

Additional File 1

Supplementary data to

“A computational approach for the functional classification of the epigenome”

Francesco Gandolfi¹, Anna Tramontano^{1,2}

¹Department of Physics, Sapienza Università di Roma, P.le Aldo Moro 00183 Rome, Italy

²Istituto Pasteur, Fondazione Cenci Bolognetti, Viale Regina Elena 291, 00161, Rome, Italy

Figure S1 – Selection of the best factorization rank



Figure S2 – Significance of chromatin profile enrichment in distinct genomic features

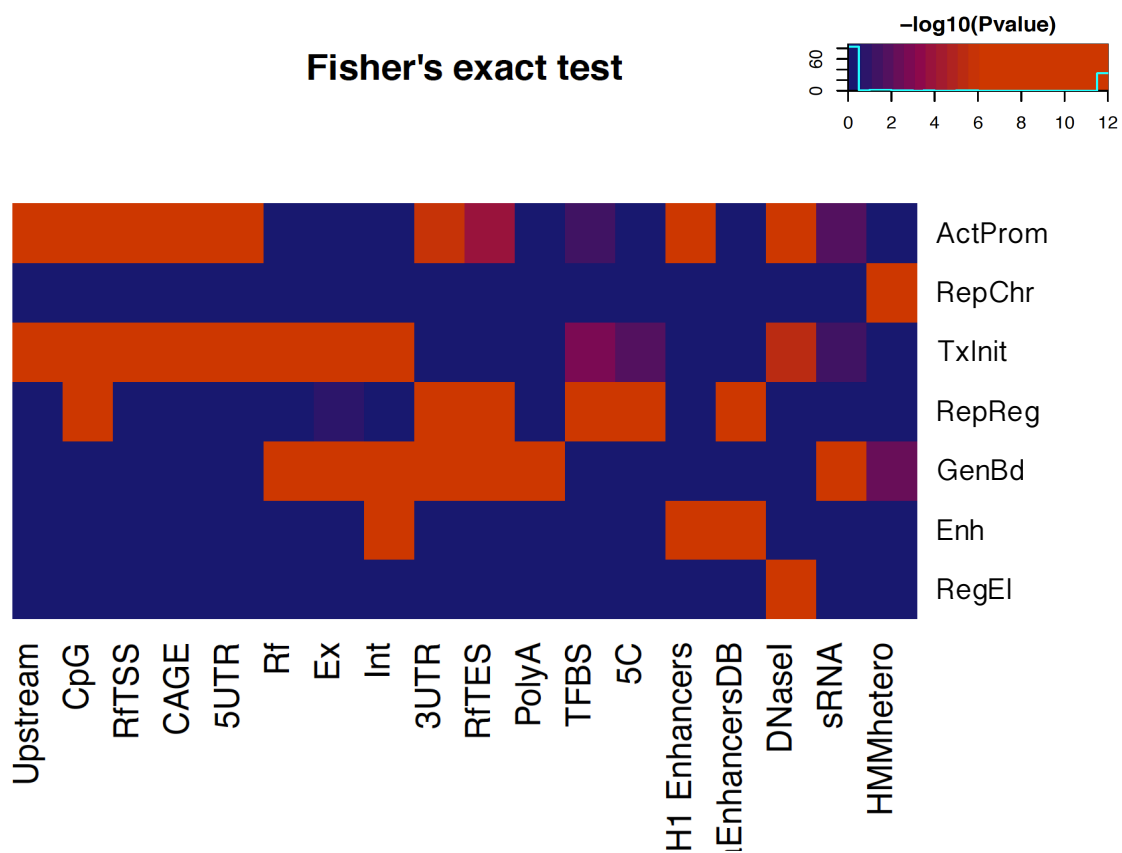


Figure S3 – Frequency of transition between epigenetic profiles

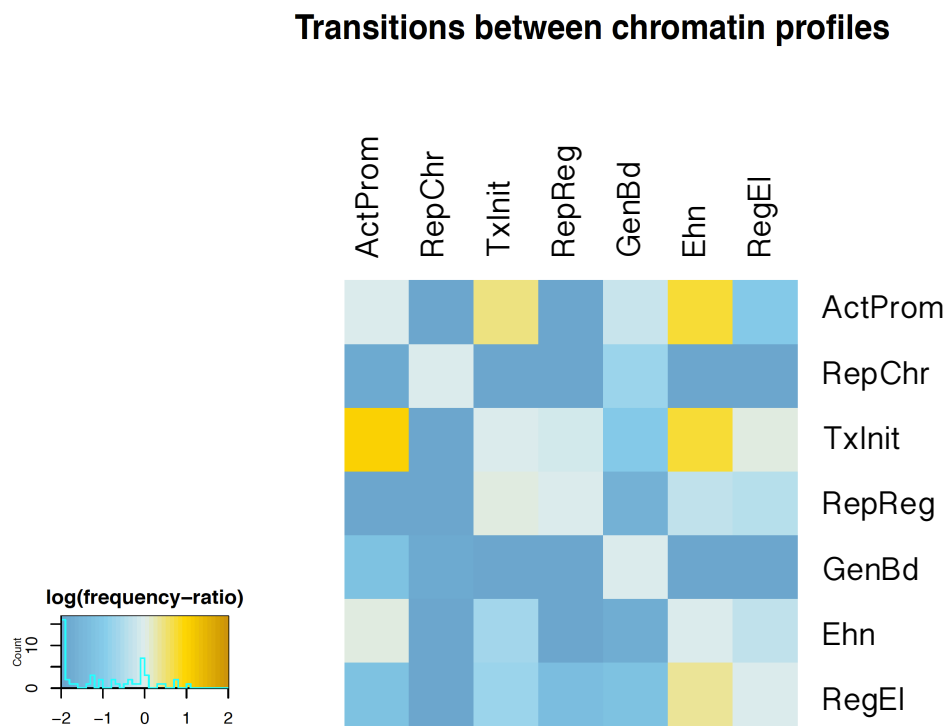


Figure S4 – Recovery power of poly-adenylation sites in chromatin profiles and single epigenetic marks

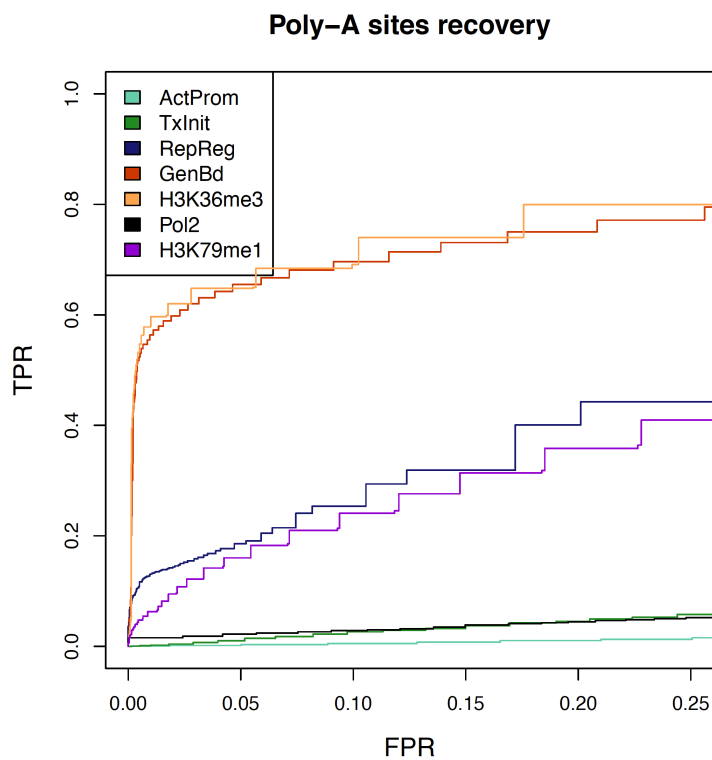


Figure S5 – Recovery of genomic information using ambiguous profile assignment

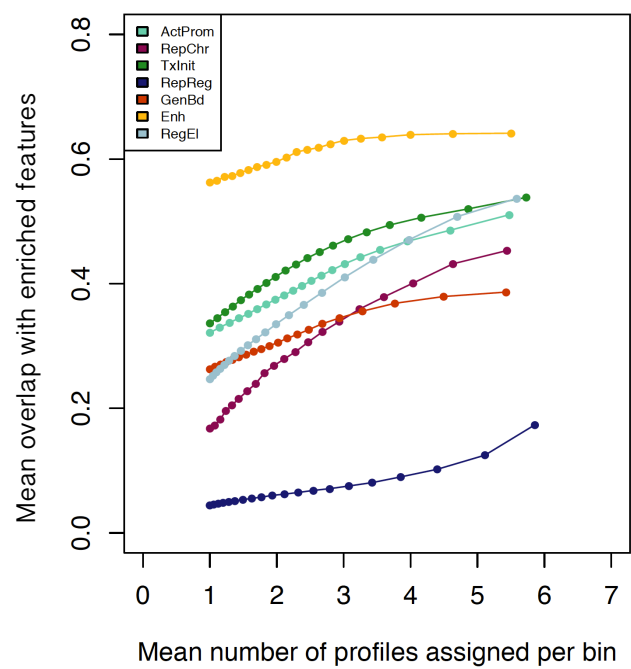


Figure S6 – Chromatin profile assignment according to expression and distance from the TSS

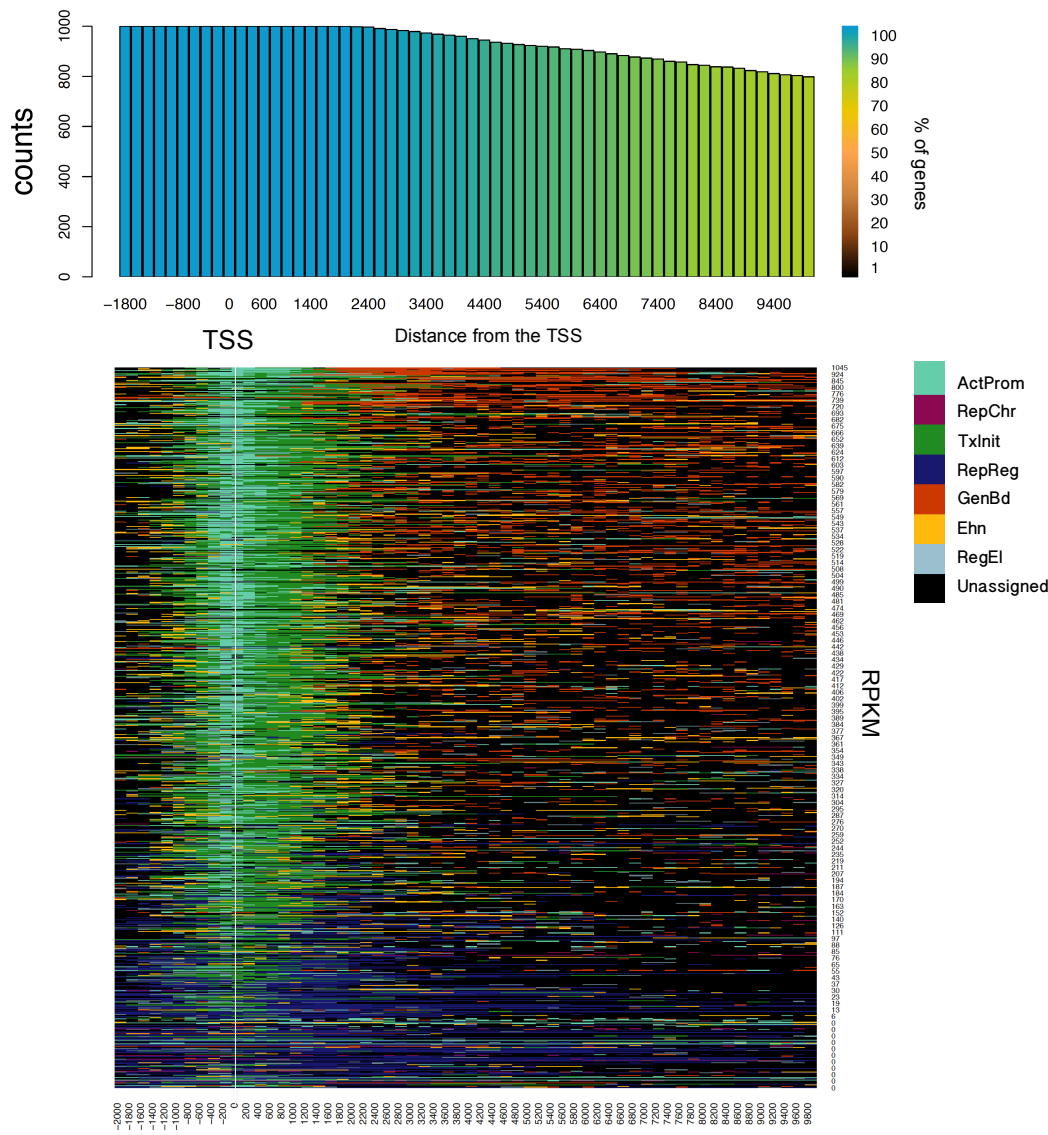


Figure S7 – Frequency of chromatin profiles according to expression and distance from the TSS

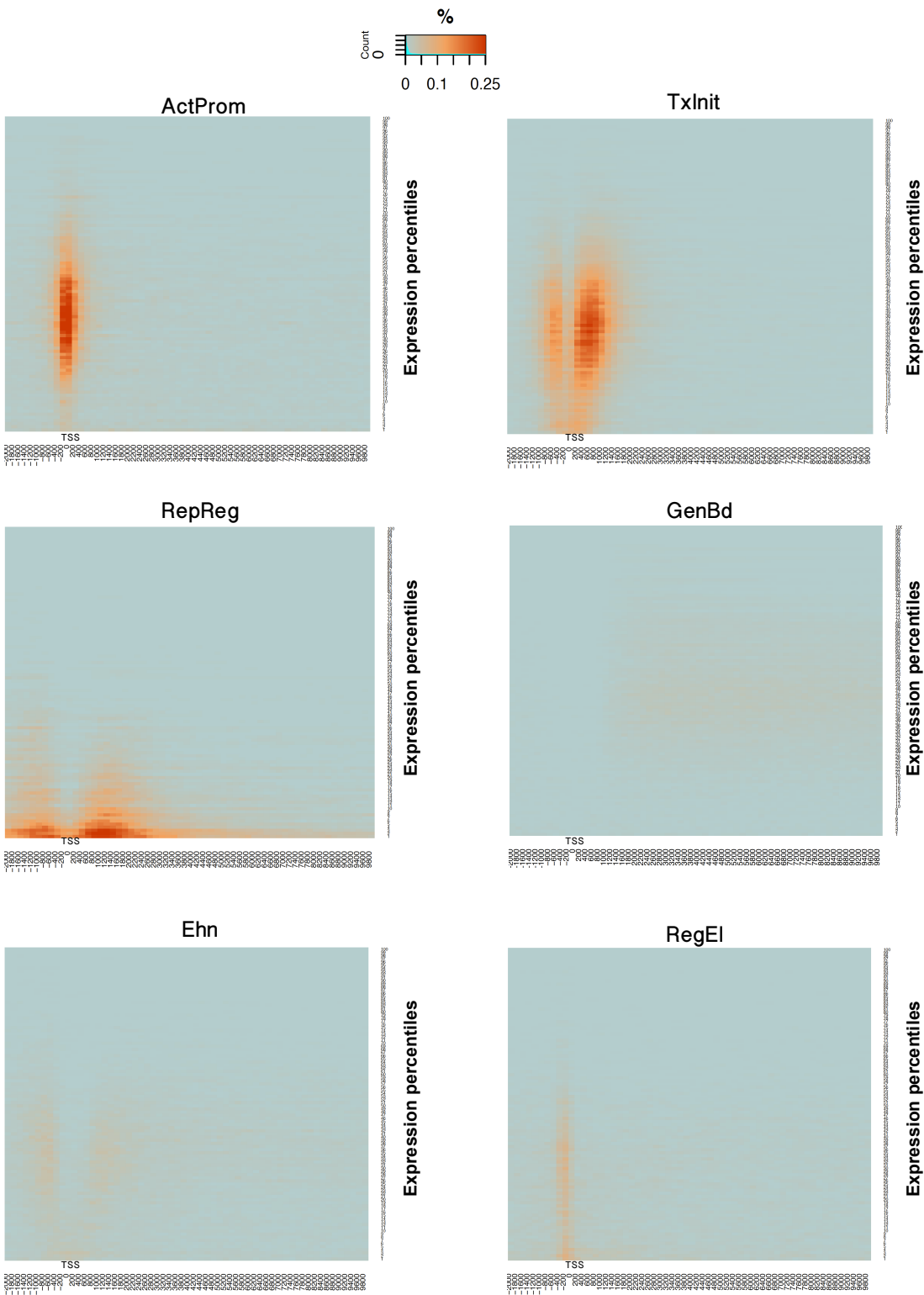


Figure S8 – Distribution of gene expression in different subclusters

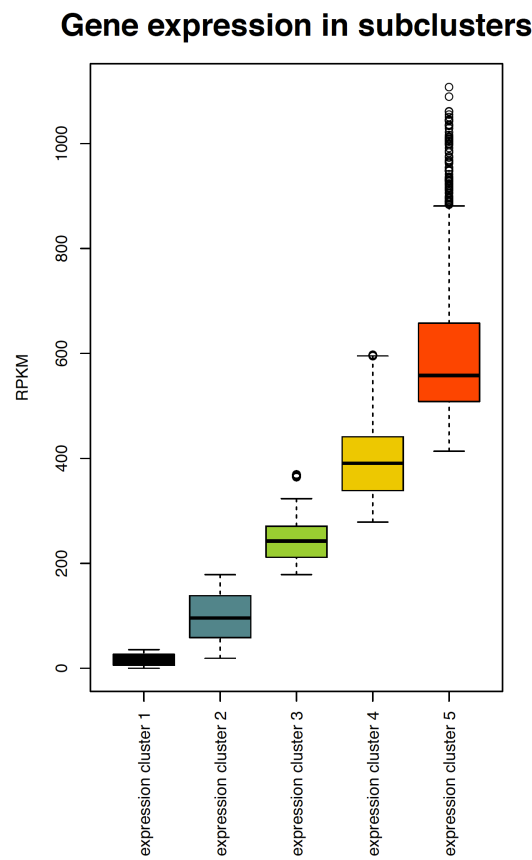


Figure S9 – Feature enrichment distributions across profiles/states using different methods

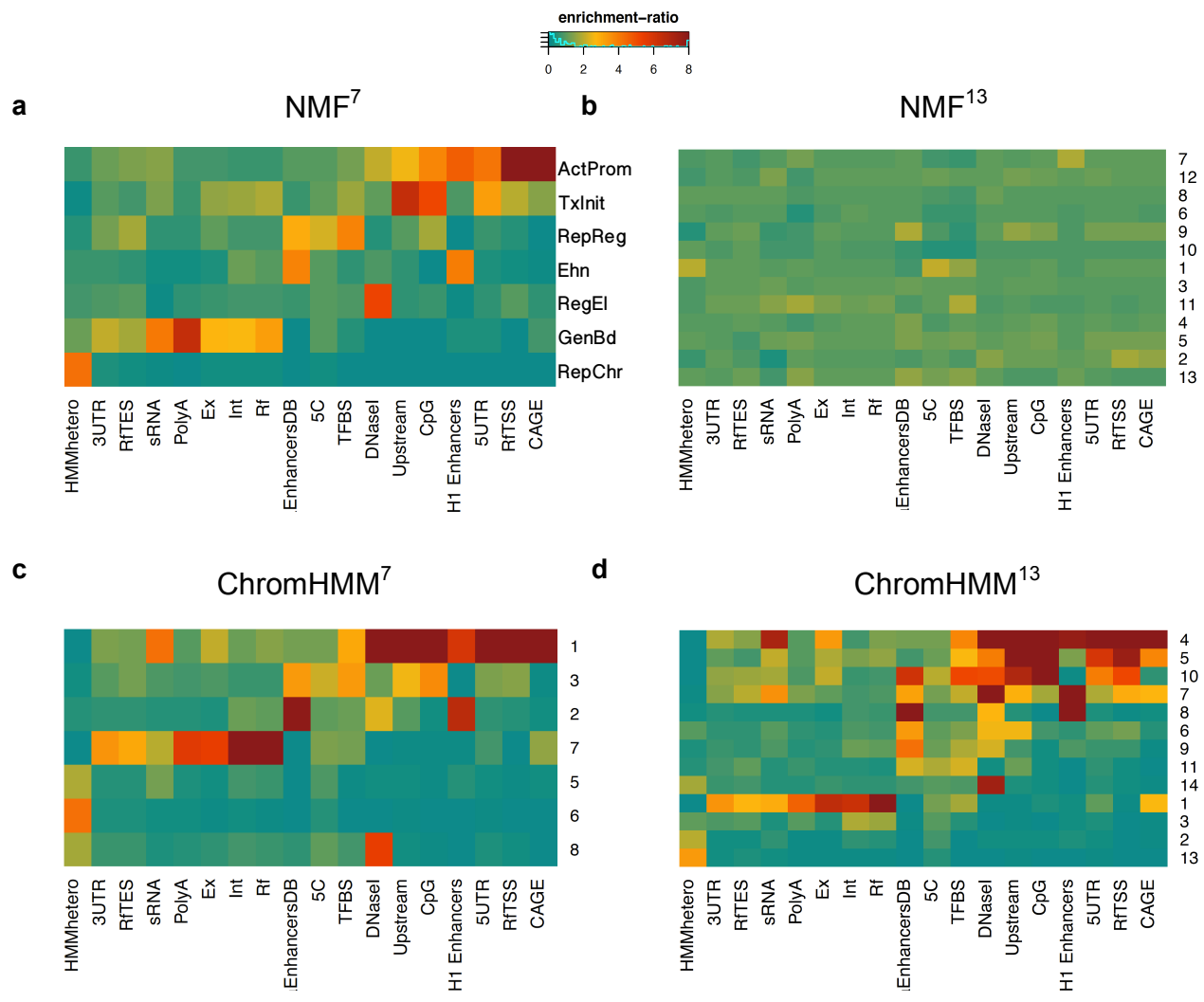


Figure S10 – Overlap and coverage of functional genomic regions with different combinations of enriched profiles/states

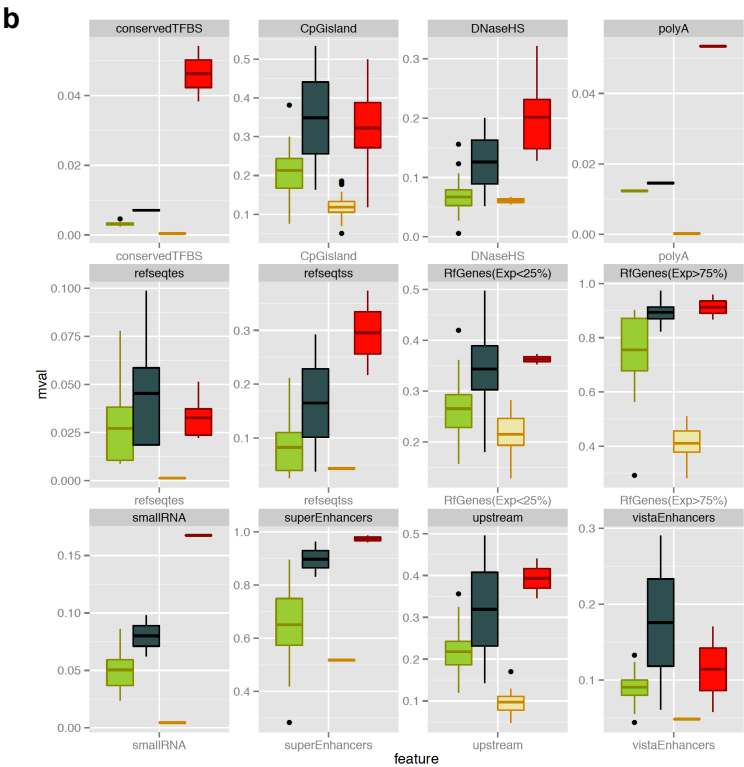
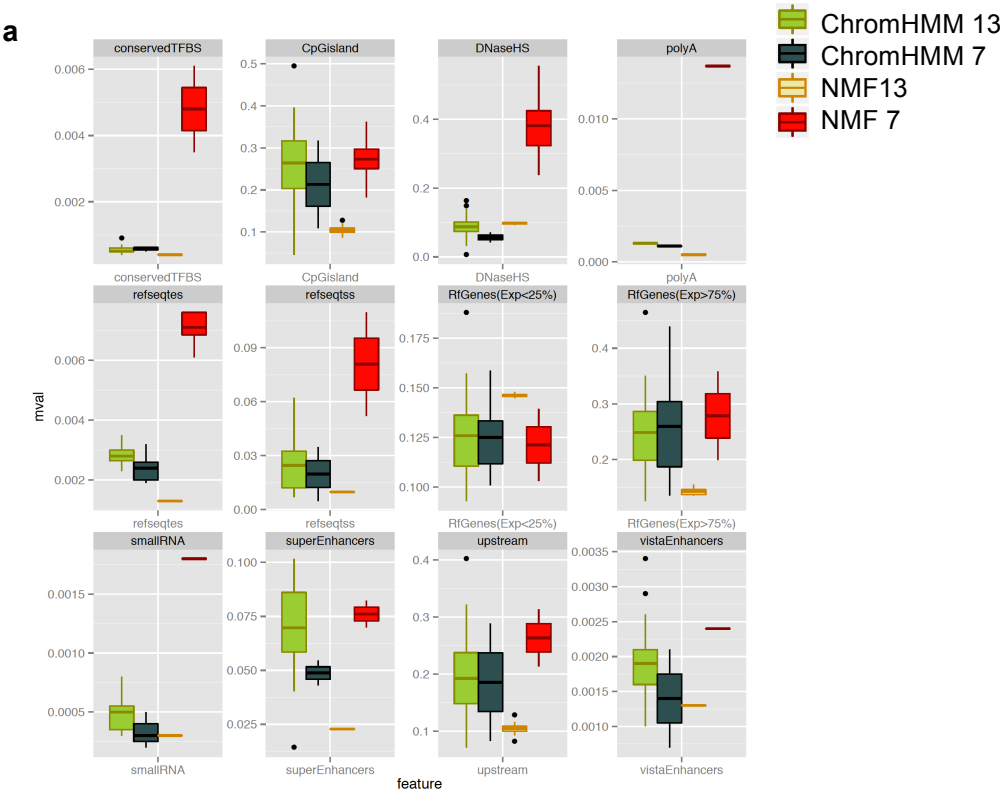


Table S1 – Combinations of enriched profiles/states used to test the overlap with genomic features in each method.

Feature	ChromHMM 13 states		ChromHMM 7 states		NMF 13 profiles		NMF 7 profiles	
	states/profiles	n.combinations	states/profiles	n.combinations	states/profiles	n.combinations	states/profiles	n.combinations
1kb-upstream	4,5,6,7,10	31	1,3	3	2,5,9,12	15	ActProm,Txlnit	3
CpG islands	4,5,7,10	15	1,3	3	2,3,4,5,9,12	63	ActProm,Txlnit, RepReg	7
RefseqTSS(+/-50bp)	4,5,7,10	15	1,3	3	2	1	ActProm,Txlnit	3
RefseqTES(+/-50bp)	1,4,7,10	15	1,3,7	7	11	1	ActProm, RepReg, GenBd	7
Refseq genes (>75%exp)	1,3,4,5,9	31	1,2,7	7	4,11,12	7	Txlnit, GenBd	3
Refseq genes (<25%exp)	1,3,4,5,9	31	1,2,7	7	4,11,12	7	Txlnit, GenBd	3
PolyAdenylation sites	1	1	7	1	11	1	GenBd	1
Conserved TFBS	4,5,9,10,11	31	1,3	3	11	1	Txlnit, RepReg,	3
superEnhancers H1	4,5,7,8	15	1,2	3	7	1	ActProm, Ehn	3
vistaEnhancersDB	6,7,8,9,10,11	63	1,2	3	13	1	RepReg, Ehn	3
Dnase-HS	6,7,8,9,10,11,14	127	2,3	3	2,8	3	ActProm,Txlnit, RegEI	7
smallRNAs	1,4,7	7	1,7	3	11	1	GenBd	1

Figure S11 – NMF and ChromHMM performance in the prediction of distinct functional elements in the genome

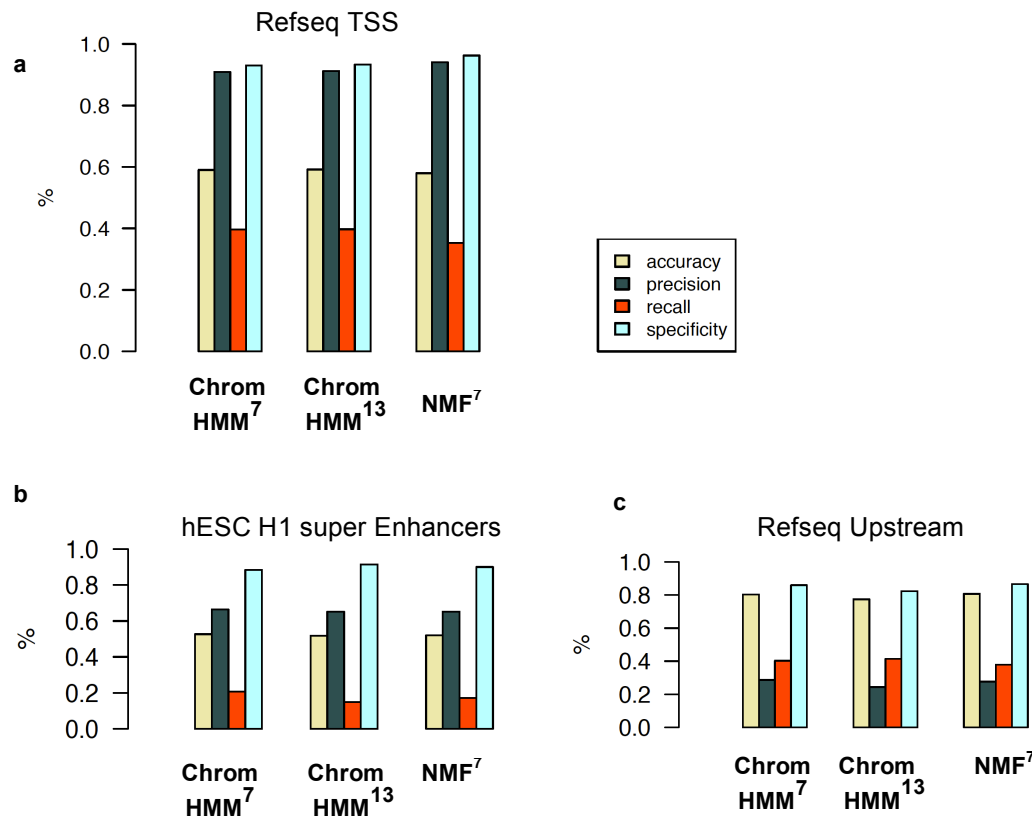


Figure S12 – Comparison of the NMF chromatin-segmentation approach in different cell lines

