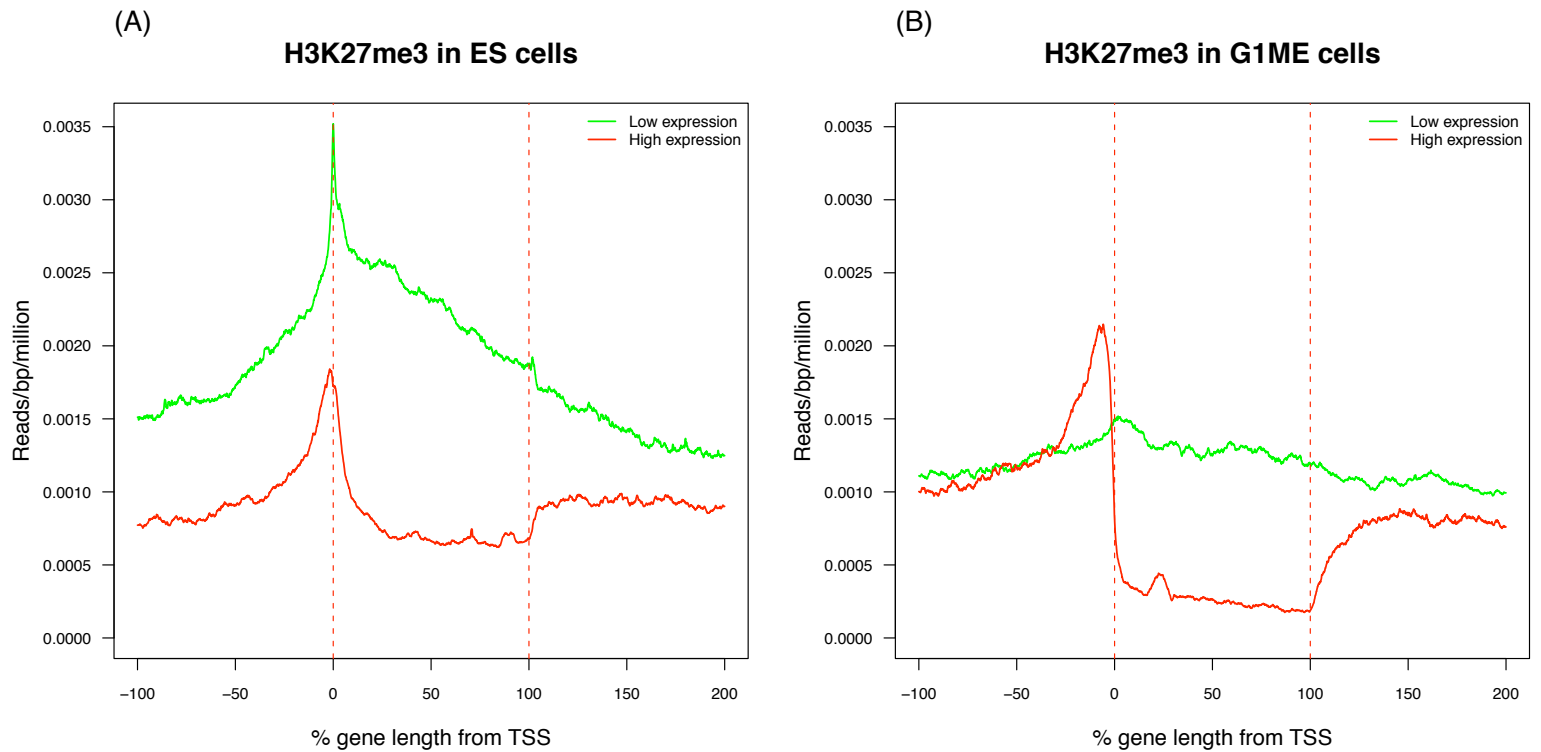
ASE plots of the H3K27Me3 signal for the most highly/lowly expressed (top/bottom 5%) genes also called as marked by H3K27me3 in ES cells (A) & G1ME cells (B). Although no classification or clustering was performed on these genes, we are still able to recover the promoter and broad binding profiles for each of the cell types.
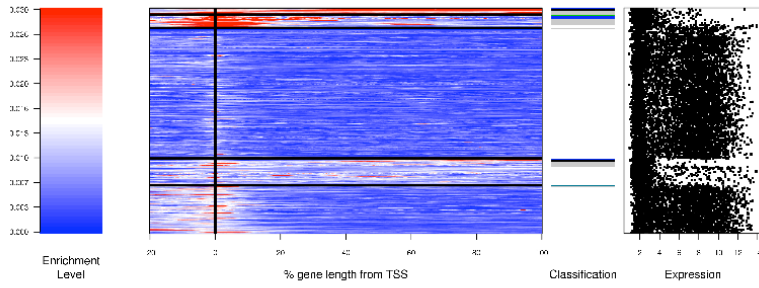
Supplementary Figure 8: K-means clustering plots for selected conditions

(A) Scaled clustering for ES H3K27Me3
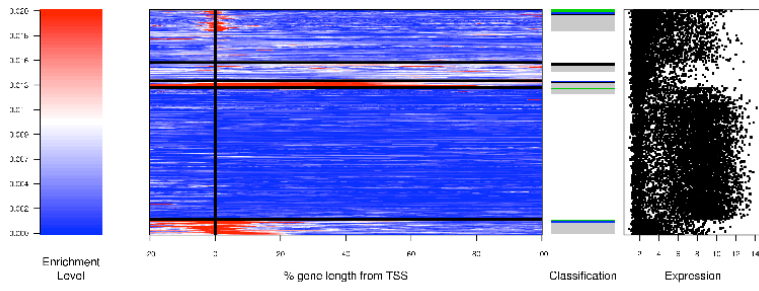
(B) Scaled clustering for G1ME H3K27Me3
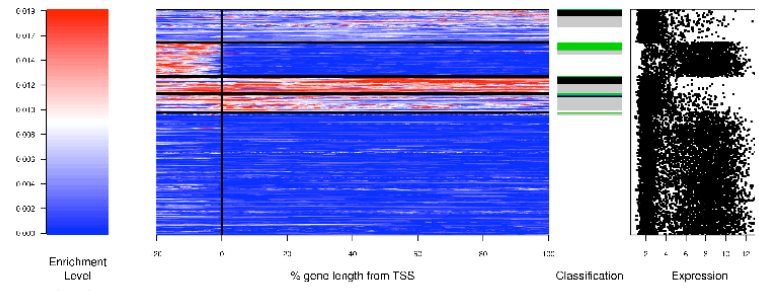
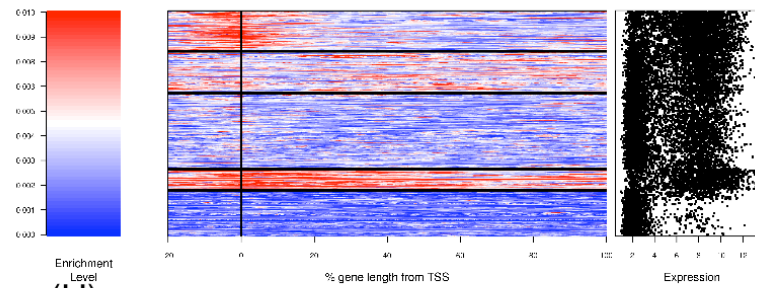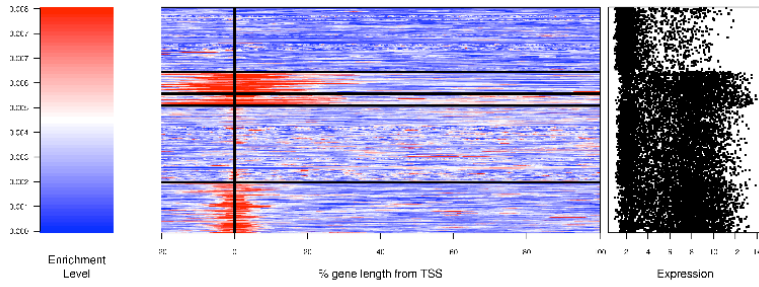(C) Scaled clustering for MEF H3K27Me3

(D) Scaled clustering for NP H3K27Me3
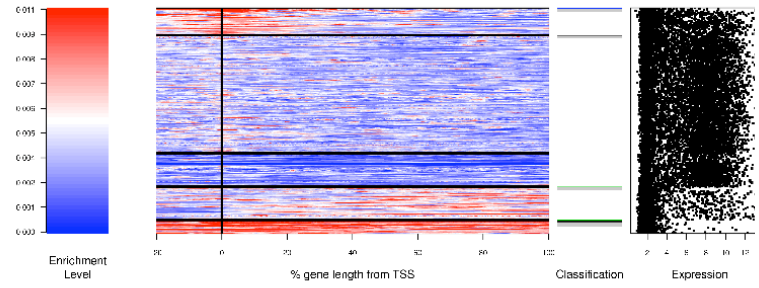
(E) Scaled clustering for ES INPUT
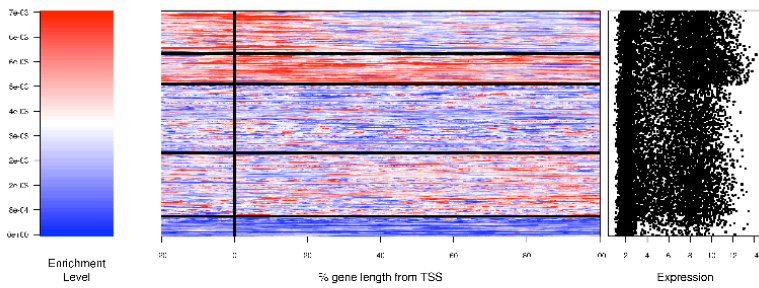
(F) Scaled clustering for G1ME INPUT

(G) Scaled clustering for MEF INPUT
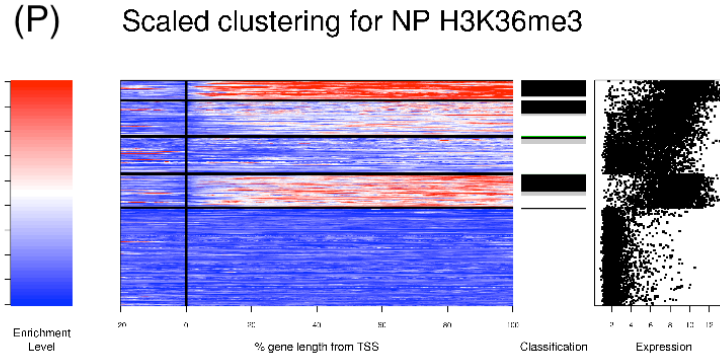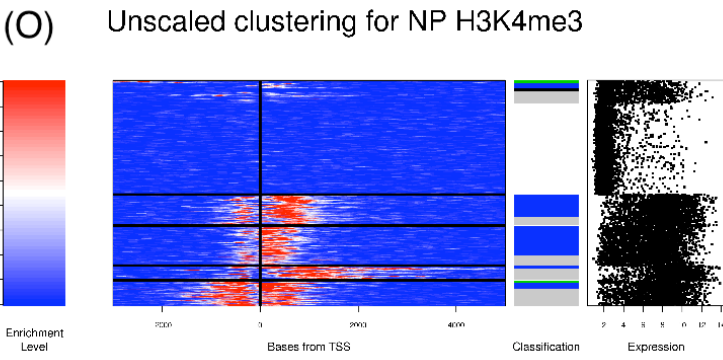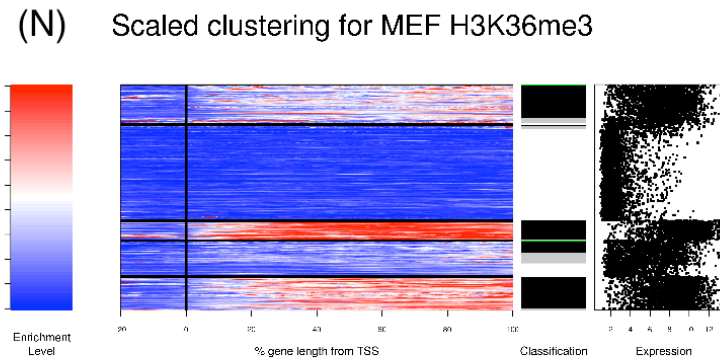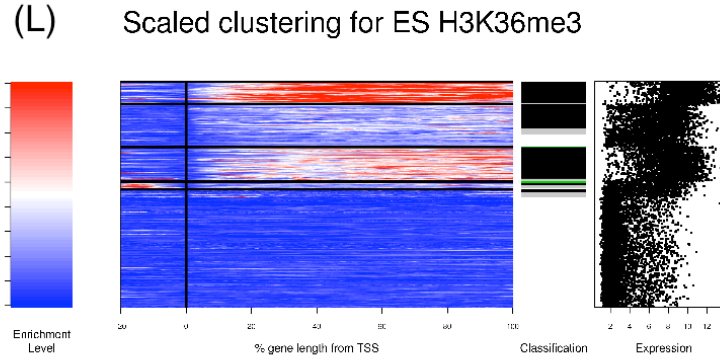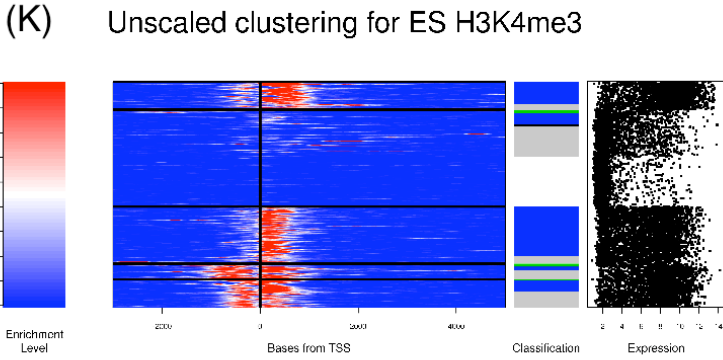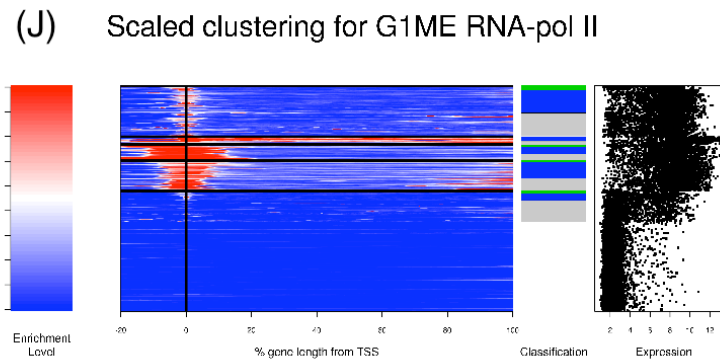
(H) Scaled clustering for NP INPUT

(I) Scaled clustering for ES RNA-pol II

(J) Scaled clustering for G1ME RNA-pol II

(K) Unscaled clustering for ES H3K4me3

(L) Scaled clustering for ES H3K36me3

(M) Unscaled clustering for MEF H3K4me3

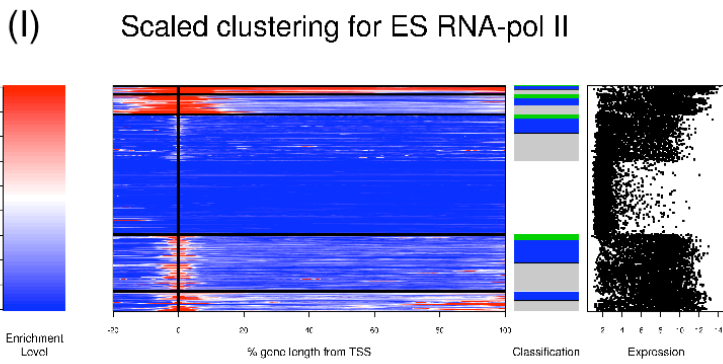**(N)** Scaled clustering for MEF H3K36me3

(O) Unscaled clustering for NP H3K4me3

**(P)** Scaled clustering for NP H3K36me3
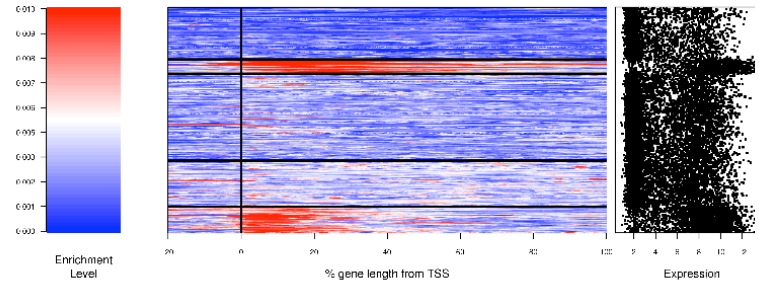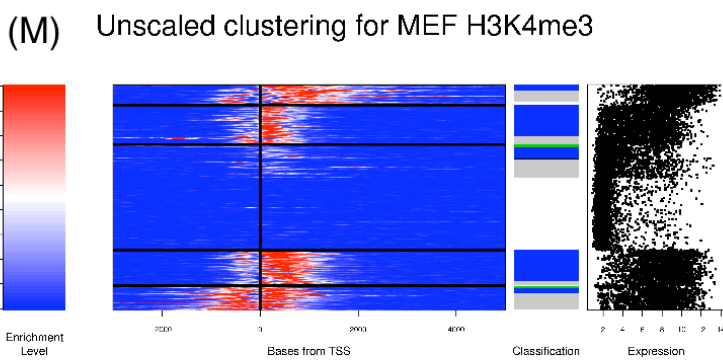
K-means cluster for a selection of cell types and histone marks. In each example the main panel shows a heatmap of the signal for all genes in the genome where genes are either scaled to the same length (K27,K36 RNApol-II & Input) or 5kb around the TSS (K4). Red is the strongest marking and blue is the weakest. The Gene status track shows the classification calls made on each gene, where blue corresponds to TSS genes, green to promoter genes, black to broad genes, grey to marked and unclassified and white to unmarked. The right most panel shows the corresponding mRNA expression level of each gene as measured on the relevant microarray.

**Supplementary Figure 9: G1ME classes plotted in other cell types**
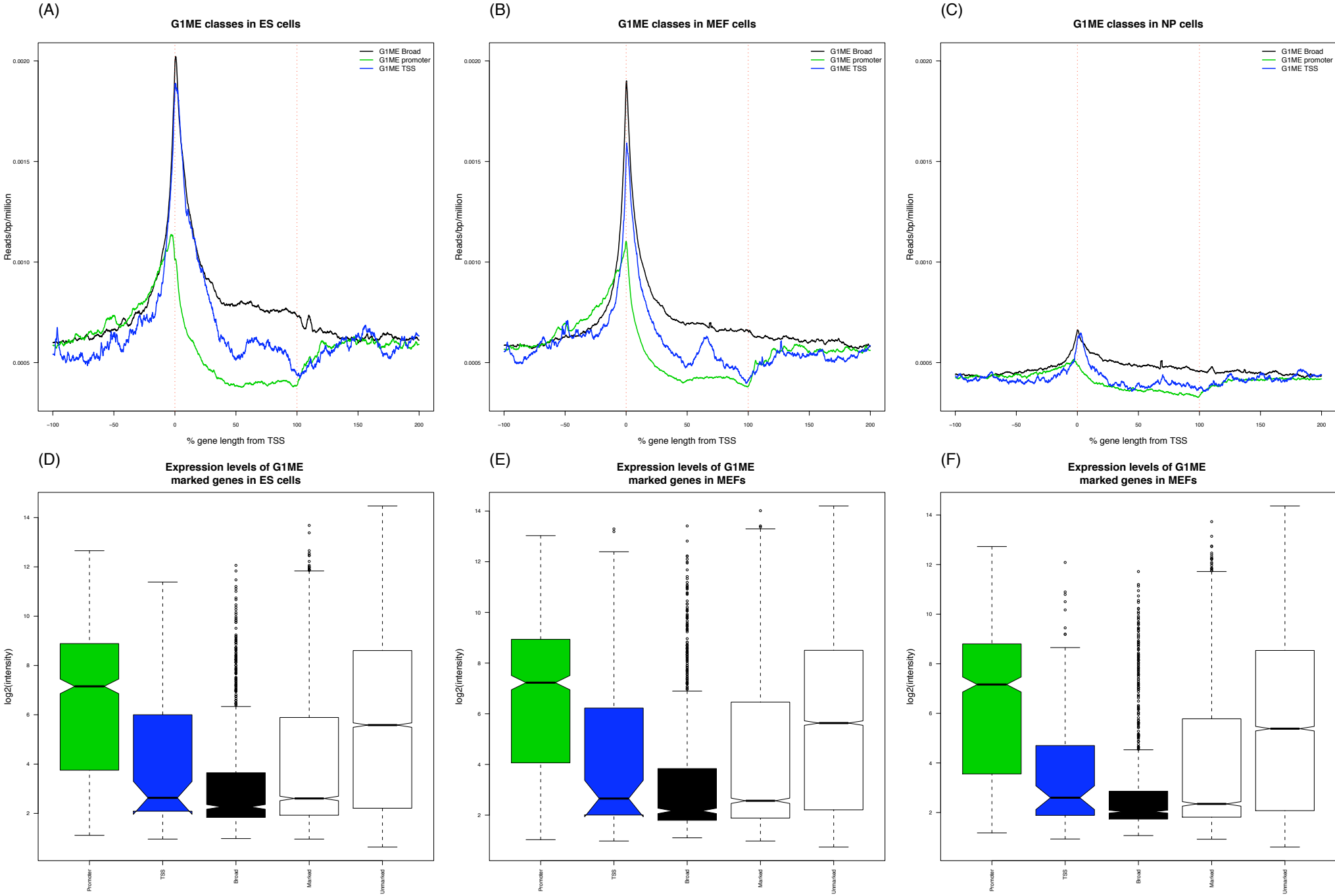
The H3K27me3 signal for those genes called as promoter, broad or TSS in G1ME cells, plotted in ES cells (A), MEFs (B) & NPs (C). The corresponding mRNA expression levels for the same sets of genes are also shown in ES cells (D), MEFs (E) and NPs (F). The genes identified as promoter genes in G1ME have a consistent enrichment for H3K27Me3 in the other two cell types and also have consistently higher expression.

## Supplementary Figure 10: Perturbation of PRC2
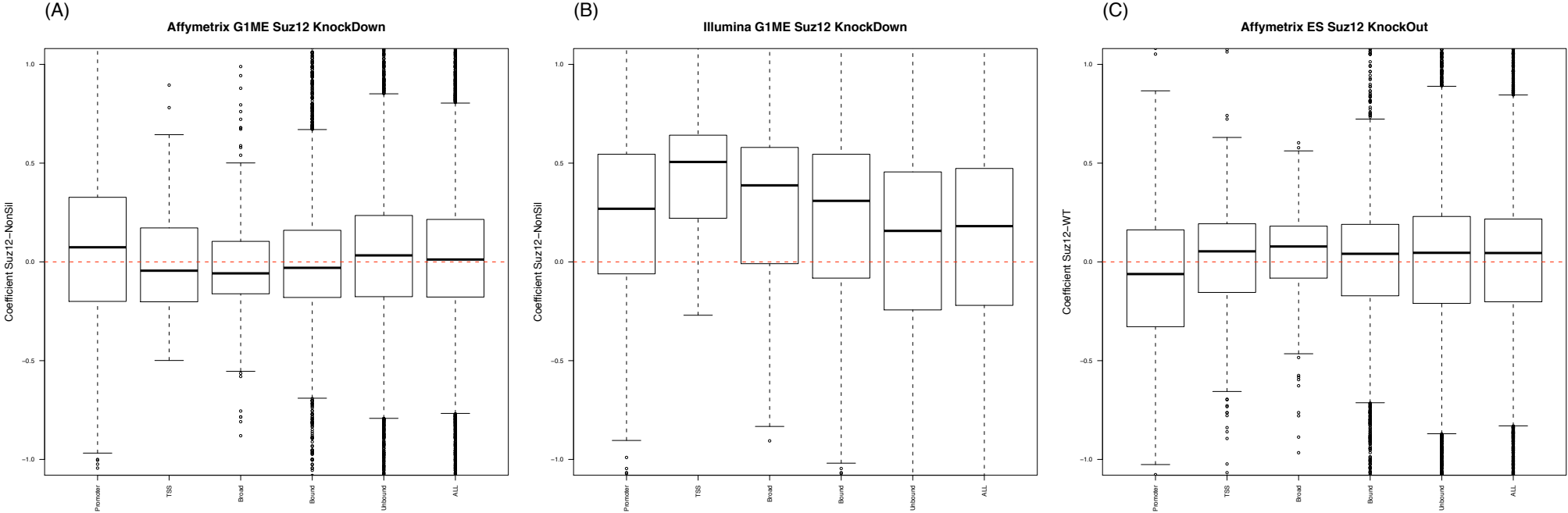
The fold changes resulting from knocking down or knocking out Suz12 in ES or G1ME cells for the promoter, TSS and broad classes of genes. Panels (A) and (B) have the same experimental setup, but use a different microarray platform and were hybridized using RNA from separate batches of G1ME cells (Affymetrix GeneChip Mouse Genome 430 2.0 were used for panel A and Illumina's Mouse Whole Genome 6 v1.1 arrays for panel B). In general the fold changes between the two conditions were modest, as were the differences between different classes (TSS, broad, promoter). The expression level of the promoter class increased slightly in G1ME cells upon inhibition of PRC2, whereas it was slightly lower in ES cells. The modest changes in expression and the inconsistent result between cell types preclude us from drawing any conclusion regarding the functional role of H3K27me3 on these genes.

**Supplementary Figure 11: The flanking genes to the promoter have more marking than the broad ones.**

(A)

(B)
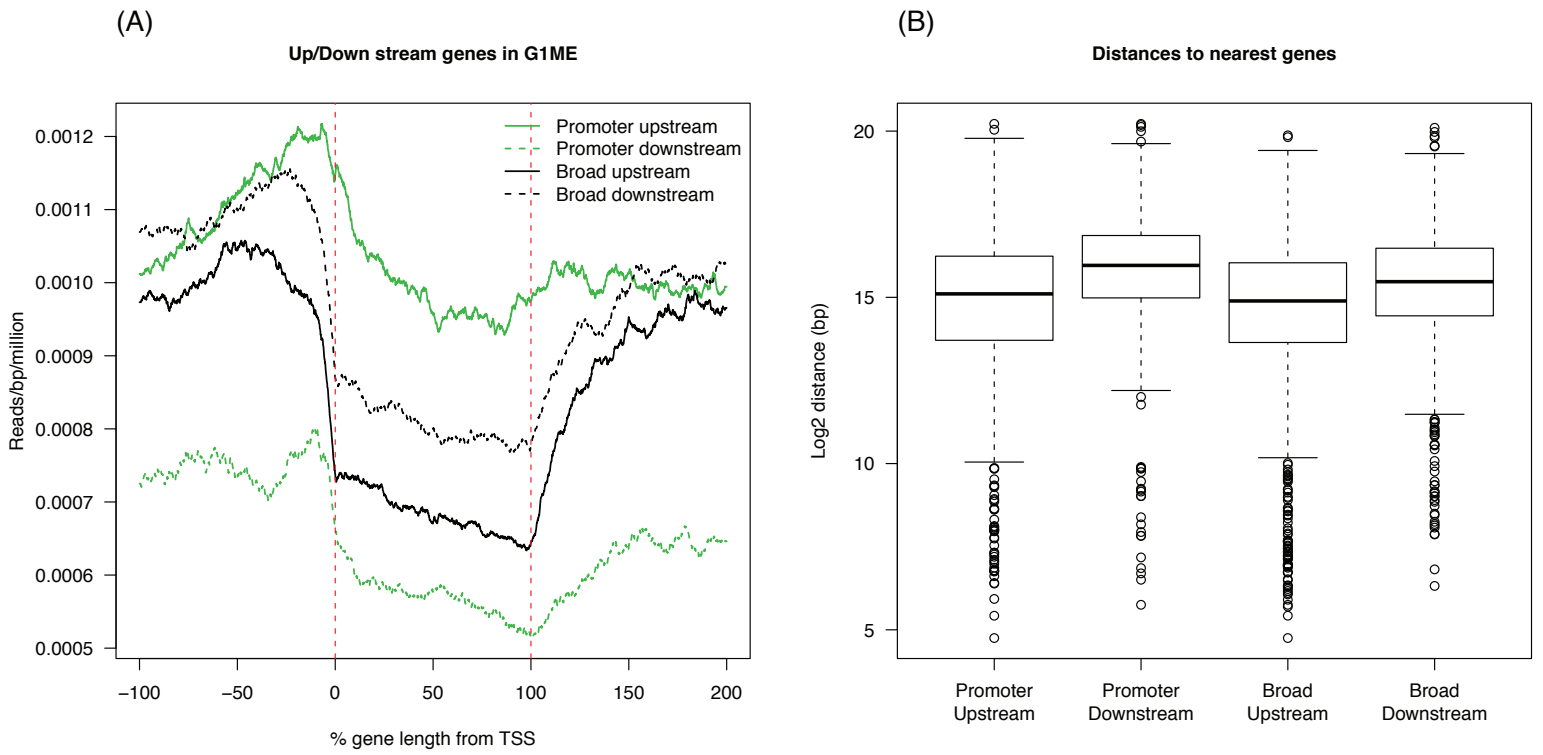
The ASE plot of H3K27me3 of the nearest, non-overlaping upstream/downstream neighbouring genes of the broad and promoter classes of genes in G1ME (A), as well as the distance to neighbouring genes (B).

Supplementary Figure 12: ASE plots stratified by expression and scaled
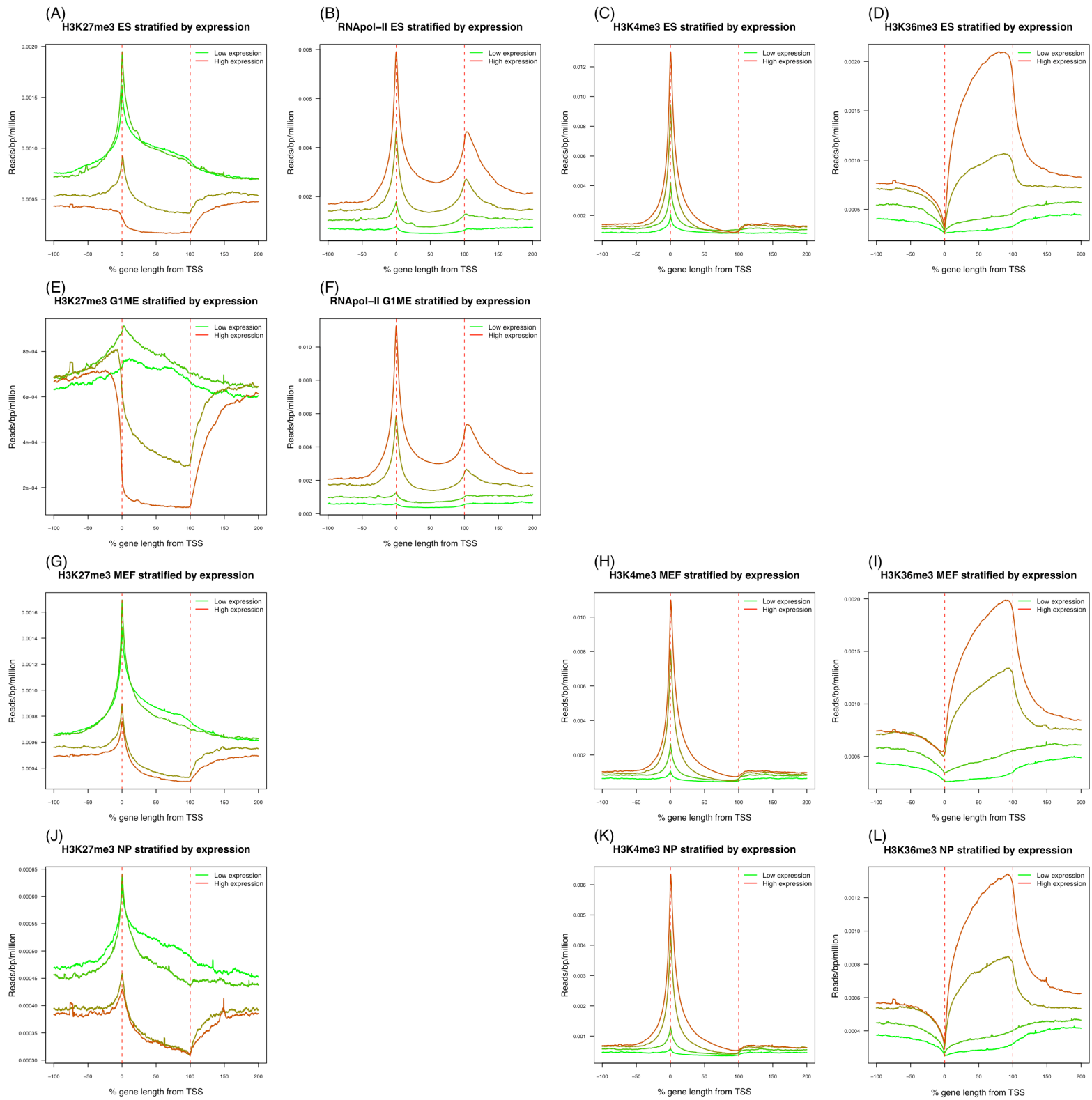
ASE plots of all genes in the genome stratified by mRNA expression level for different marks in ES cells (A-D), G1ME (E-F), MEF (G-I) and NP (J-L).

Supplementary Figure 13: TSS centered plots stratified by expression
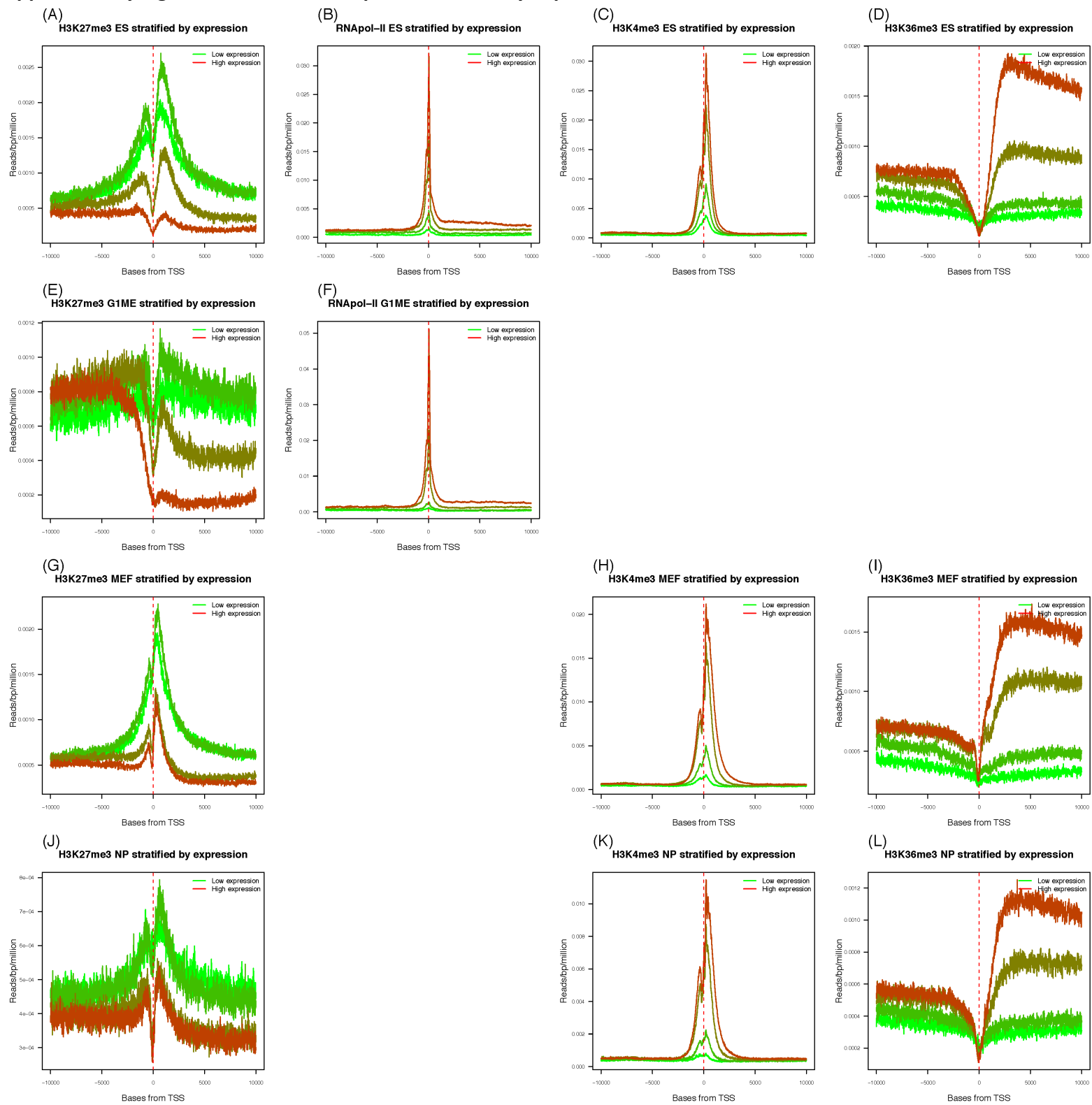
TSS centered plots of all genes in the genome stratified by mRNA expression level for different marks in ES cells (A-D), G1ME (E-F), MEF (G-I) and NP (J-L).