Figure S1

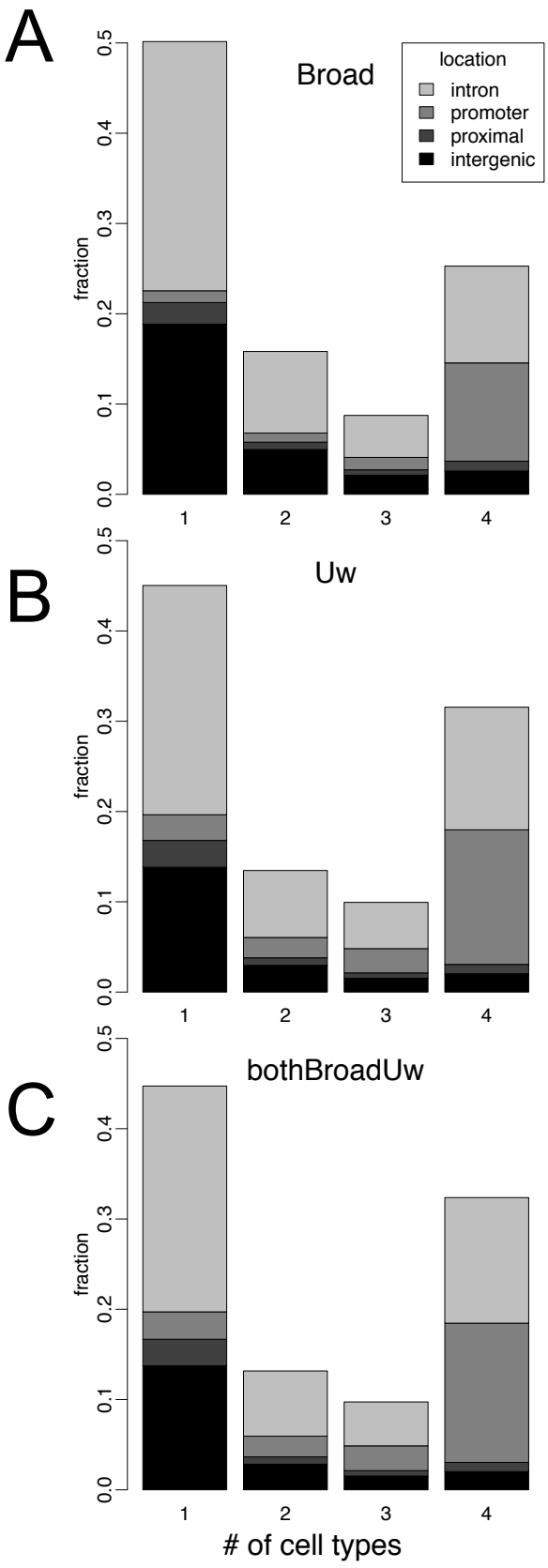

A
Broad
location
intron
promoter
proximal
intergenic
fraction
0.0
0.1
0.2
0.3
0.4
0.5
1
2
3
4
B
Uw
C
bothBroadUw
# of cell types

Figure S2

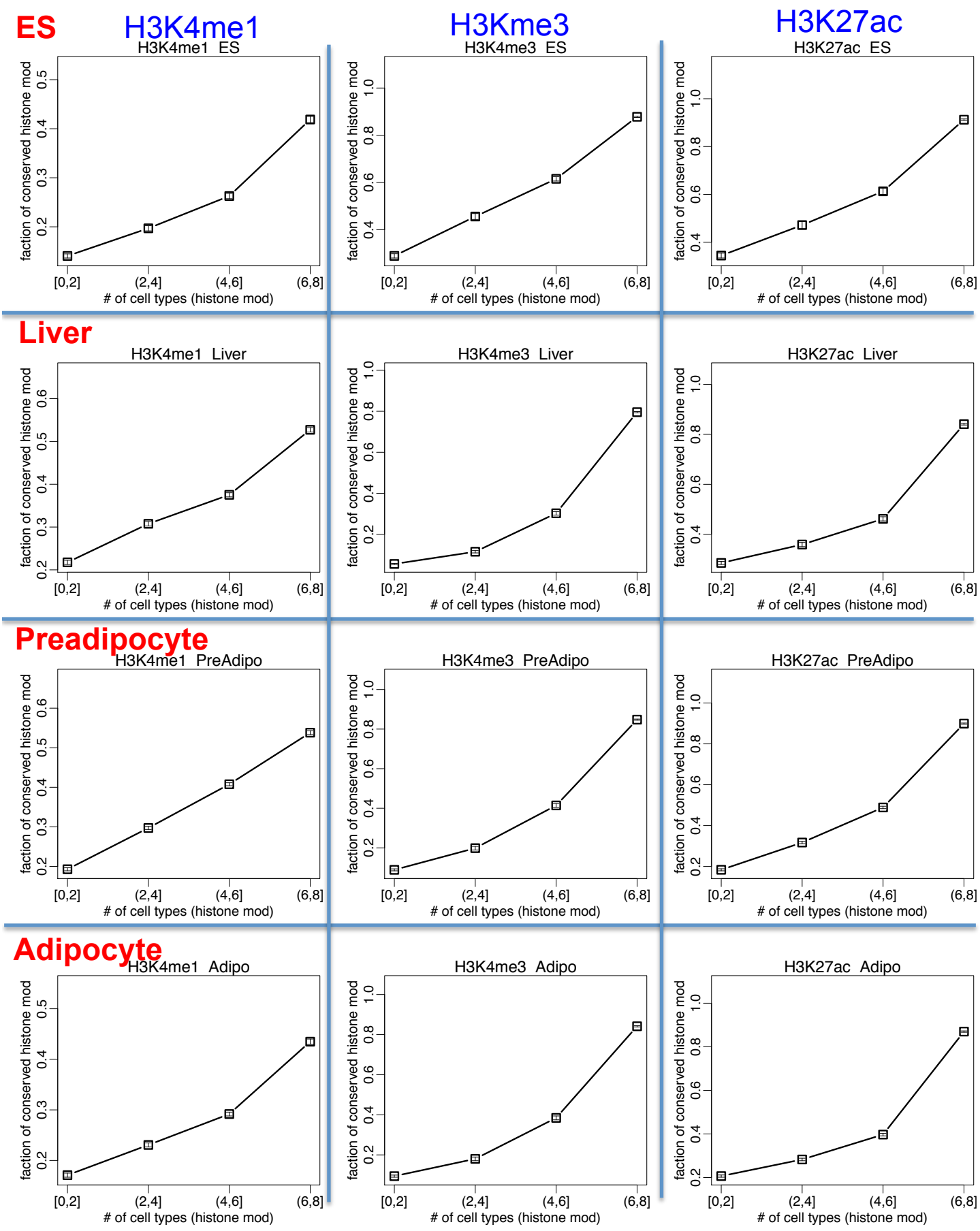

ES
H3K4me1
H3Kme3
H3K27ac
Liver
Preadipocyte
Adipocyte
H3K4me1 ES
H3K4me3 ES
H3K27ac ES
H3K4me1 Liver
H3K4me3 Liver
H3K27ac Liver
H3K4me1 PreAdipo
H3K4me3 PreAdipo
H3K27ac PreAdipo
H3K4me1 Adipo
H3K4me3 Adipo
H3K27ac Adipo
faction of conserved histone mod
# of cell types (histone mod)
[0,2]
(2,4]
(4,6]
(6,8]

Figure S3

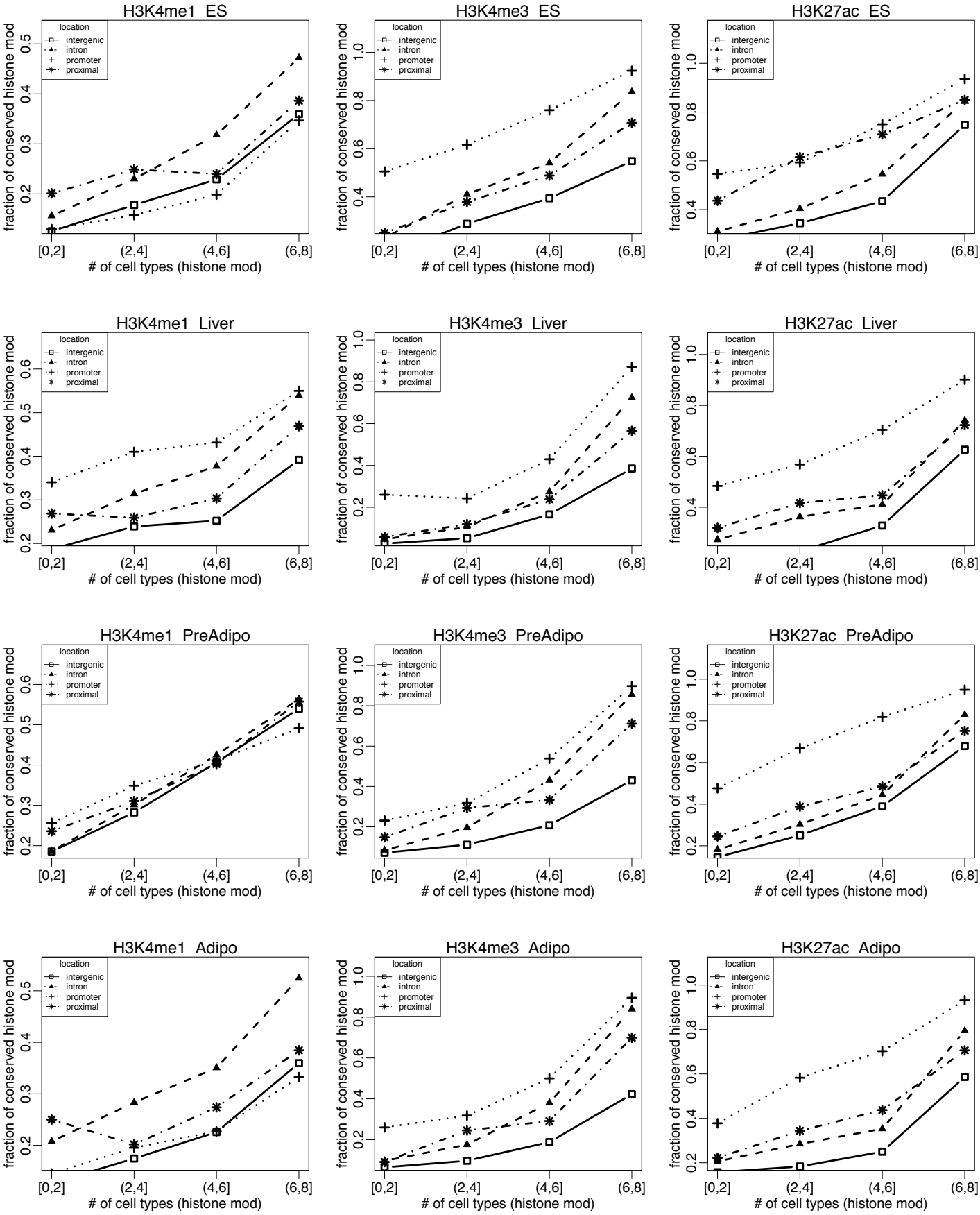

H3K4me1 ES
H3K4me3 ES
H3K27ac ES
H3K4me1 Liver
H3K4me3 Liver
H3K27ac Liver
H3K4me1 PreAdipo
H3K4me3 PreAdipo
H3K27ac PreAdipo
H3K4me1 Adipo
H3K4me3 Adipo
H3K27ac Adipo
location
intergenic
intron
promoter
proximal
fraction of conserved histone mod
# of cell types (histone mod)
[0,2]
(2,4]
(4,6]
(6,8]

Figure S4

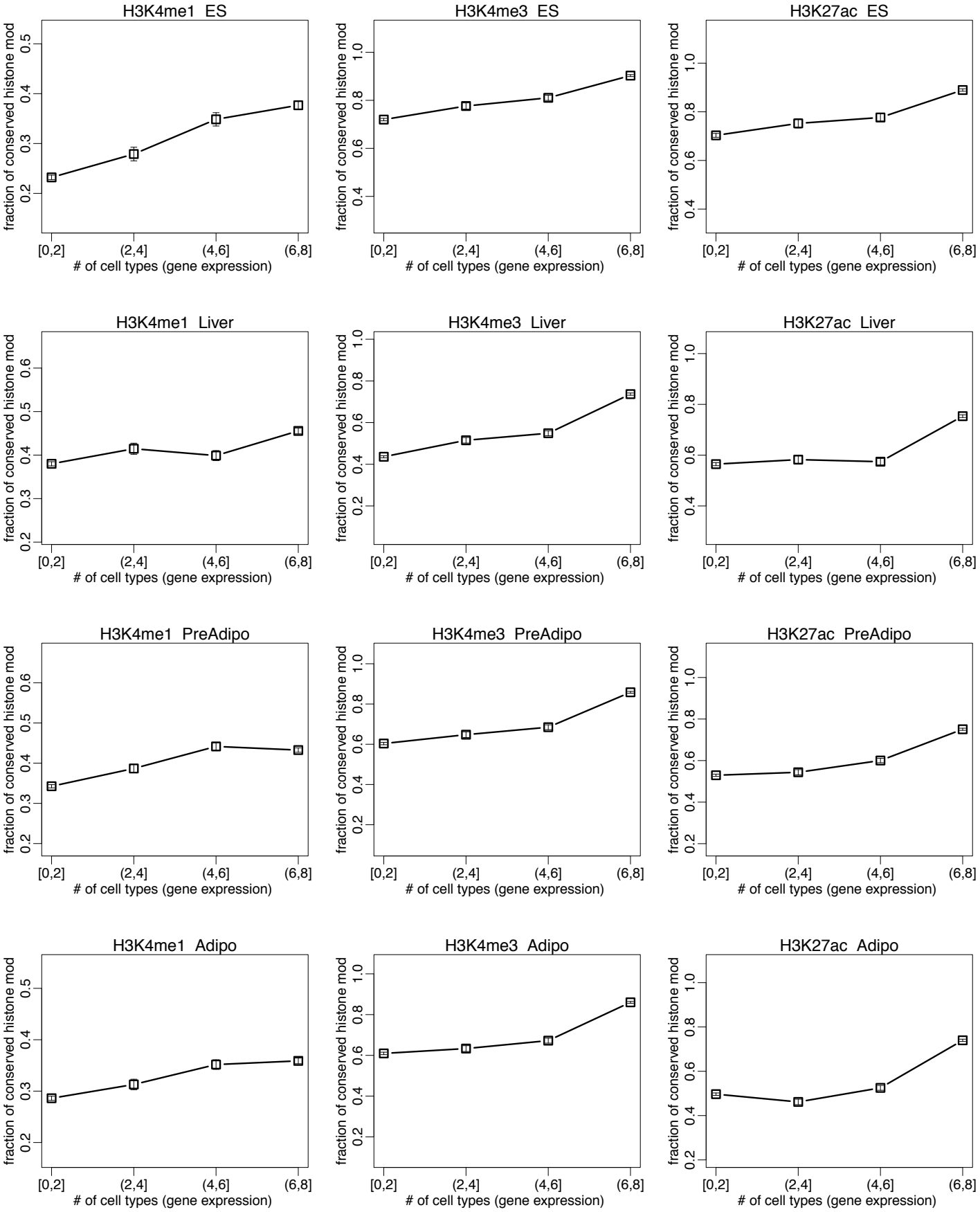

H3K4me1 ES
H3K4me3 ES
H3K27ac ES
H3K4me1 Liver
H3K4me3 Liver
H3K27ac Liver
H3K4me1 PreAdipo
H3K4me3 PreAdipo
H3K27ac PreAdipo
H3K4me1 Adipo
H3K4me3 Adipo
H3K27ac Adipo
fraction of conserved histone mod
[0,2]
(2,4]
(4,6]
(6,8]
# of cell types (gene expression)

Figure S5

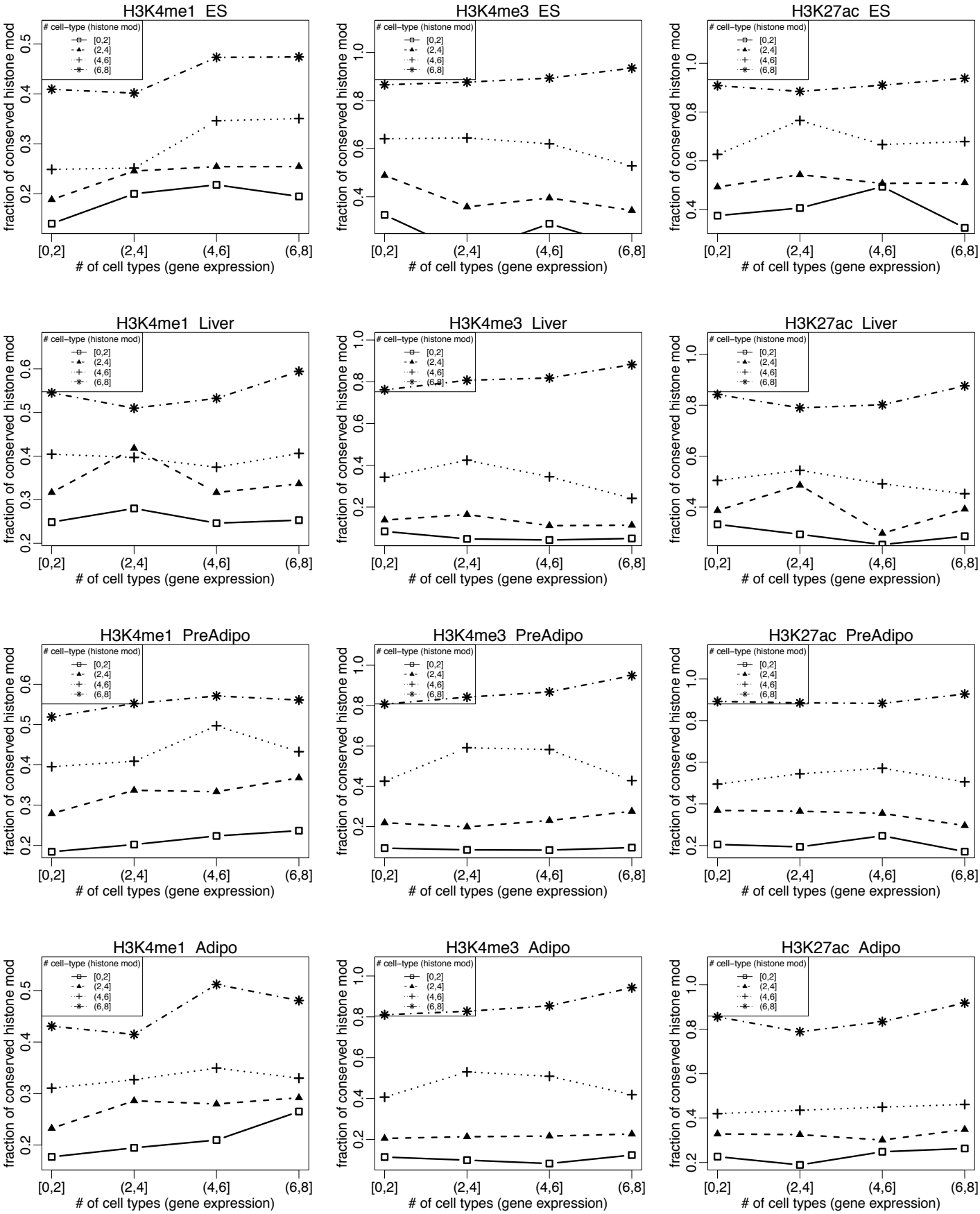

H3K4me1 ES
H3K4me3 ES
H3K27ac ES
H3K4me1 Liver
H3K4me3 Liver
H3K27ac Liver
H3K4me1 PreAdipo
H3K4me3 PreAdipo
H3K27ac PreAdipo
H3K4me1 Adipo
H3K4me3 Adipo
H3K27ac Adipo
cell−type (histone mod)
[0,2]
(2,4]
(4,6]
(6,8]
fraction of conserved histone mod
# of cell types (gene expression)

Figure S6

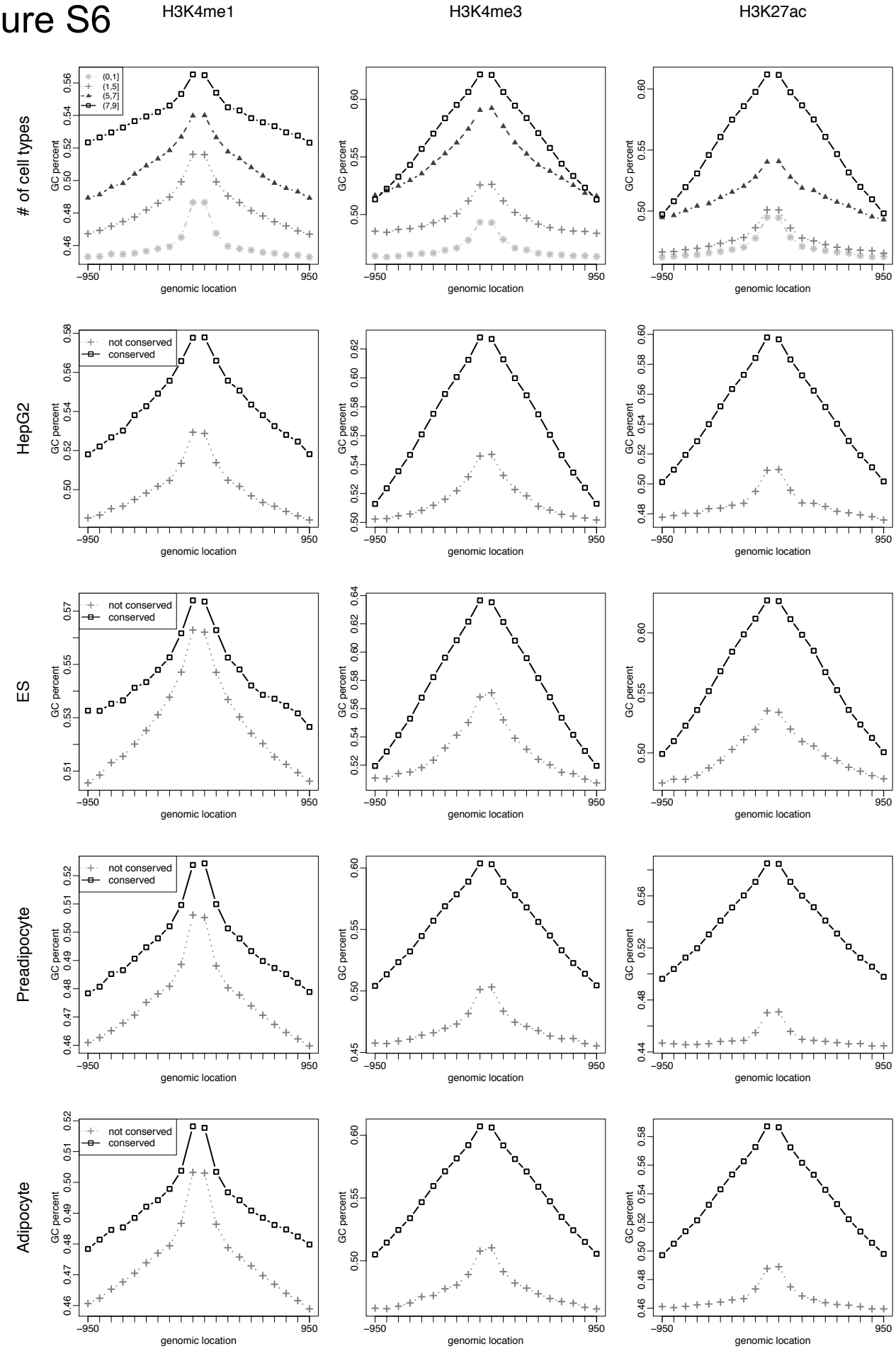

H3K4me1
H3K4me3
H3K27ac
# of cell types
HepG2
ES
Preadipocyte
Adipocyte
(0,1]
(1,5]
(5,7]
(7,9]
not conserved
conserved
GC percent
genomic location
−950
950

Figure S7

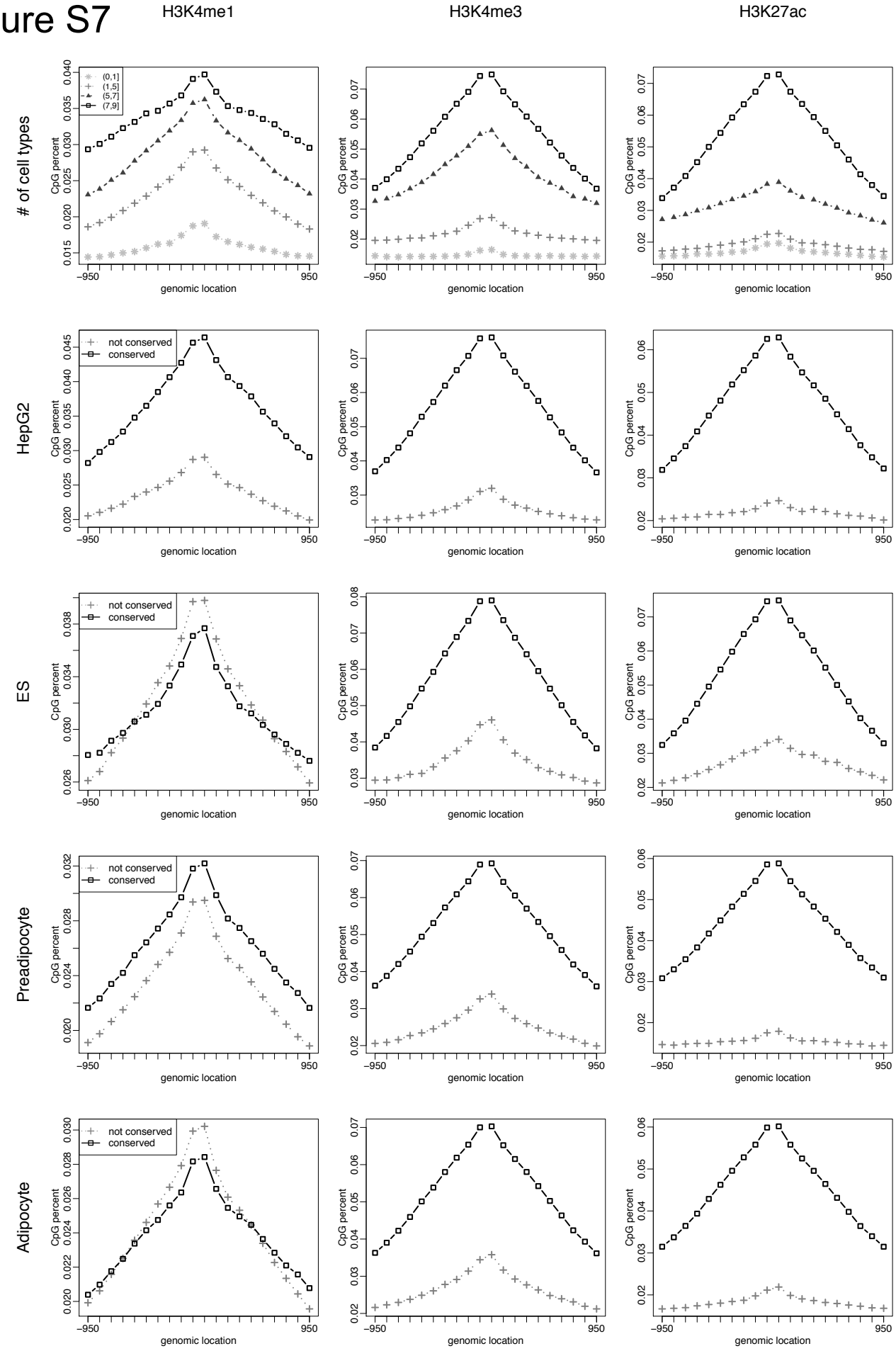

H3K4me1
H3K4me3
H3K27ac
# of cell types
HepG2
ES
Preadipocyte
Adipocyte
(0,1]
(1,5]
(5,7]
(7,9]
not conserved
conserved
CpG percent
genomic location
−950
950

Figure S8

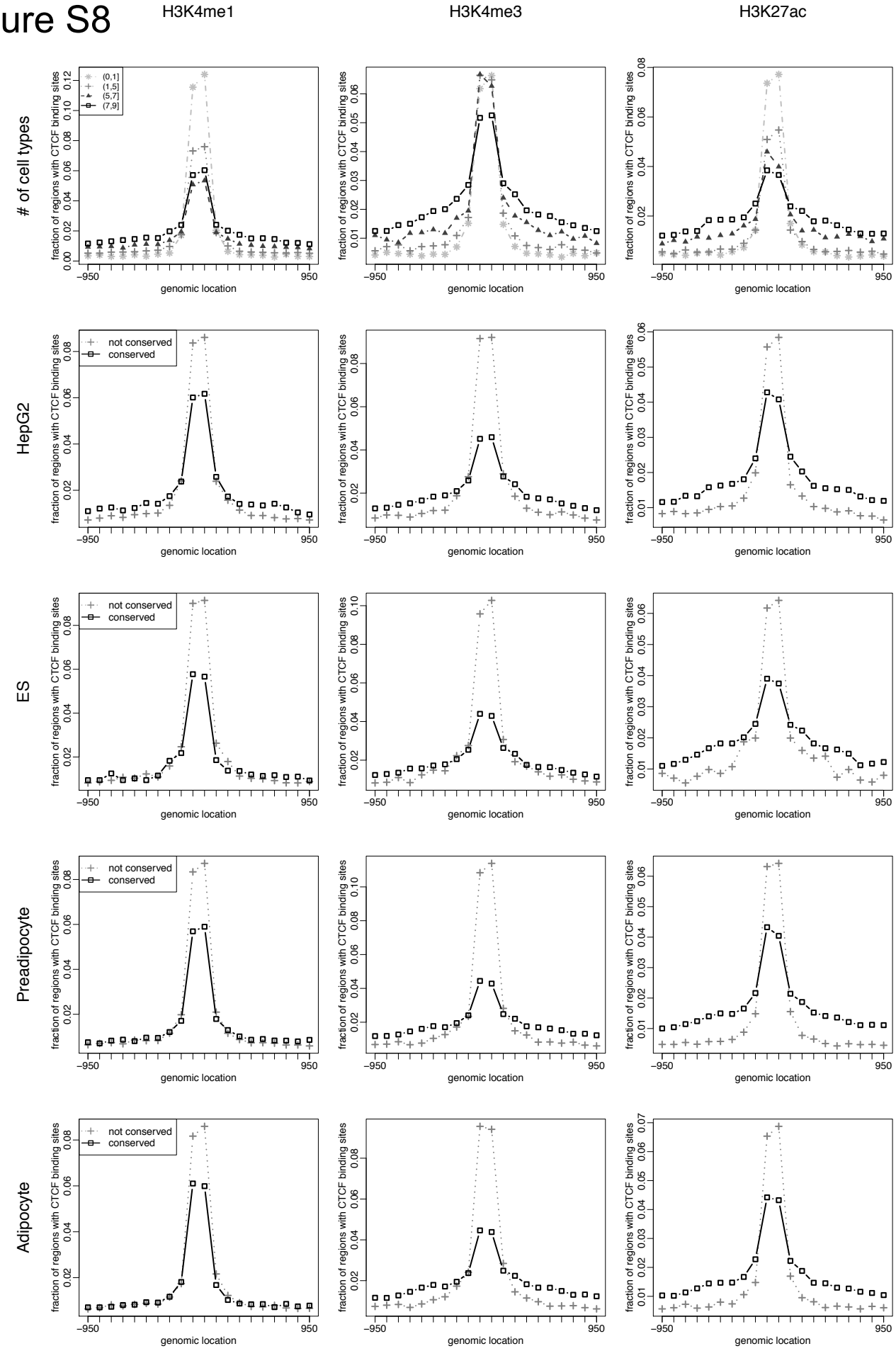

H3K4me1
H3K4me3
H3K27ac
# of cell types
HepG2
ES
Preadipocyte
Adipocyte
(0,1]
(1,5]
(5,7]
(7,9]
not conserved
conserved
fraction of regions with CTCF binding sites
genomic location
−950
950

Figure S9

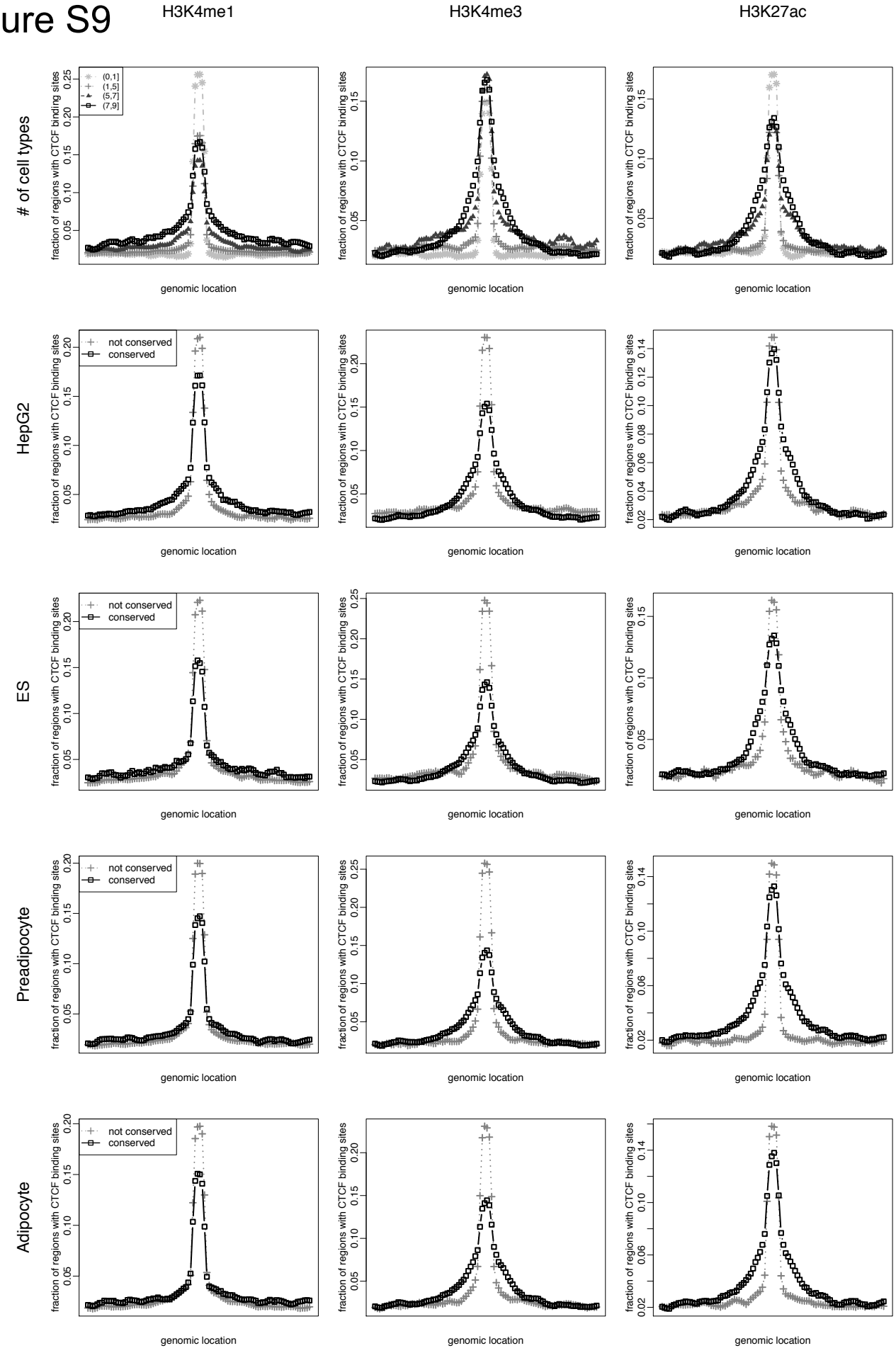
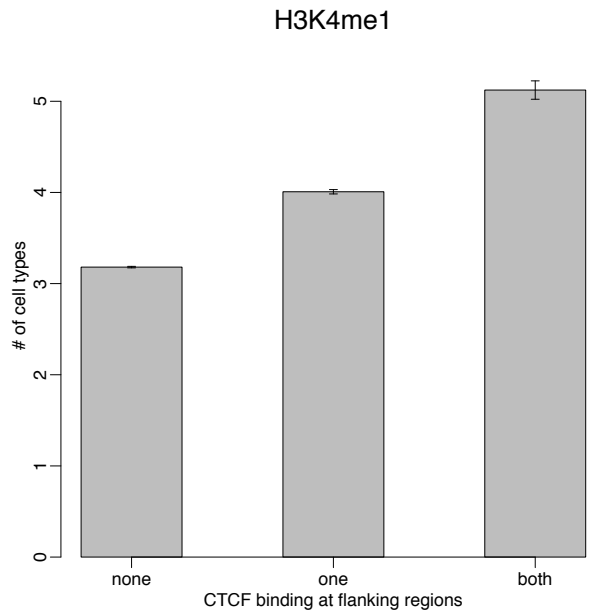
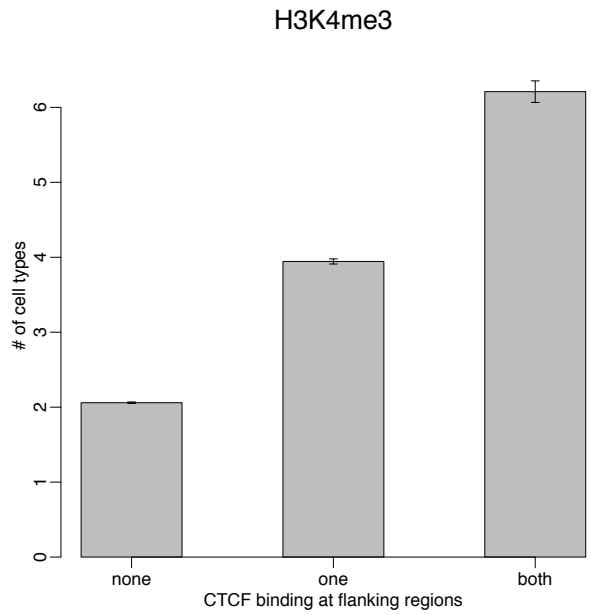

H3K4me1
H3K4me3
H3K27ac
# of cell types
HepG2
ES
Preadipocyte
Adipocyte
(0,1]
(1,5]
(5,7]
(7,9]
not conserved
conserved
fraction of regions with CTCF binding sites
genomic location

Figure S10


H3K4me1
# of cell types
0
1
2
3
4
5
none
one
both
CTCF binding at flanking regions


H3K4me3
# of cell types
0
1
2
3
4
5
6
none
one
both
CTCF binding at flanking regions

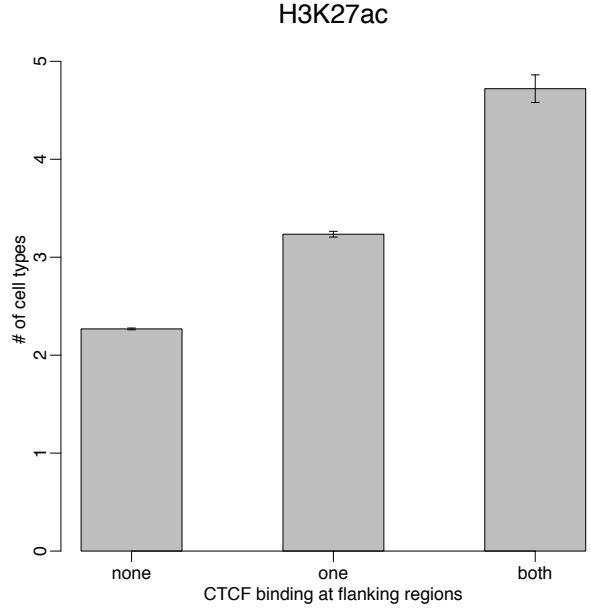

H3K27ac
# of cell types
0
1
2
3
4
5
none
one
both
CTCF binding at flanking regions

Figure S11

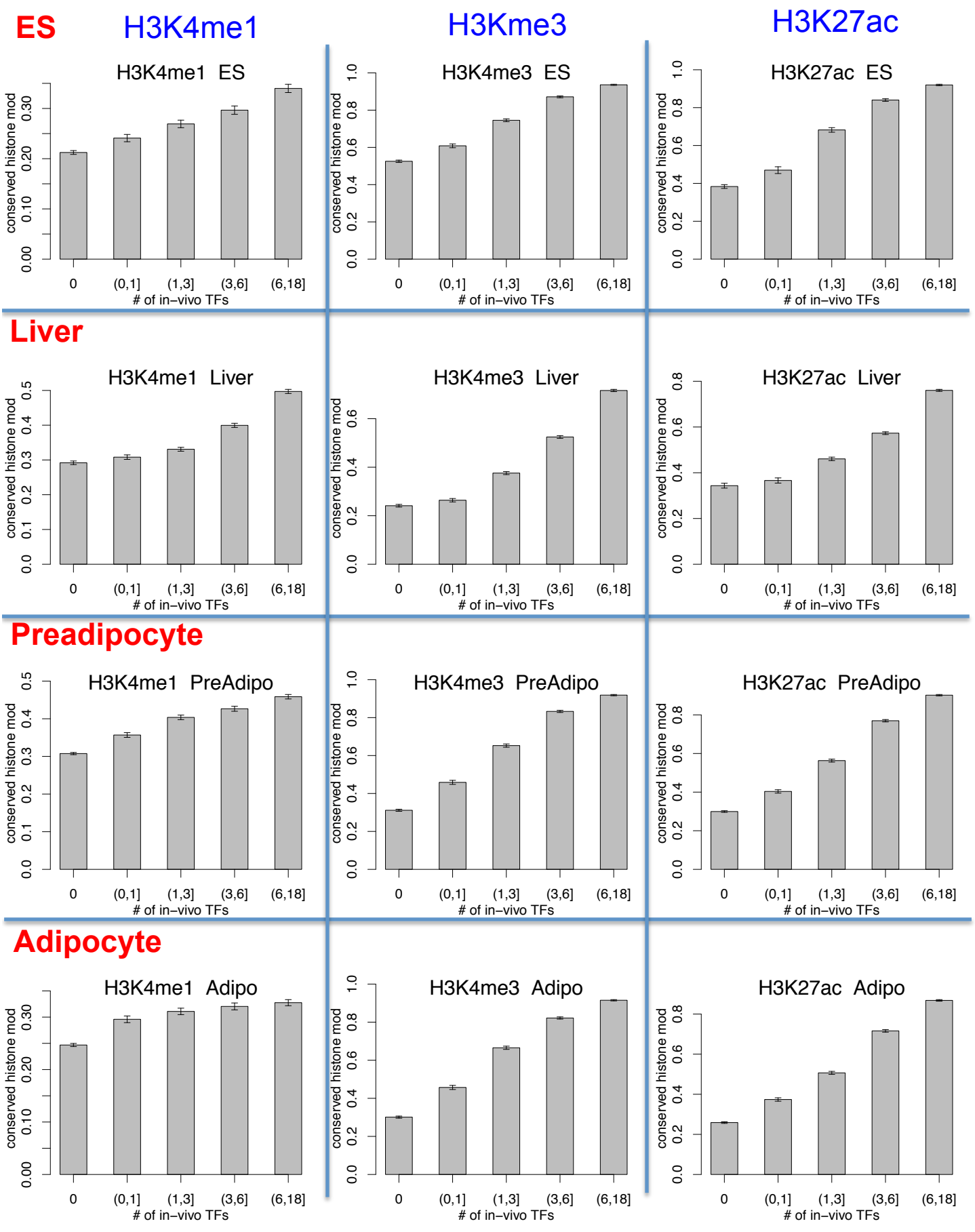

ES
H3K4me1
H3Kme3
H3K27ac
Liver
Preadipocyte
Adipocyte
H3K4me1 ES
H3K4me3 ES
H3K27ac ES
H3K4me1 Liver
H3K4me3 Liver
H3K27ac Liver
H3K4me1 PreAdipo
H3K4me3 PreAdipo
H3K27ac PreAdipo
H3K4me1 Adipo
H3K4me3 Adipo
H3K27ac Adipo
conserved histone mod
# of in−vivo TFs
0
(0,1]
(1,3]
(3,6]
(6,18]

Figure S12

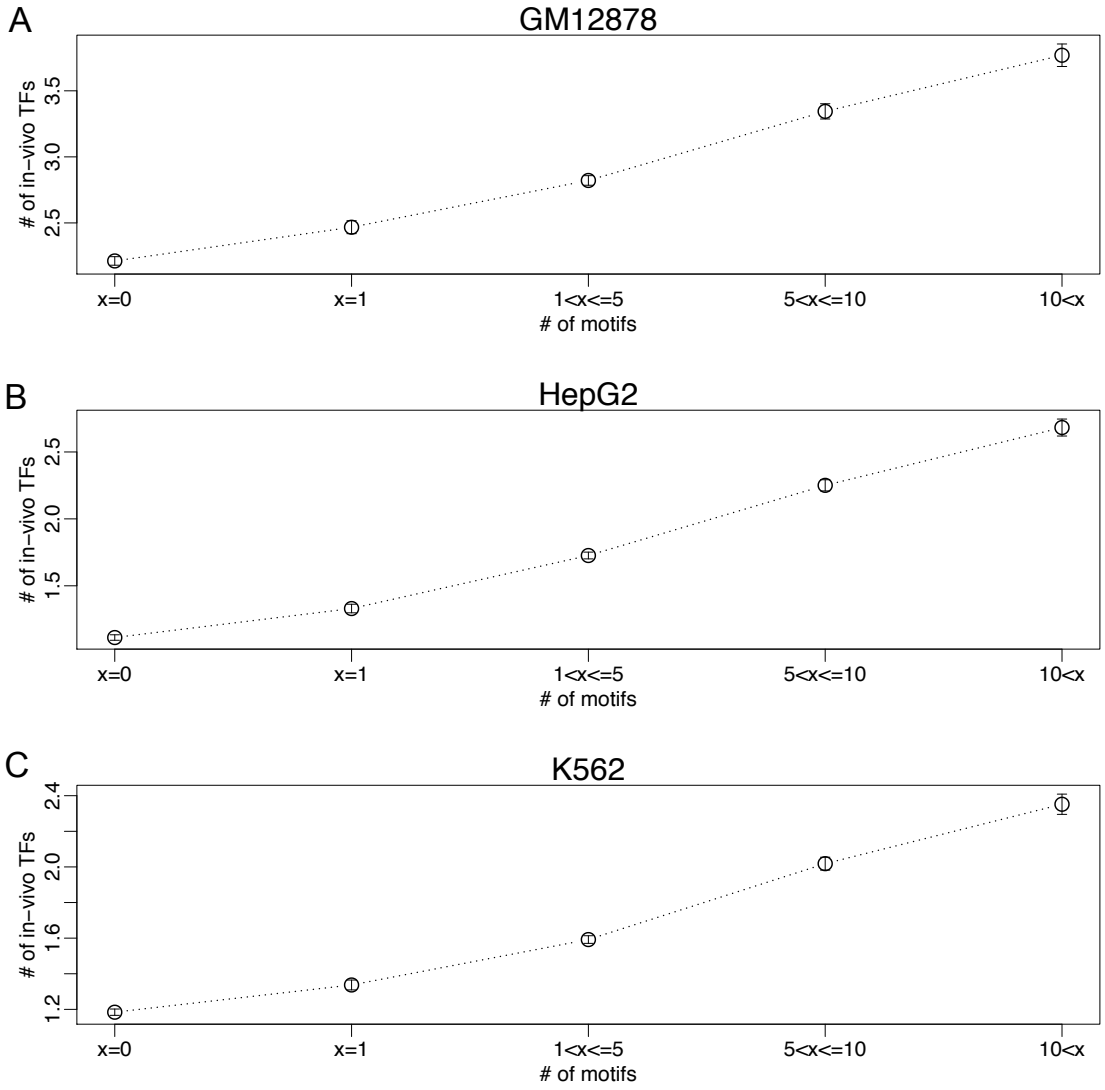

A
GM12878
# of in−vivo TFs
2.5
3.0
3.5
x=0
x=1
1<x<=5
5<x<=10
10<x
# of motifs
B
HepG2
# of in−vivo TFs
1.5
2.0
2.5
x=0
x=1
1<x<=5
5<x<=10
10<x
# of motifs
C
K562
# of in−vivo TFs
1.2
1.6
2.0
2.4
x=0
x=1
1<x<=5
5<x<=10
10<x
# of motifs

Figure S13

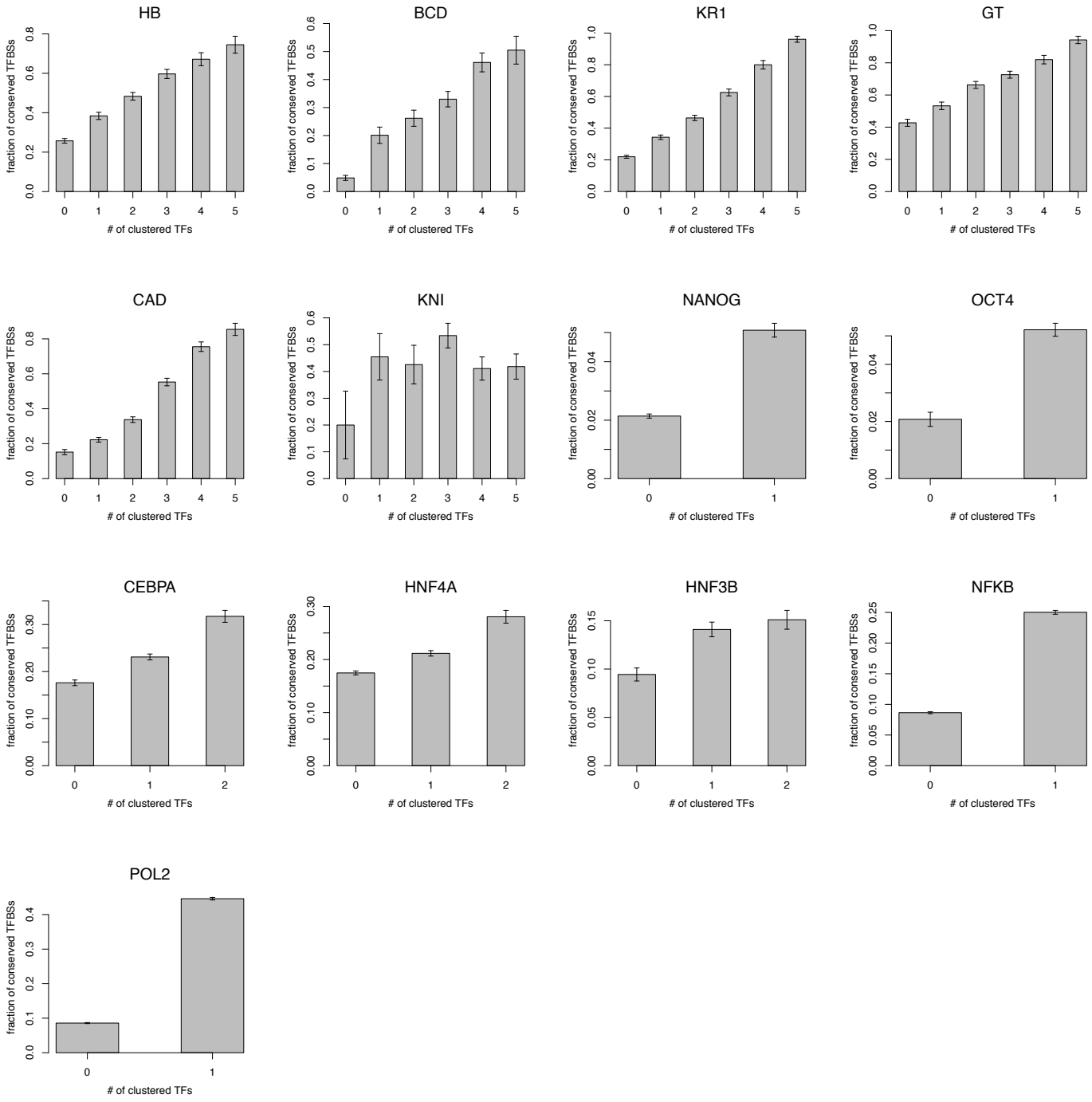

HB
BCD
KR1
GT
CAD
KNI
NANOG
OCT4
CEBPA
HNF4A
HNF3B
NFKB
POL2
fraction of conserved TFBSs
# of clustered TFs

Figure S14

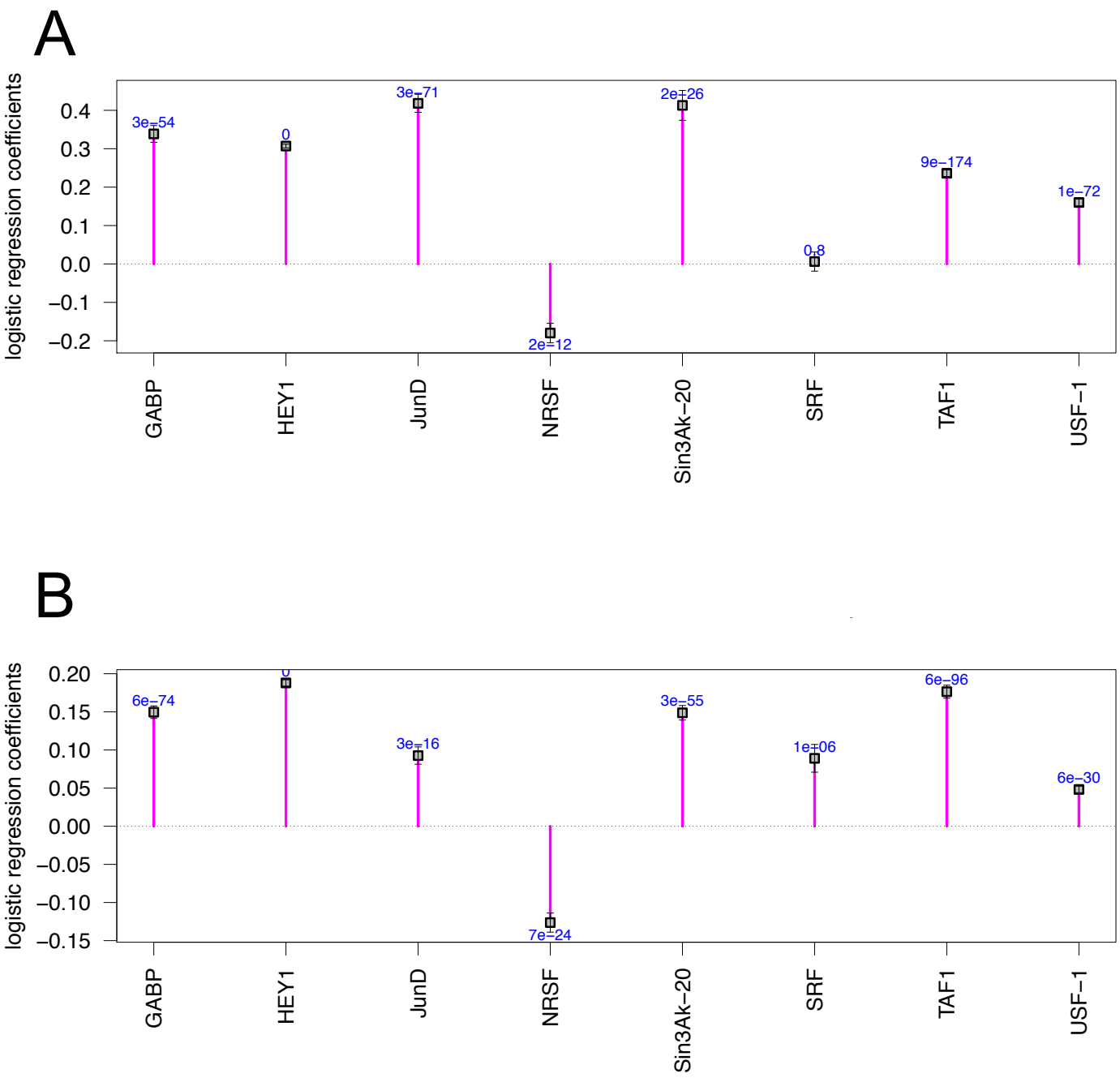

A
logistic regression coefficients
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
3e−54
0
3e−71
2e−12
2e−26
0.8
9e−174
1e−72
GABP
HEY1
JunD
NRSF
Sin3Ak−20
SRF
TAF1
USF−1
B
logistic regression coefficients
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
6e−74
3e−16
7e−24
3e−55
1e−06
6e−96
6e−30
GABP
HEY1
JunD
NRSF
Sin3Ak−20
SRF
TAF1
USF−1

Figure S15

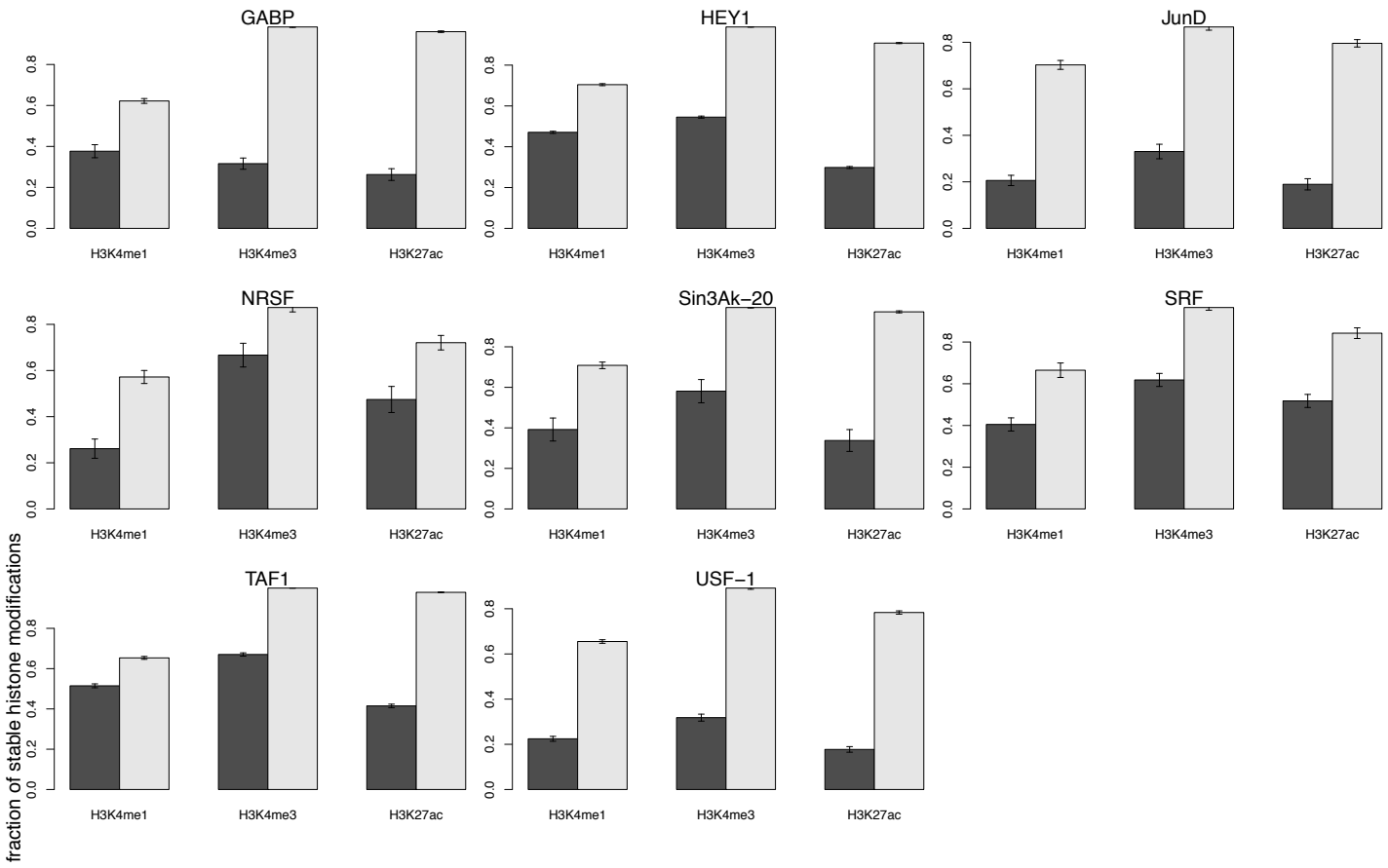

GABP
HEY1
JunD
NRSF
Sin3Ak−20
SRF
TAF1
USF−1
fraction of stable histone modifications
0.0
0.2
0.4
0.6
0.8
H3K4me1
H3K4me3
H3K27ac

Figure S16

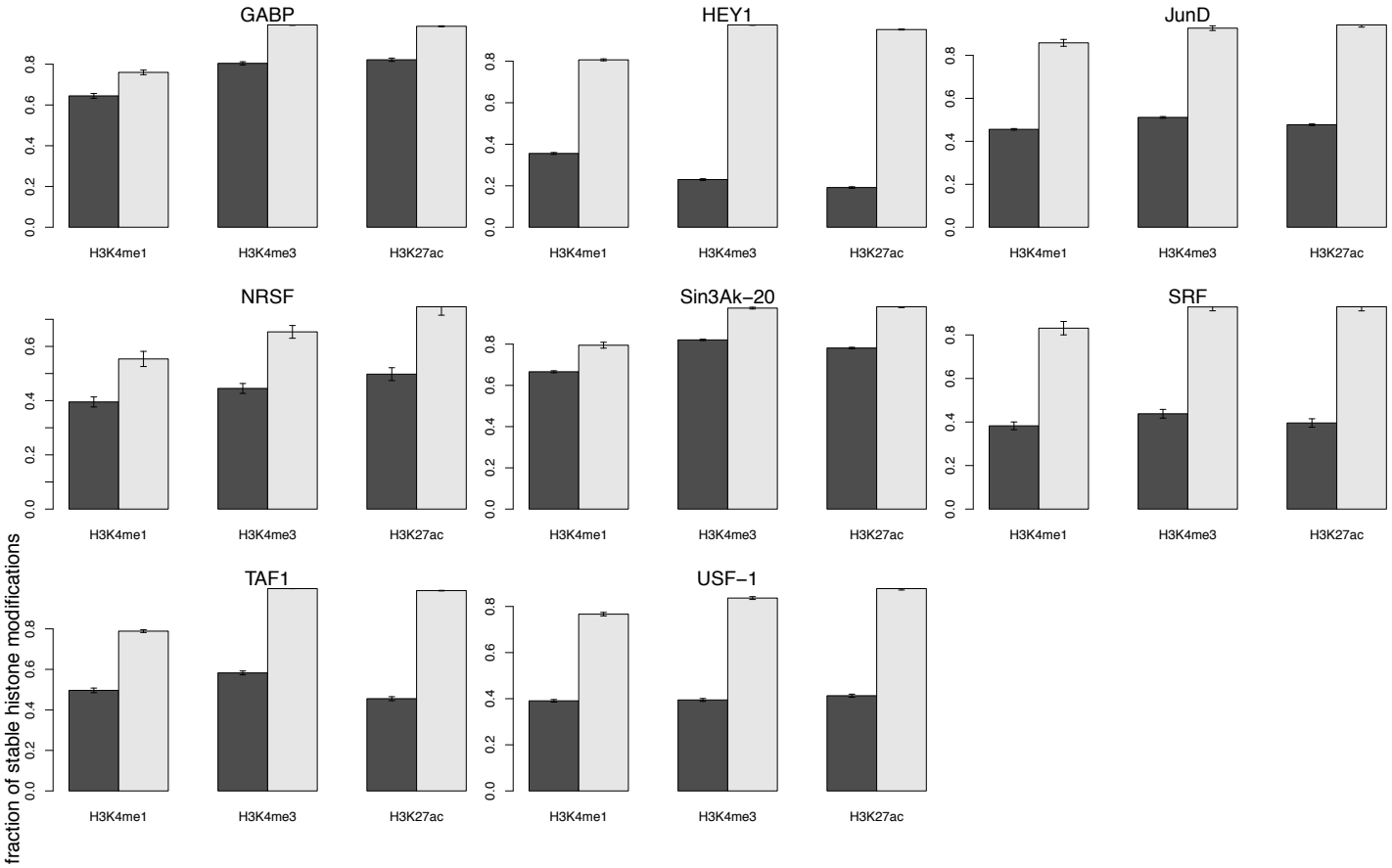
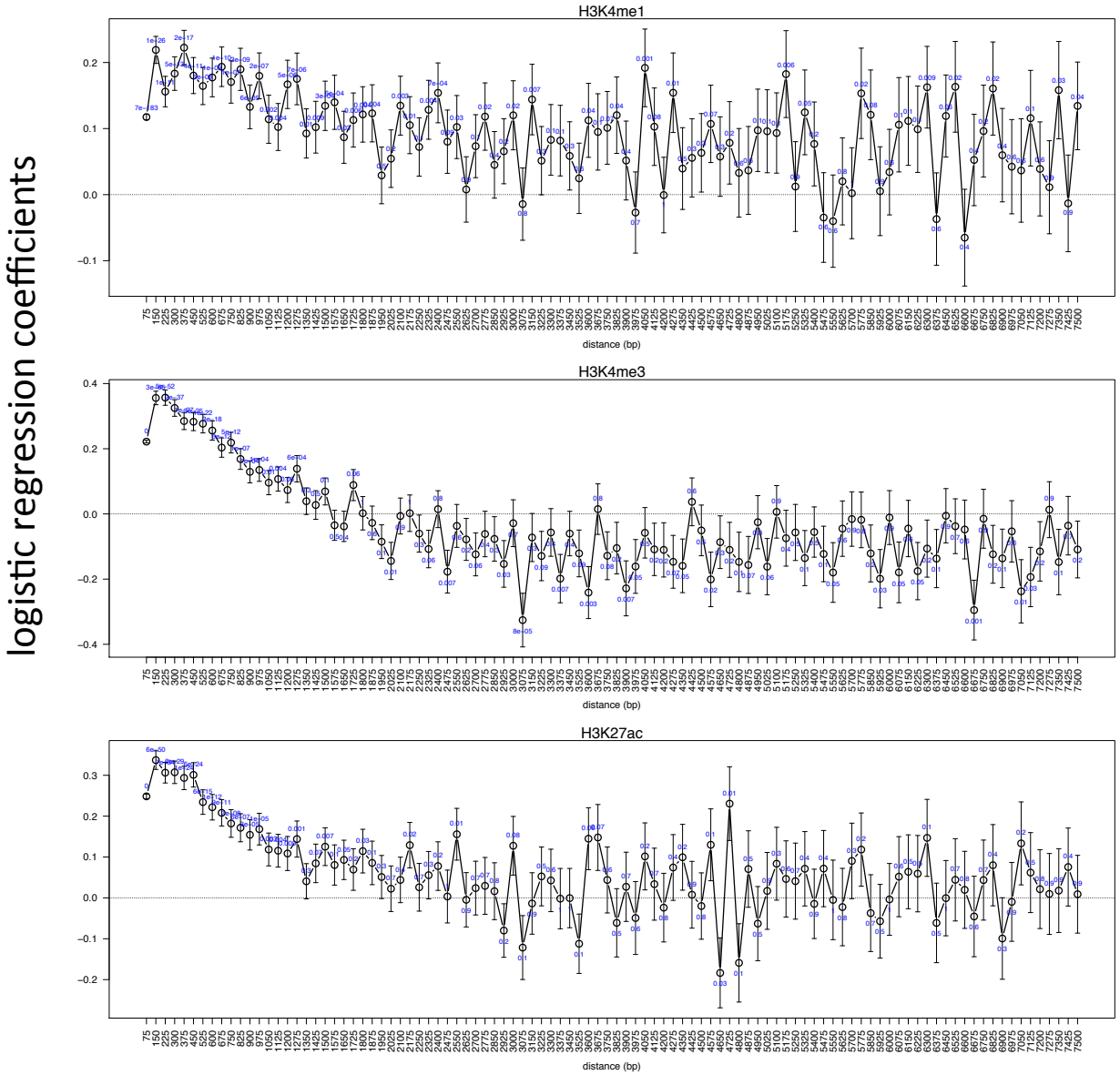

GABP
HEY1
JunD
NRSF
Sin3Ak−20
SRF
TAF1
USF−1
fraction of stable histone modifications
0.0
0.2
0.4
0.6
0.8
H3K4me1
H3K4me3
H3K27ac

Figure S17


logistic regression coefficients
H3K4me1
H3K4me3
H3K27ac
distance (bp)

Supplementary Figure Legends

Figure S1. Distribution of H3K4me3 in 4 human cell types (GM12878, HepG2, HUVEC, NHEK) generated by Broad Institute (A), University of Washington (B), or both (C). The shade of the bar indicates different types of genomic regions: intergenic, proximal promoter, promoter, and intron. We included only regulatory sites with histone modification in at least one cell type.

Figure S2. Evolutionary conservation of histone modifications between species vs. stability of histone modifications among cell types. For all sites with histone modifications in the human genome (the reference genome), we identified the fraction (Y-axis) of the sites whose orthologous region in the mouse genome showed the same histone modification (the test genome). Y-axis: the fraction of evolutionarily conserved histone modifications. X-axis: the number of cell types with histone modifications. Error bars indicate the standard error based on a binomial distribution. The analysis was repeated for 4 cell types (embryonic stem cell, liver, preadipocyte, and adipocyte) (rows) for 3 histone marks (H3K4me1, H3K4me3, and H3K27ac) (columns).

Figure S3. Evolutionary conservation of histone modifications vs. stability of histone modifications among cell types, stratified by the type of genomic region. Y-axis: the fraction of evolutionarily conserved histone modifications. X-axis: the number of cell types with histone modifications. The histone marks tested were H3K4me1 and H3K27ac. The genomic sites were classified as promoters, proximal promoter, intronic, or intergenic regions. For H3K4me1 and K3K27ac, we repeated the analysis in intronic and intergenic regions, after filtering out sites with H3K4me3 modifications, and we still detected the same trend (data not shown).

Figure S4. Evolutionary conservation of histone modifications versus gene expression breadth. Y-axis: the fraction of evolutionarily conserved histone modifications. X-axis: gene expression breadth of the nearby genes. We removed sites that do not have a known gene within 10kb. Expression breadth was defined as the number of cell types in which the gene shows an expression

level above a threshold (log2-transformed *rma* expression value from Affymetrix GeneChip > 5). Error bars indicate the standard error based on a binomial distribution. Rows indicate histone marks tested and the cell-type of the cross-species comparison of the histone modification. We repeated the analysis with tissue expression profiles from another source (http:// http://biogps.org/downloads) and found the same result (data not shown).

Figure S5. Evolutionary conservation of histone modifications (Y-axis) and gene expression breath (X-axis), stratified by histone modification stability (individual lines). Y-axis: the fraction of evolutionarily conserved histone modifications. X-axis: gene expression breadth of the nearby genes. We removed sites that do not have a known gene within 10kb. Expression breadth was defined as the number of cell types in which the gene shows an expression level above a threshold (log2-transformed *rma* expression value from Affymetrix GeneChip > 5). The regulatory sites were stratified based on their histone modification stability (individual lines).

Figure S6-S8. Evolutionary conservation of histone modifications and stability of histone modifications among cell types vs. nucleotide composition and CTCF binding distribution. X-axis: genomic distance (bp) from the regulatory site. Each point represents a 100bp interval; for example, the leftmost point (X= -950bp) indicates genomic regions that are 900bp to 1000bp upstream of the regulatory site on the positive strand of the genome. Y-axis: fraction of G and C nucleotides (S6), fraction of CpG dinucleotides (S7), or fraction of sites bound by CTCF (S8). Individual lines indicate sites with varying degrees of histone modification stability (1^{ST} row) or conservation (2^{ND} TO 5^{TH} row), where the lines indicate sites whose histone modifications are conserved or not. Histone marks tested were H3K4me1 (1^{ST} column), H3K4me3 (2^{ND} column) and H3K27ac (3^{RD} column).

Figure S9. Evolutionary conservation of histone modifications and stability of histone modifications among cell types vs CTCF binding distribution. X-axis: genomic distance (bp) from the regulatory site in a 10kb window. Each point represents a 500bp interval, overlapped with intervals of an adjacent point by

400bp; for example, leftmost point would indicate genomic regions that are 4500bp to 5000bp upstream of the regulatory site on the positive strand of the genome, and the point that is immediate right of the leftmost point would indicate regions 4400bp to 4900bp. See Figure S8 for legends.

Figure S10. Histone modification stability as a function of CTCF binding at the flanking regions. X-axis: binding at the flanking regions, from 100 to 1000bp on either side. The possible values are none, one, or both. Y-axis: the number of cell types with histone modifications.

Figure S11. Evolutionary conservation of histone modifications vs. TFBS clustering. X-axis: the number of *in-vivo* TFBSs within a 1kb window. We obtained *in-vivo* TFBSs from the HepG2 cell line generated as a part of the ENCODE project. We counted the number of different TFs bound to the region, instead of the total number of TFBSs, thus removing clustered bindings of the same TF. Y-axis: the fraction of evolutionarily conserved histone modifications between human and mouse ES cells (1^{st} row), liver cells (2^{nd} row), undifferentiated adipocytes (3^{rd} row) and differentiated adipocytes (4^{th} row). Columns indicate the three histone marks tested. Figure 6 is identical to the 2^{nd} row.

Figure S12. Genetic basis for *in-vivo* TFBSs. The number of *in-vivo* bound TFs (Y-axis) in ENCODE cell lines GM12878 (A), HepG2 (B), and K562 (C) as a function of TF binding sequence motifs in the region (X-axis). Genome-wide TF binding motifs were obtained from the UCSC genome browser, and the number of unique TFs was counted, i.e. a TF with multiple binding motifs in the region was counted as one.

Figure S13. Evolutionary conservation of TFBS versus TFBS clustering. The fraction of evolutionarily conserved TFBSs (Y-axis) was examined as a function of the number of clustered TFBSs (X-axis). Human was the reference, and mouse the test. X-axis: the number of TFs that overlapped with the TFBSs tested. For example, to calculate the number of clustered TFBSs around CEBPA binding sites, we examined whether the binding sites of HNF4A, those of HNF3B,

or both overlapped with CEBPA binding sites. For NANOG and OCT4, we added 500bp to the boundary of the peaks because the conservation rate was extremely low using the original boundary.

Figure S14. Conservation of TFBSs among cell types versus TFBS clustering. For each TF, conservation of the binding sites across HepG2 and K562 was examined as a function of the number of clustered TFBSs in the reference. K562 and HepG2 were used as the reference and the test, respectively, in (A) which were reversed in (B). Y-axis: coefficient from the logistic regression, which measures the extent of changes in conservation when clustered with other TFBSs. Positive and negative values indicate increases or decreases in the conserved fraction when clustered with other TFSBs, respectively. P-values for statistical significance of the coefficient were given at each X value. X-axis: the number of individual TFs.

Figure S15, S16. Conservation of histone modifications among cell types versus conservation of TFBSs among cell types. Barplots show the fraction of stable histone modifications as a function of TF binding stability between two human cell types, K562 and HepG2 (X-axis). Black and gray bars indicate TF binding sites whose orthologous regions in the mouse showed conservation. K562 and HepG2 were used as the reference and the test, respectively, in Figure S15, which were reversed in Figure S16.

Figure S17. Conservation of histone modifications versus TF clustering at various distances. Conservation between mouse and human liver cells (Y-axis) as a function of the number of TFBSs bound at various genomic distances from the site (X-axis). Y-axis: coefficients from a logistic regression, whose positive and negative values would indicate increase or decrease in the conservation of histone modifications, respectively, as the number of clustered TFBS at the given genomic distance increases. Error bars indicate the standard error of the coefficient. P-values for statistical significance of the coefficient were given at each X value. We detected the same result for histone modification stability across human cell types (data not shown).

Supplementary Tables

Table S1. List of histone modifications used in the study.

	Human (reference genome)		**Mouse (test genome)**	
Cell type	**Histone Mark**	**Source**	**Histone mark**	**Source**
Embryonic Stem Cell	H3K4me1,H3K4me3,H3K27ac	ENCODE project (H1)	H3K4me1,H3K4me3,H3K27ac	GSE24164
Preadipocyte (before differentiation)	H3K4me1,H3K4me3,H3K27ac	GSE21365	H3K4me1,H3K4me3,H3K27ac	GSE21365
Adipocyte (after differentiation)	H3K4me1,H3K4me3,H3K27ac	GSE21365	H3K4me1,H3K4me3,H3K27ac	GSE21365
Liver	H3K4me1,H3K4me3,H3K27ac	ENCODE project (HepG2)	H3K4me1,H3K4me3,H3K27ac	GSE24164

Table S2. List of cross-species *in-vivo* TF binding data sets.

Species	***Antibody***	***Cell type***	***Source***
human (individuals)	NFKB	B lymphoid	http://www.embl.de/~korbel/ChIP_Seq/Snyder_Korbel_Labs_Variation (Kasowski et al. 2010)
human (individuals)	PolII	B lymphoid	http://www.embl.de/~korbel/ChIP_Seq/Snyder_Korbel_Labs_Variation (Kasowski et al. 2010)
human vs mouse	NANOG	Embryonic Stem Cell	Kunarso et al. 2010 (hsap) (Kunarso et al. 2010) and GSE11431 (mmus)
human vs mouse	OCT4	Embryonic Stem Cell	Kunarso et al. 2010 (hsap) (Kunarso et al. 2010) and GSE11431 (mmus)
human vs mouse	CEBPA	Liver	http://www.ebi.ac.uk/~benoit/cebpa_science2010/ (Schmidt et al. 2010)
human vs mouse	HNF4A	Liver	http://www.ebi.ac.uk/~benoit/cebpa_science2010/ (Schmidt et al. 2010)
human vs mouse	FOXA2 (HNF3B)	Liver	GSE25836
human vs mouse	Pparg	Adipocyte	GSE21365
D. melanogaster vs D. yakuba	HB	embryo (2-4h)	GSE20369
D. melanogaster vs D. yakuba	BCD	embryo (2-4h)	GSE20369
D. melanogaster vs D. yakuba	KR1	embryo (2-4h)	GSE20369
D. melanogaster vs D. yakuba	GT	embryo (2-4h)	GSE20369
D. melanogaster vs D. yakuba	CAD	embryo (2-4h)	GSE20369
D. melanogaster vs D. yakuba	KNI	embryo (2-4h)	GSE20369

Literature Cited For Supplementary Tables

Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, Waszak SM, Habegger L, Rozowsky J, Shi M, Urban AE et al. (x co-authors. 2010. Variation in transcription factor binding among humans. Science 328:232-235.

Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat. Genet. 42:631-634.

Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S et al. (x co-authors. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science 328:1036-1040.