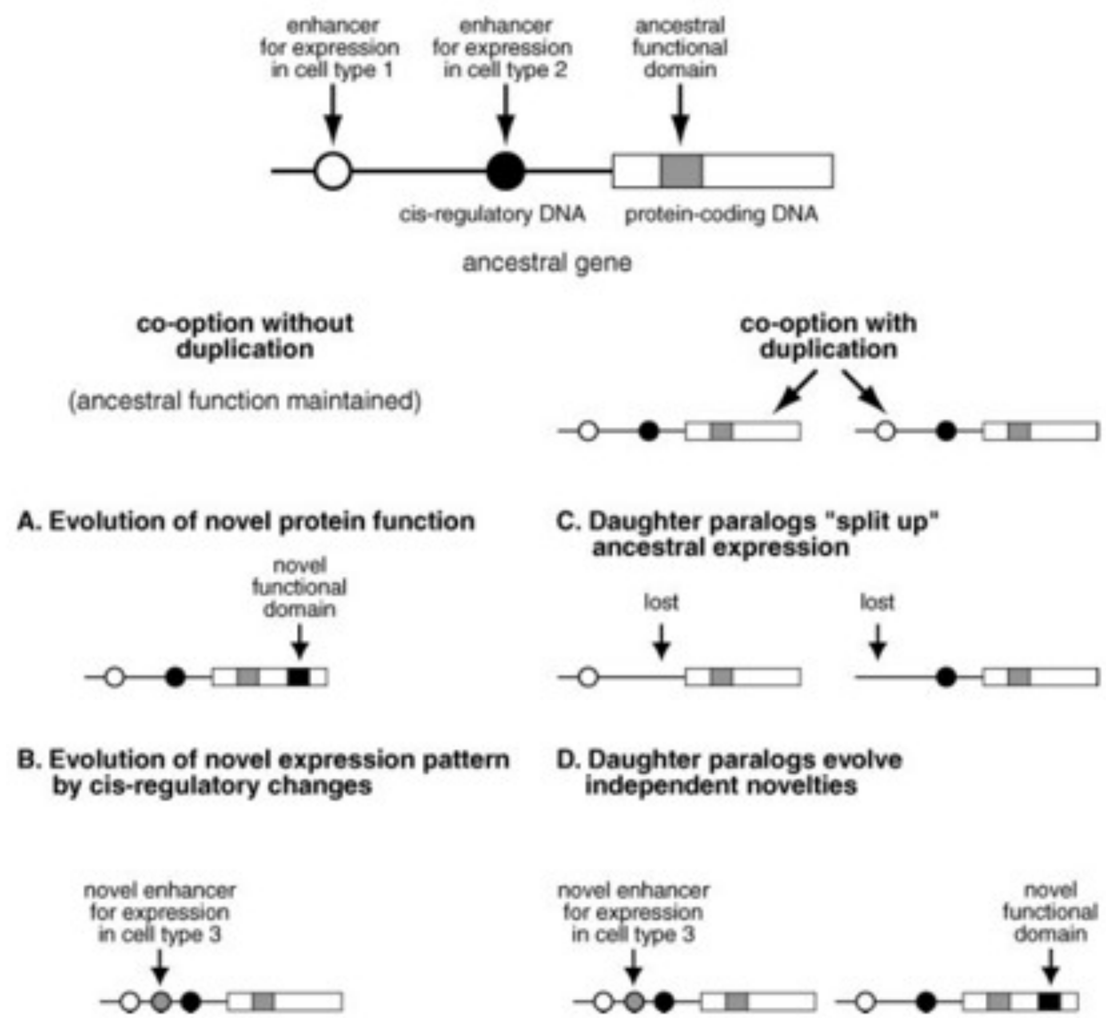
- **H3K27ac ChIP-seq on early embryonic limb tissues in human, rhesus and mouse, at various developmental stages**
- **Matched RNA-seq in future**

Human

Rhesus

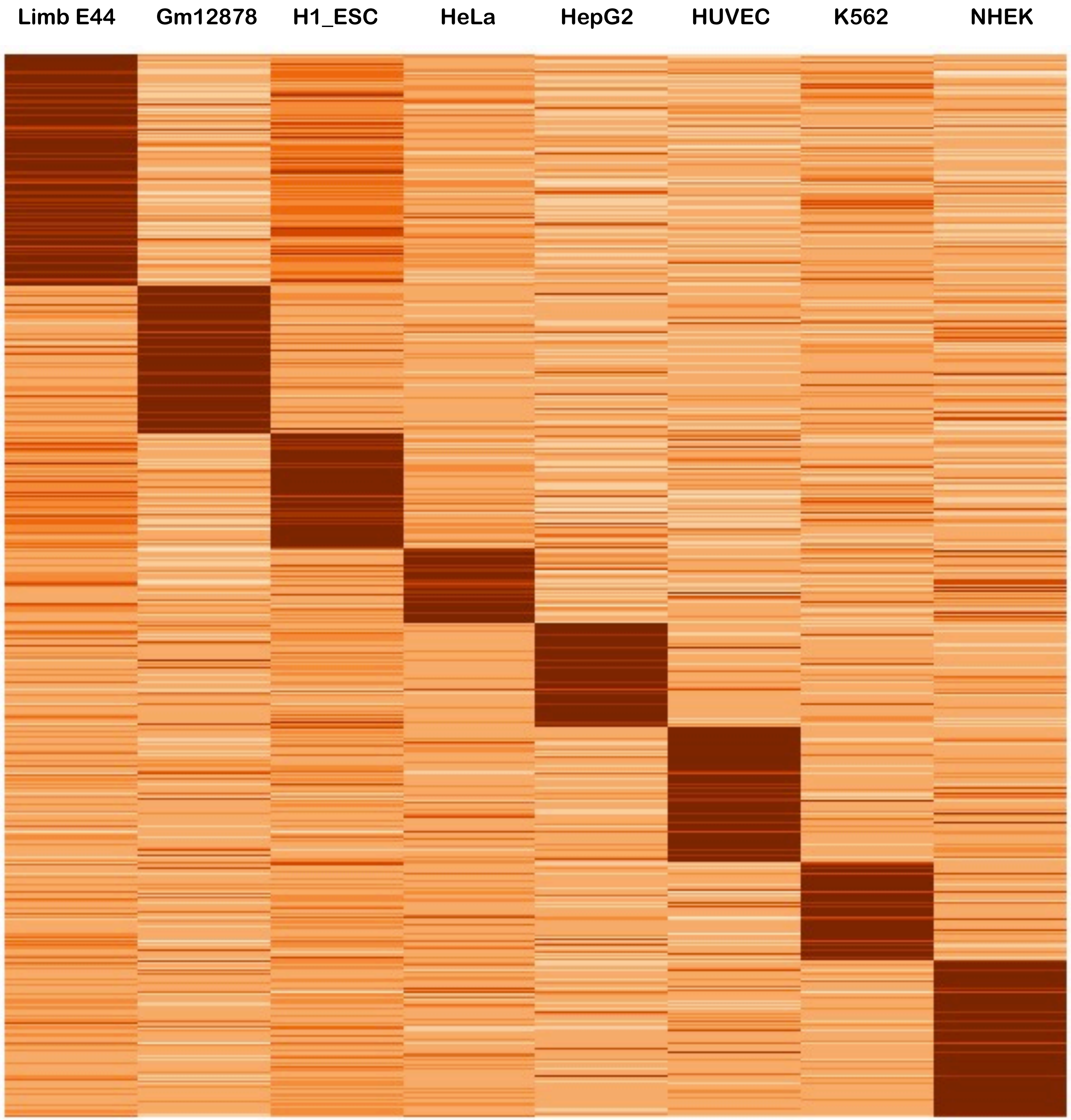
Mouse

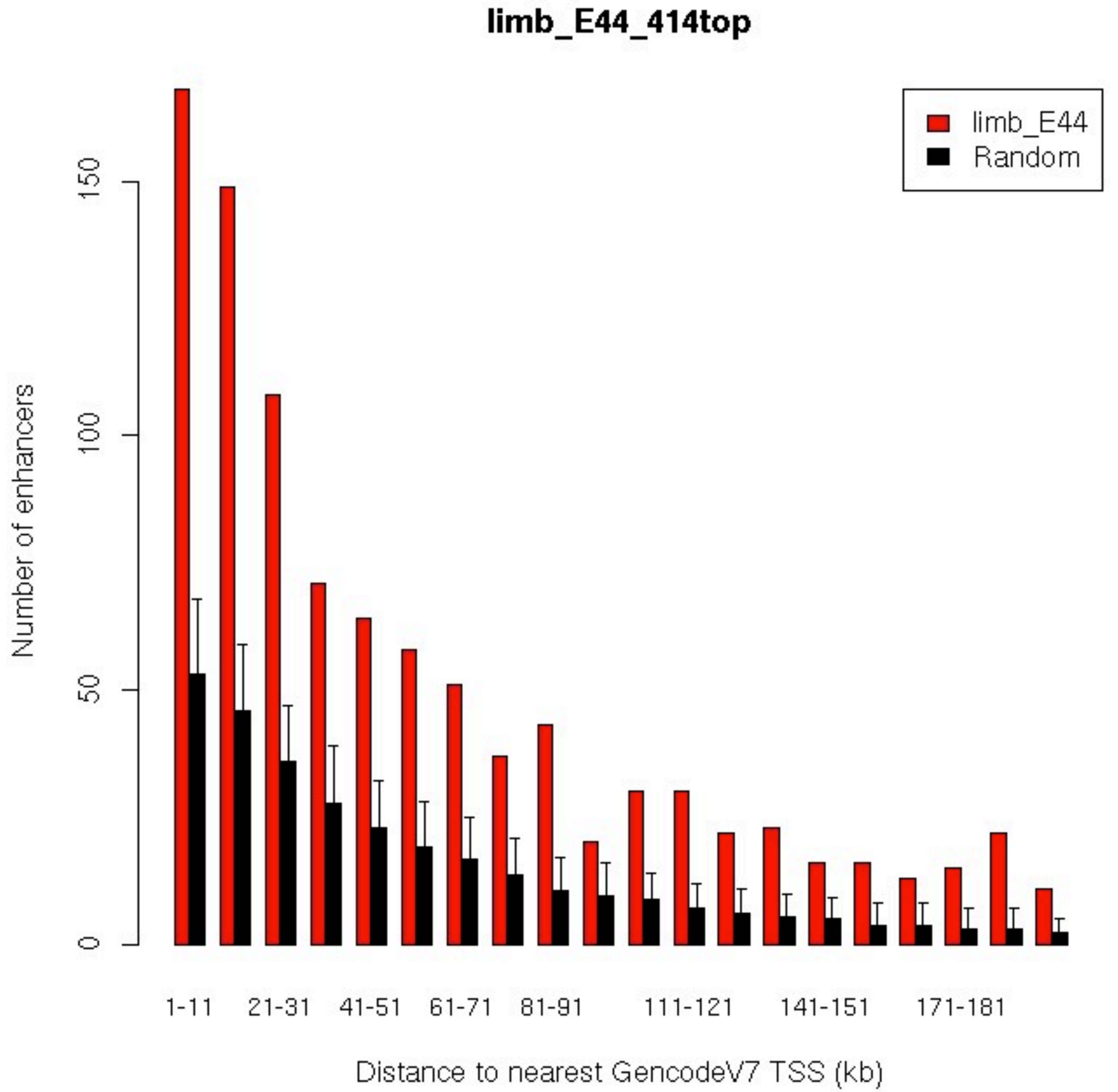


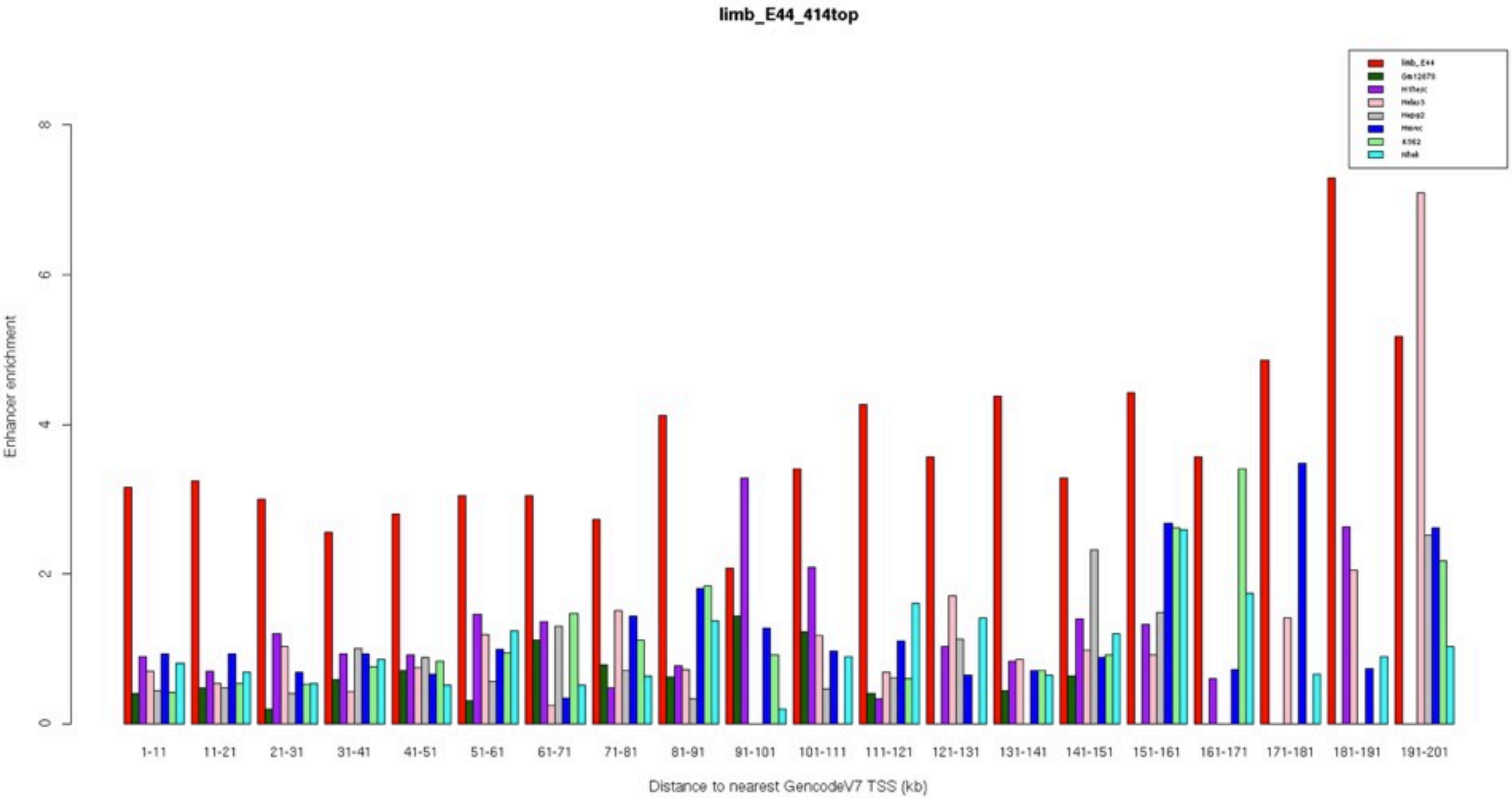
H3K27ac marking and gene expression ?

- Evidence for some validity of genomic proximity from tissue and temporal comparison in mouse
- Using signal correlation

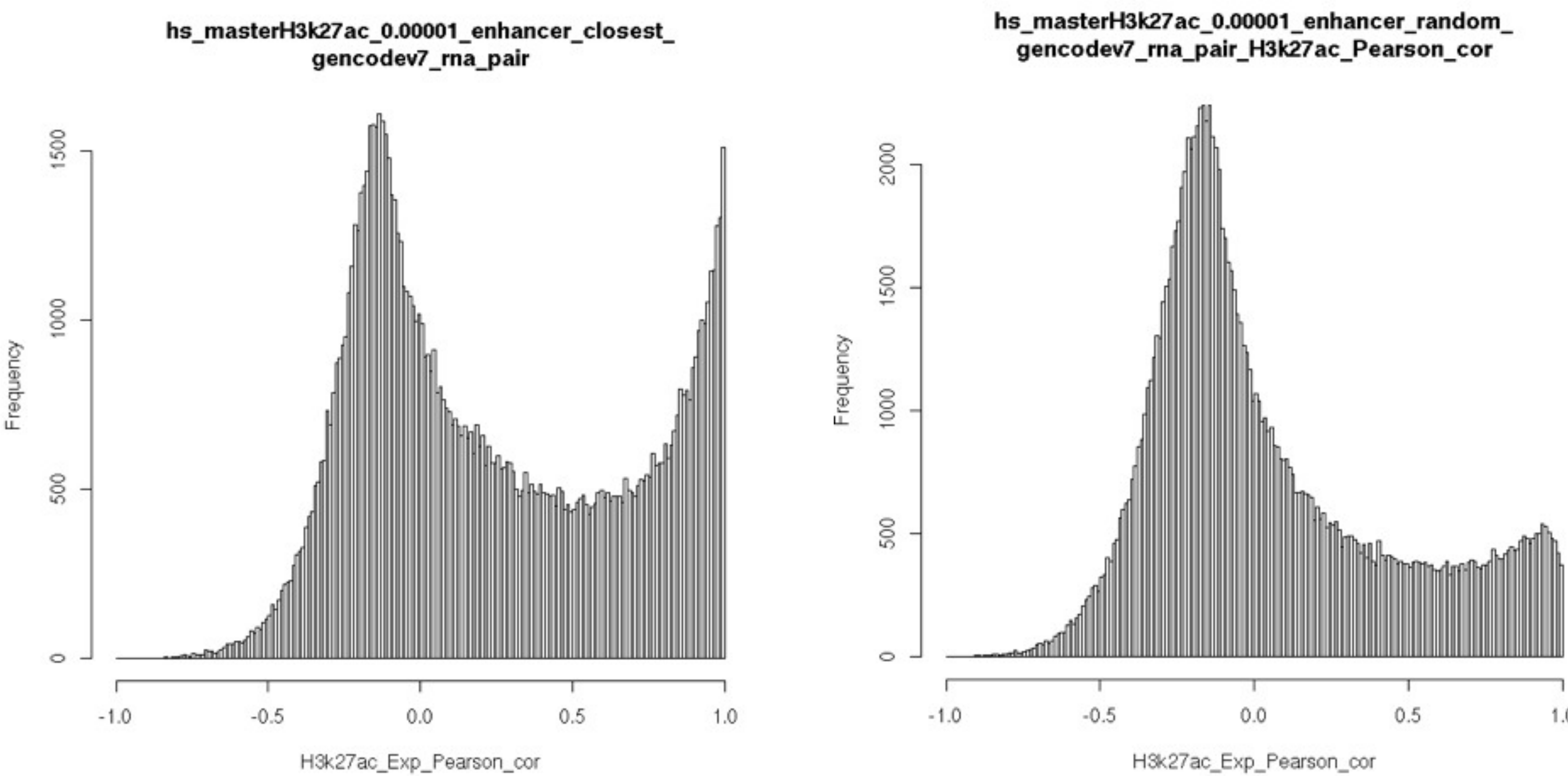
enhancer



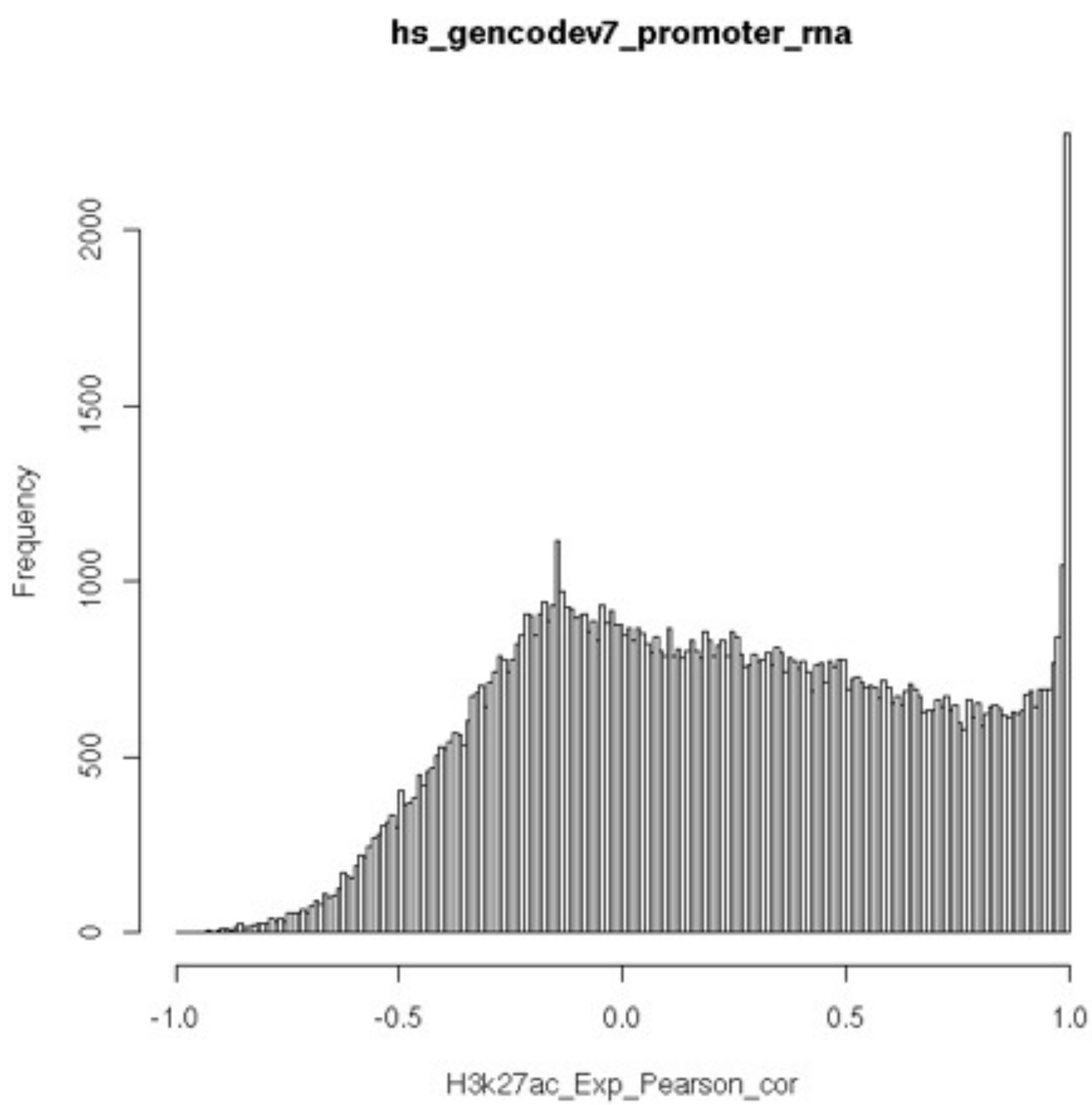

limb_E44_414top
limb_E44
Random
Number of enhancers
0
50
100
150
1-11
21-31
41-51
61-71
81-91
111-121
141-151
171-181
Distance to nearest GencodeV7 TSS (kb)

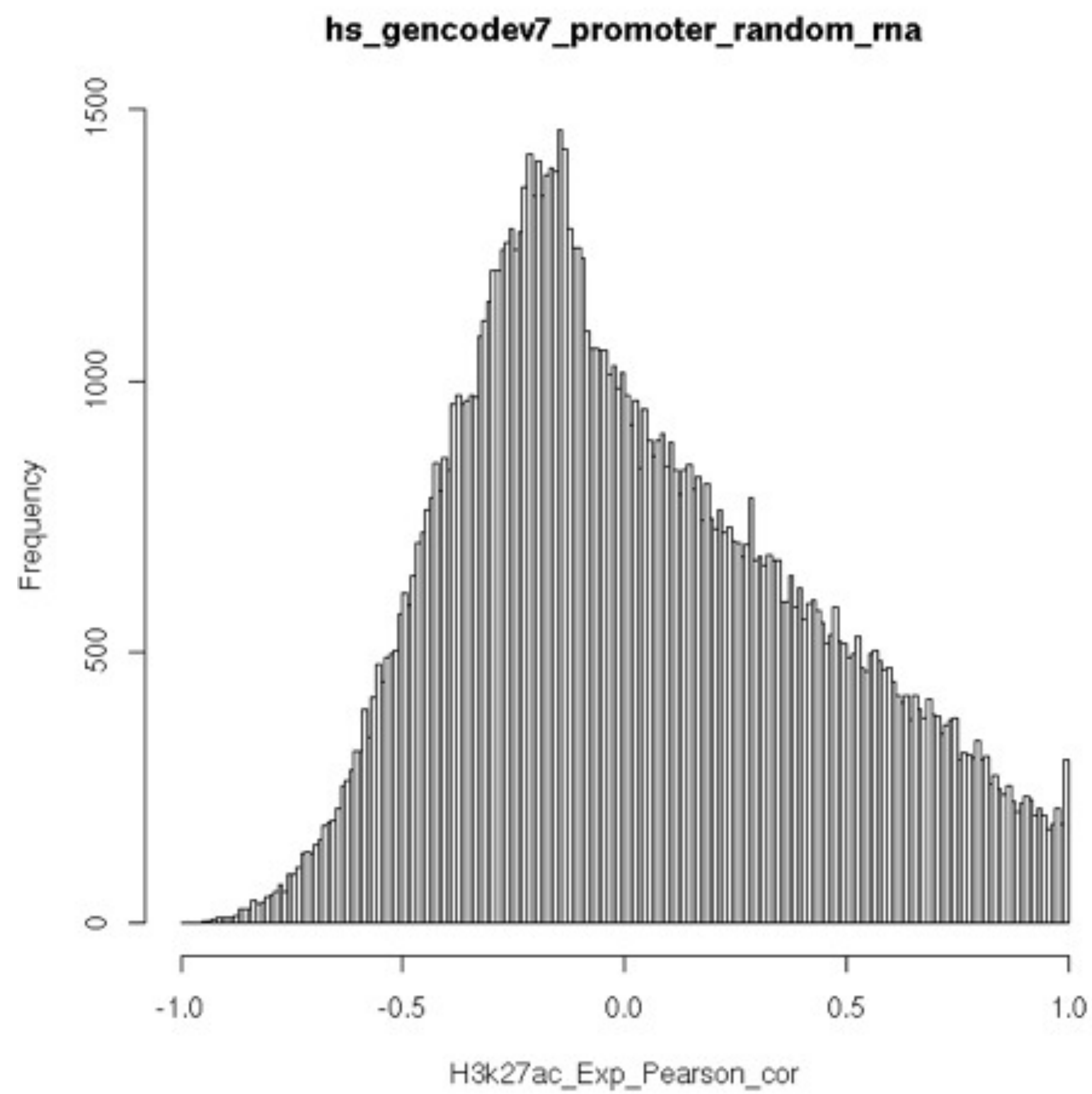

limb_E44_414top
limb_E44
Gm12878
H1hesc
Helas3
Hepg2
Huvec
K562
Nhek
Enhancer enrichment
0
2
4
6
8
1-11
11-21
21-31
31-41
41-51
51-61
61-71
71-81
81-91
91-101
101-111
111-121
121-131
131-141
141-151
151-161
161-171
171-181
181-191
191-201
Distance to nearest GencodeV7 TSS (kb)

Signal correlation in closest pairs: enhancer marking and TSS expression

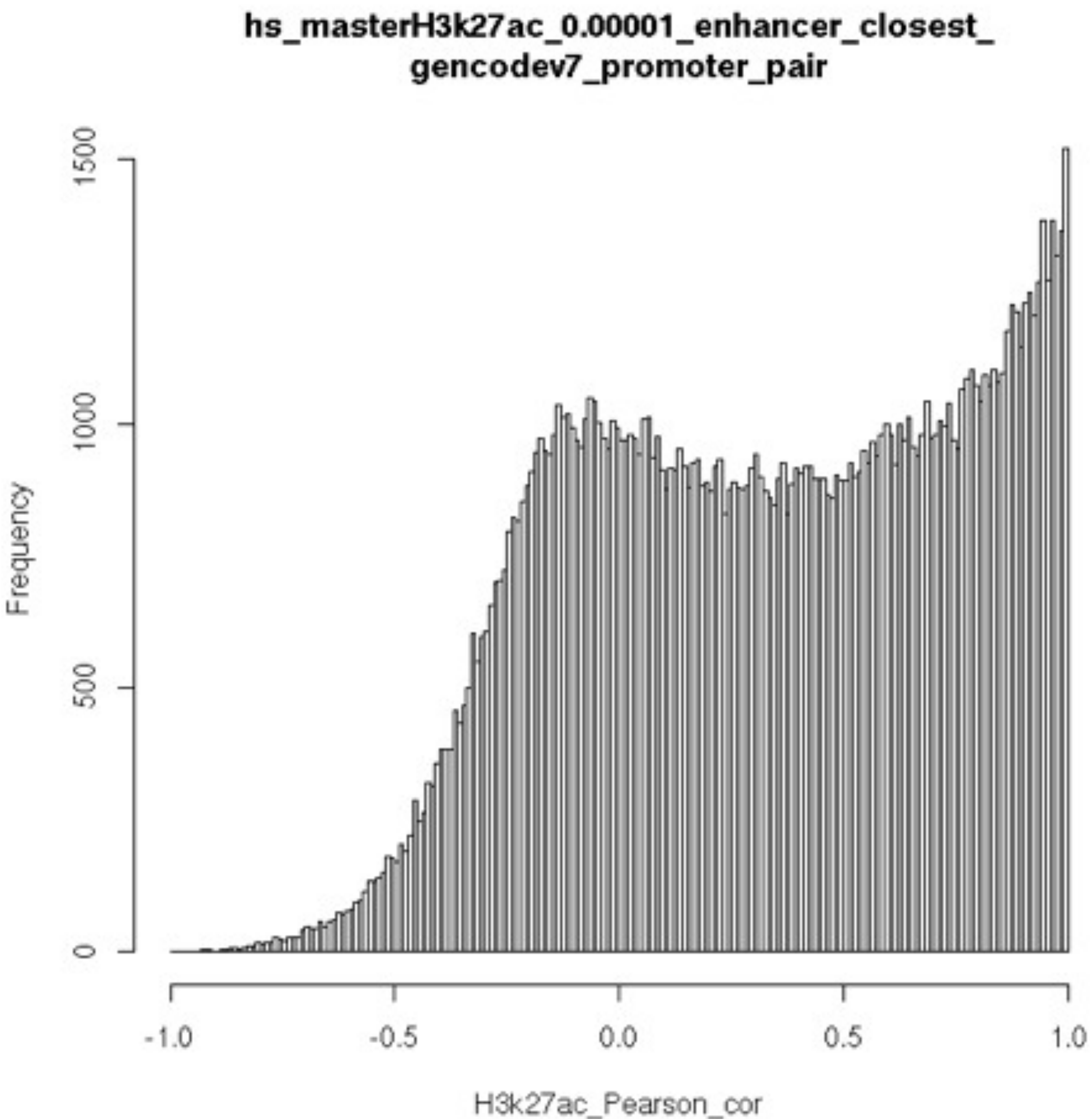


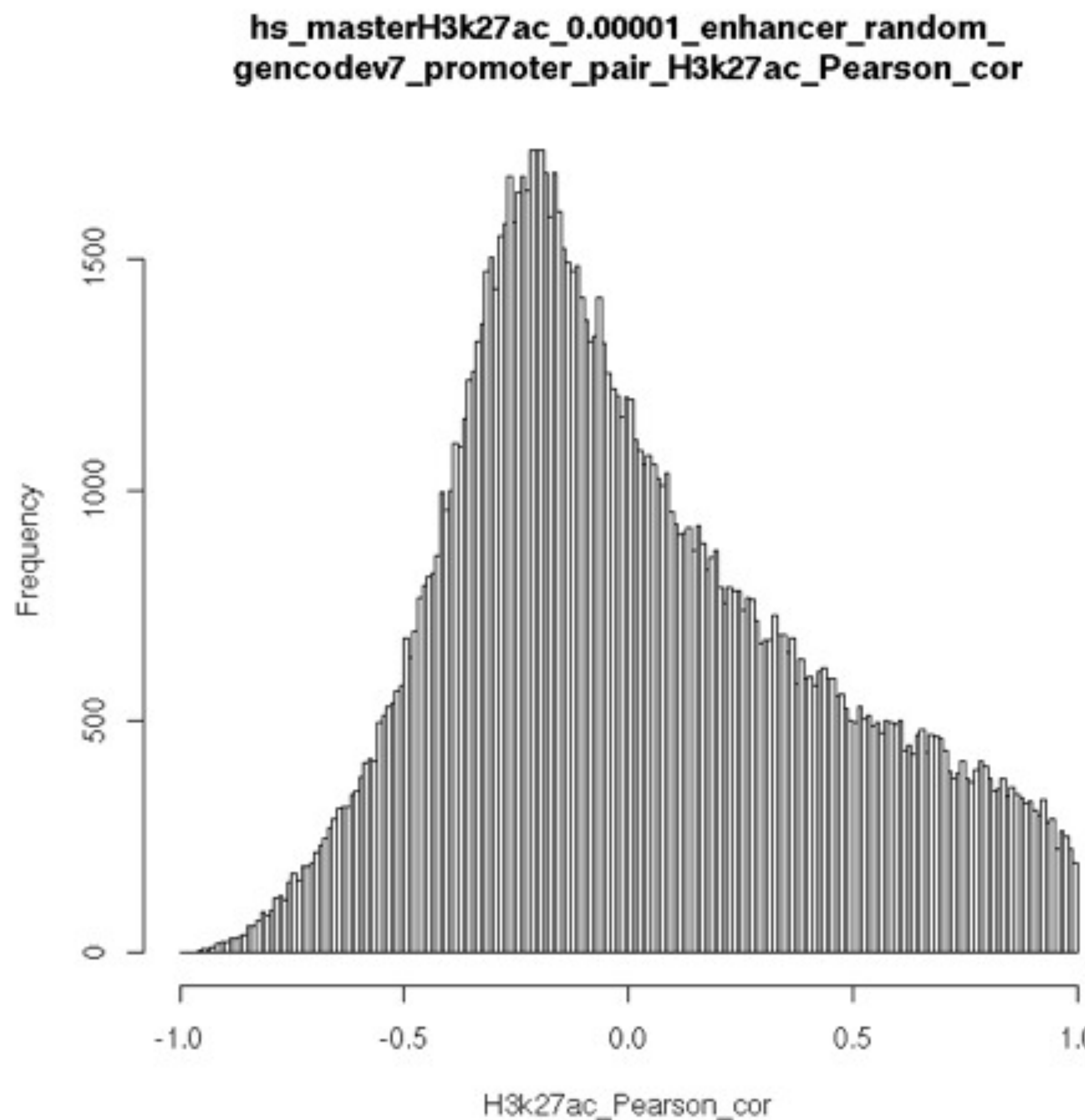
Signal correlation: promoter marking and TSS expression

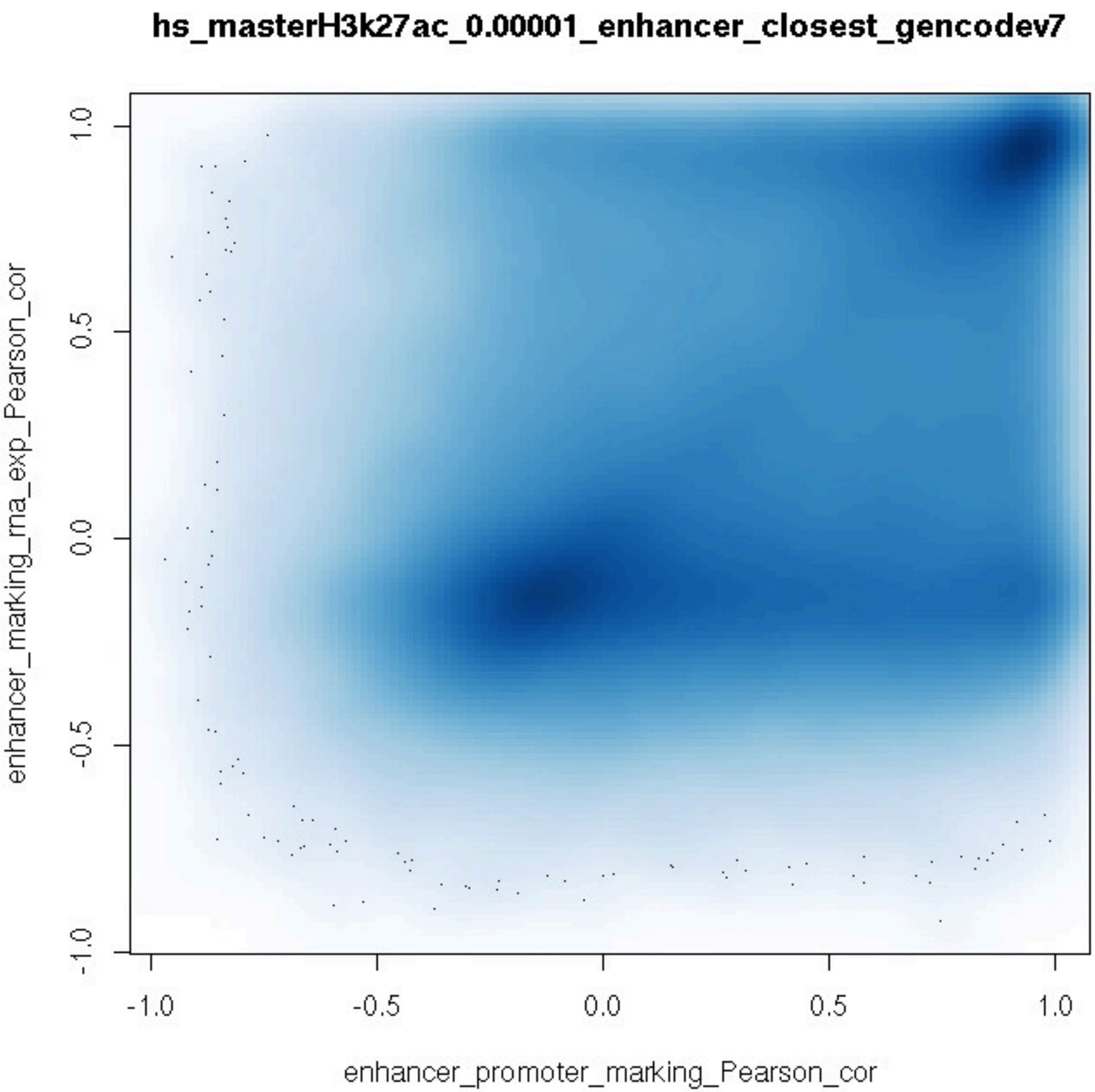




Signal correlation in closest pairs: enhancer and promoter marking







Spearman/Pearson_cor: 0.33

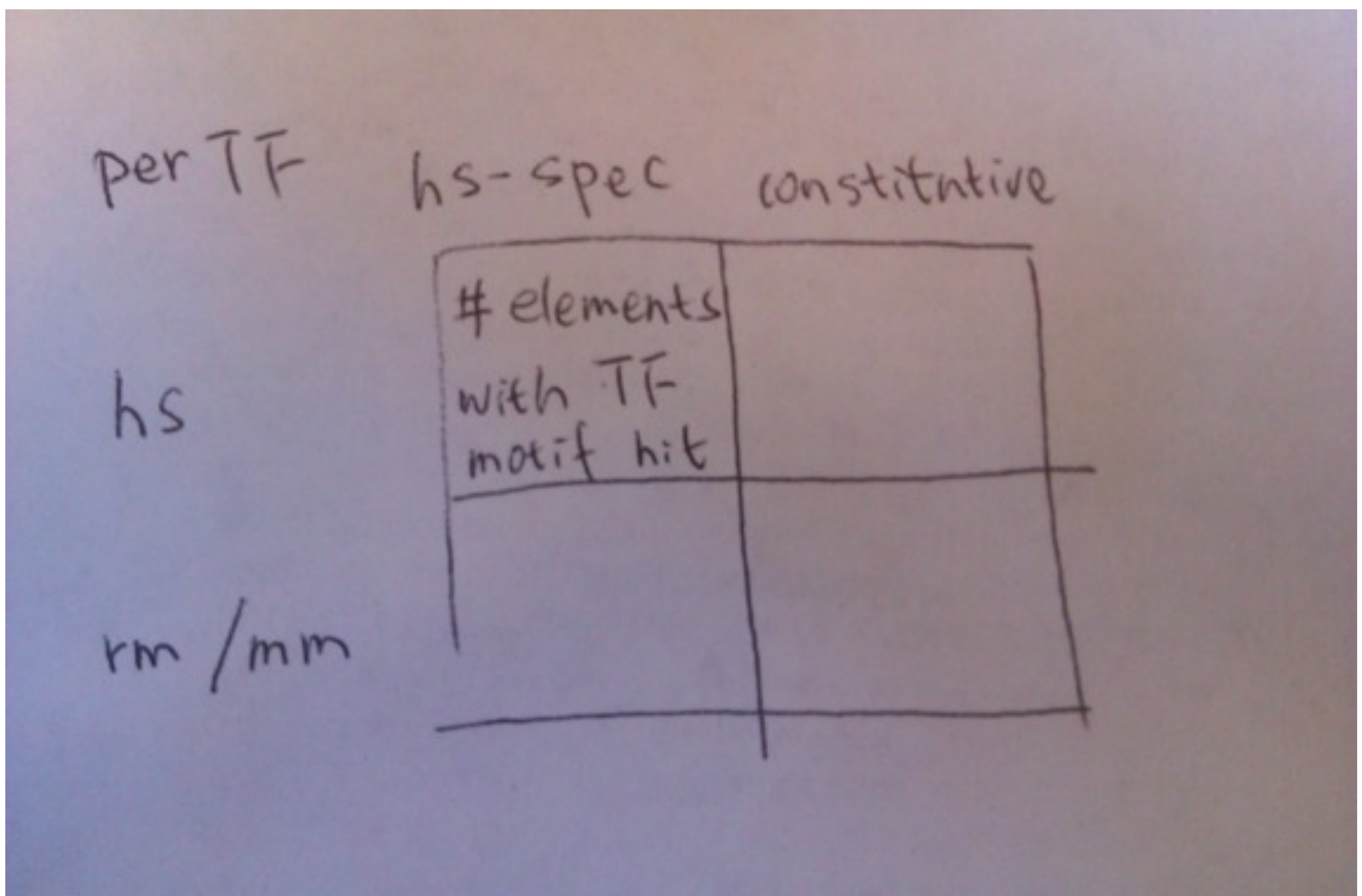
CRE cross-species dynamics

1. Event characterization
Call peaks and map orthologous CREs between species (Use pairwise genome chain files, and categorize elements into species-specific, reciprocally mappable, duplication-related)

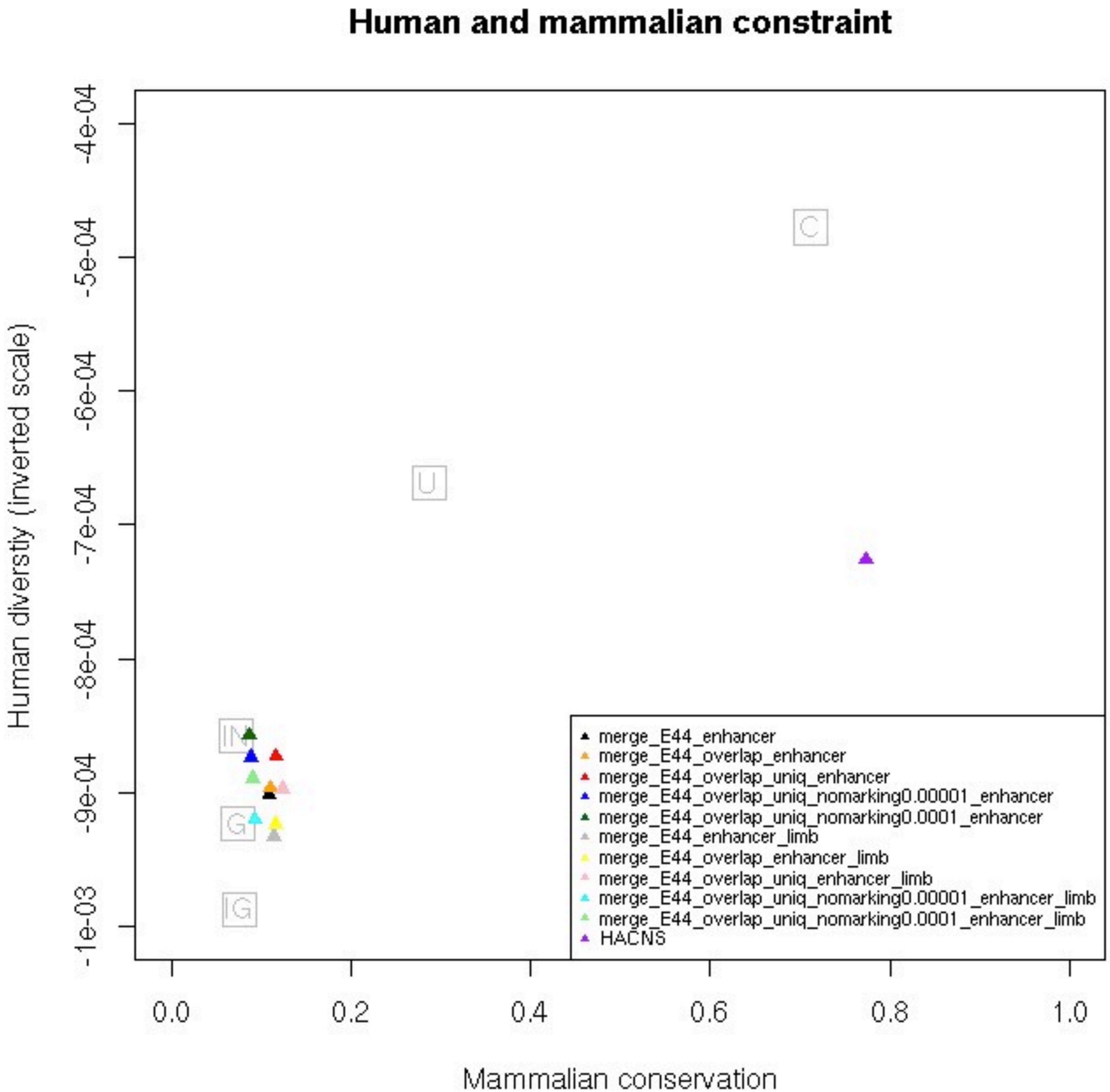
hg19	rm2	mm9	intersection	promoter	enhancer	HACNS
merge_E44_overlap	33963	33963	33963	14380	17026	47
fail	803	2762	251			
unique_chain	32365(32261)	30192(29701)	29124(28595)	12771(12606)	14021(13690)	47(47)
unique_chain_nomarking(e-5)	4895(4807)	8914(8593)	2514(2392)	343(328)	1888(1790)	3(3)
unique_chain_nomarking(e-4)	3381(3295)	7009(6716)	1512(1426)	194(183)	1153(1085)	1(1)
unique_chain_nomarking(e-3)	1995(1919)	4515(4268)	682(632)	82(75)	522(483)	0(0)
multiple_chain	795	1009	143			
multiple_chain_nomarking(e-5)	101	239	6			
multiple_chain_nomarking(e-4)	79	193	5			
multiple_chain_nomarking(e-3)	47	128	4			
merge_E47_overlap	31591	31592	31593	14292	14996	35
unique_chain	30083(29976)	28079(27654)	27058(26586)	12706(12537)	12256(11992)	35(35)
unique_chain_nomarking(e-5)	4415(4327)	7669(7388)	2239(2137)	346(336)	1620(1538)	3(3)
unique_chain_nomarking(e-4)	3062(2980)	5908(5655)	1299(1223)	199(191)	958(896)	1(1)
unique_chain_nomarking(e-3)	1810(1736)	3842(3630)	579(539)	81(77)	433(401)	1(1)

2. Genetic drivers of CRE dynamics: motif and TE

- E44 hs-specific enhancer gain: FIMO_hs_gain_hs_mm & FIMO_hs_gain_hs_rm
- mm_e12.5 hs-specific enhancer loss: FIMO_hs_loss_hs_mm



3. CRE age and constraint



E44 hs-specific enhancer & 1KG_het p-values
C1: merge_E44_enhancer
C2: merge_E44_overlap_enhancer
C3: merge_E44_overlap_uniq_enhancer
C4: merge_E44_overlap_uniq_nomarking0.00001_enhancer
C5: merge_E44_overlap_uniq_nomarking0.0001_enhancer

label permutation 1000 times two-sided | one-sided
C1-C2 VS C2: diff=1.25099e-05, p=0.089 | p=0.039
C2-C3 VS C3: diff=0.0001623484, p=0 | p=0
C3-C4 VS C4: diff=-2.933e-07, p=0.984 | p=0.496
C4-C5 VS C5: diff=3.88617e-05, p=0.294 | p=0.144

C1_nonlimb VS C1_limb: diff=-6.85693e-05, p=0 | p=0
C2_nonlimb VS C2_limb: diff=-5.77755e-05, p=0 | p=0
C3_nonlimb VS C3_limb: diff=-5.08017e-05, p=0 | p=0
C4_nonlimb VS C4_limb: diff=-0.0001125775, p=0 | p=0
C5_nonlimb VS C5_limb: diff=-8.03307e-05, p=0.111 | p=0.053

E44 hs-specific enhancer & phastCons p-values
C1: merge_E44_enhancer
C2: merge_E44_overlap_enhancer
C3: merge_E44_overlap_uniq_enhancer
C4: merge_E44_overlap_uniq_nomarking0.00001_enhancer
C5: merge_E44_overlap_uniq_nomarking0.0001_enhancer

label permutation 1000 times two-sided | one-sided
C1-C2 VS C2: diff=-0.002911497, p=0.032 | p=0.017
C2-C3 VS C3: diff=-0.03359694, p=0 | p=0
C3-C4 VS C4: diff=0.02745162, p=0 | p=0
C4-C5 VS C5: diff=0.004911789, p=0.479 | p=0.255

C1_nonlimb VS C1_limb: diff=-0.01225132, p=0 | p=0
C2_nonlimb VS C2_limb: diff=-0.01235962, p=0 | p=0
C3_nonlimb VS C3_limb: diff=-0.01487142, p=0 | p=0
C4_nonlimb VS C4_limb: diff=-0.01035097, p=0.099 | p=0.052
C5_nonlimb VS C5_limb: diff=-0.006129281, p=0.493 | p= 0.232

E44 hs-specific enhancer & phastCons.placental p-values
C1: merge_E44_enhancer
C2: merge_E44_overlap_enhancer
C3: merge_E44_overlap_uniq_enhancer
C4: merge_E44_overlap_uniq_nomarking0.00001_enhancer
C5: merge_E44_overlap_uniq_nomarking0.0001_enhancer

label permutation 1000 times two-sided | one-sided
C1-C2 VS C2: diff=-0.003011569, p=0.022 | p=0.014
C2-C3 VS C3: diff=-0.04528415, p=0 | p=0
C3-C4 VS C4: diff=0.02999886, p=0 | p=0
C4-C5 VS C5: diff=0.008712642, p=0.357 | p=0.176

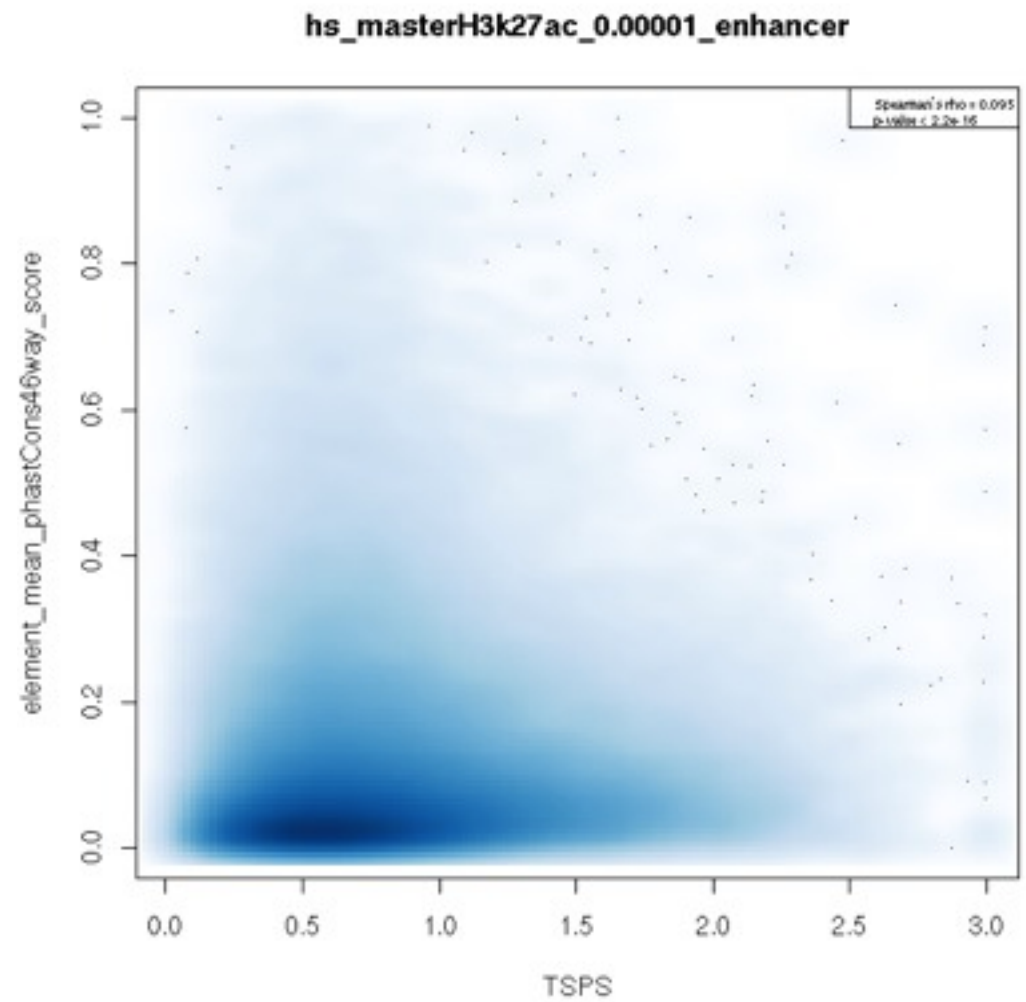
C1_nonlimb VS C1_limb: diff=-0.01117592, p=0 | p=0
C2_nonlimb VS C2_limb: diff=-0.01141631, p=0 | p=0
C3_nonlimb VS C3_limb: diff=-0.01399019, p=0 | p=0
C4_nonlimb VS C4_limb: diff=-0.01070869, p=0.123 | p=0.067
C5_nonlimb VS C5_limb: diff=-0.006847943, p=0.474 | p= 0.233

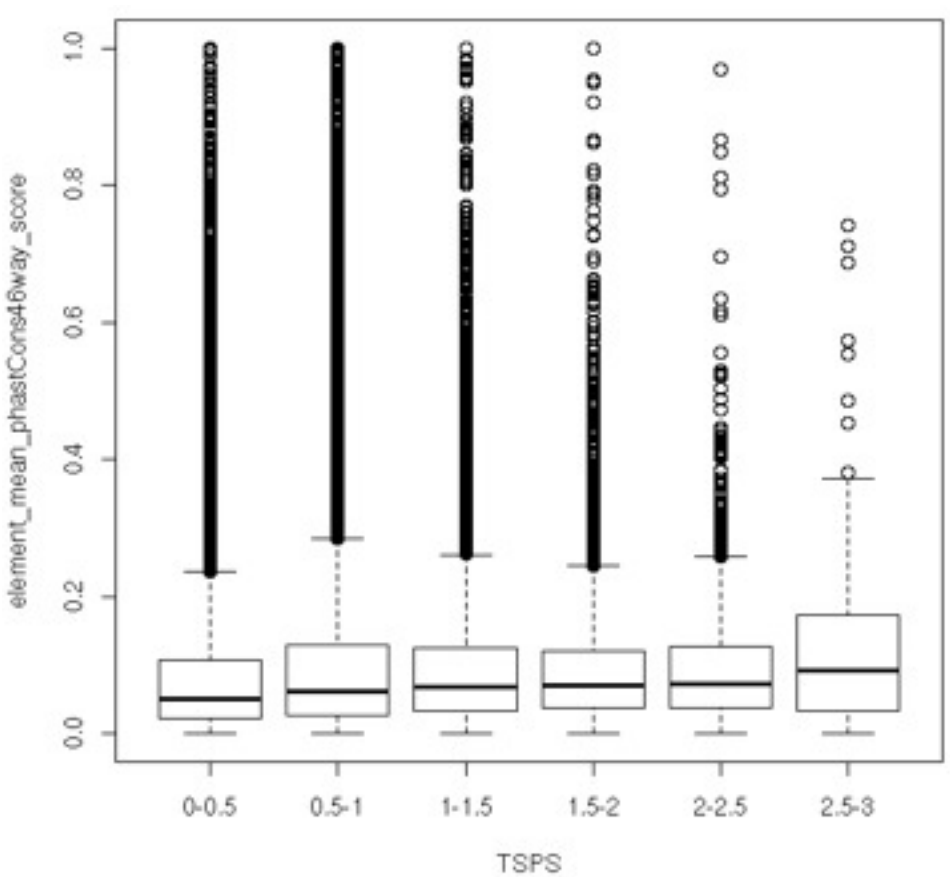
E44 hs-specific enhancer & phastCons.primates p-values
C1: merge_E44_enhancer
C2: merge_E44_overlap_enhancer
C3: merge_E44_overlap_uniq_enhancer
C4: merge_E44_overlap_uniq_nomarking0.00001_enhancer
C5: merge_E44_overlap_uniq_nomarking0.0001_enhancer

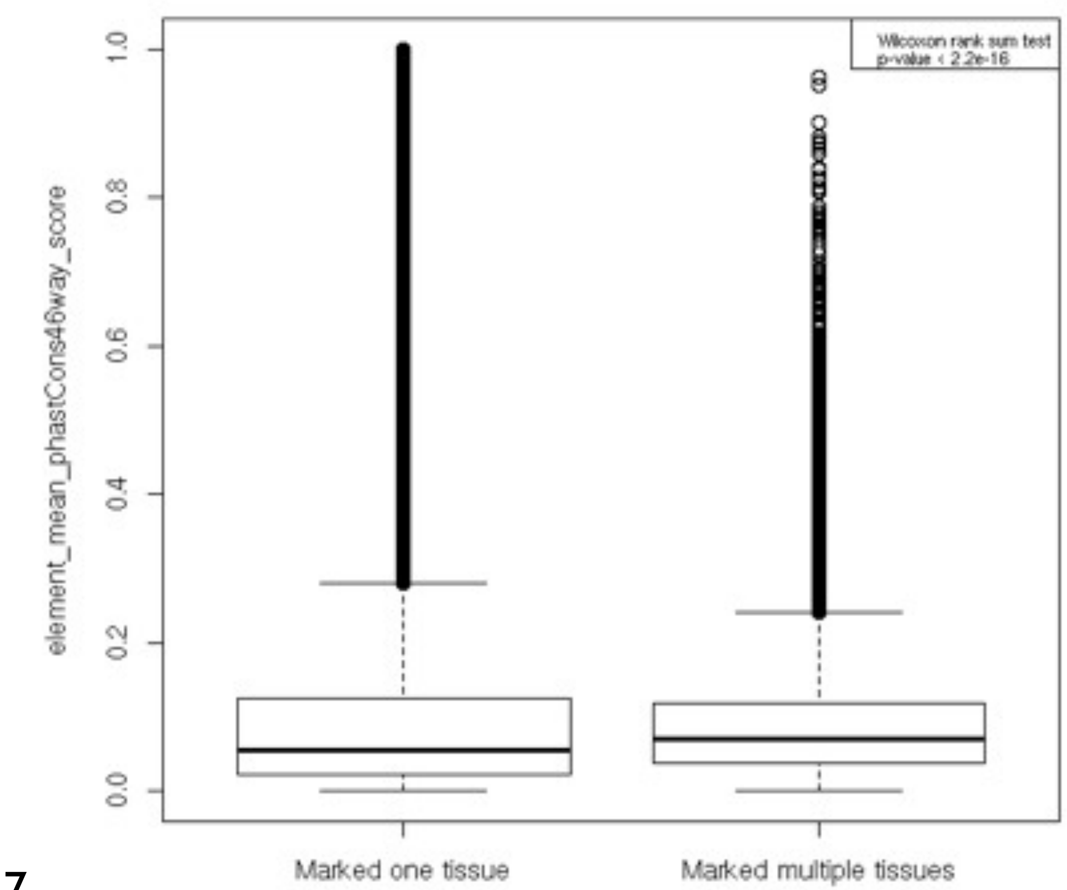
label permutation 1000 times two-sided | one-sided
C1-C2 VS C2: diff=-0.005337331, p=0 | p=0
C2-C3 VS C3: diff=-0.05894642, p=0 | p=0
C3-C4 VS C4: diff=0.03073280, p=0 | p=0
C4-C5 VS C5: diff=0.007728525, p=0.282 | p=0.14

C1_nonlimb VS C1_limb: diff=-0.002880342, p=0.033 | p=0.014
C2_nonlimb VS C2_limb: diff=-0.003786427, p=0.072 | p=0.036
C3_nonlimb VS C3_limb: diff=-0.006386742, p=0.01 | p=0.006
C4_nonlimb VS C4_limb: diff=-0.003348254, p=0.635 | p=0.304
C5_nonlimb VS C5_limb: diff=-0.002775451, p=0.78 | p= 0.364

TSPS & phastCons

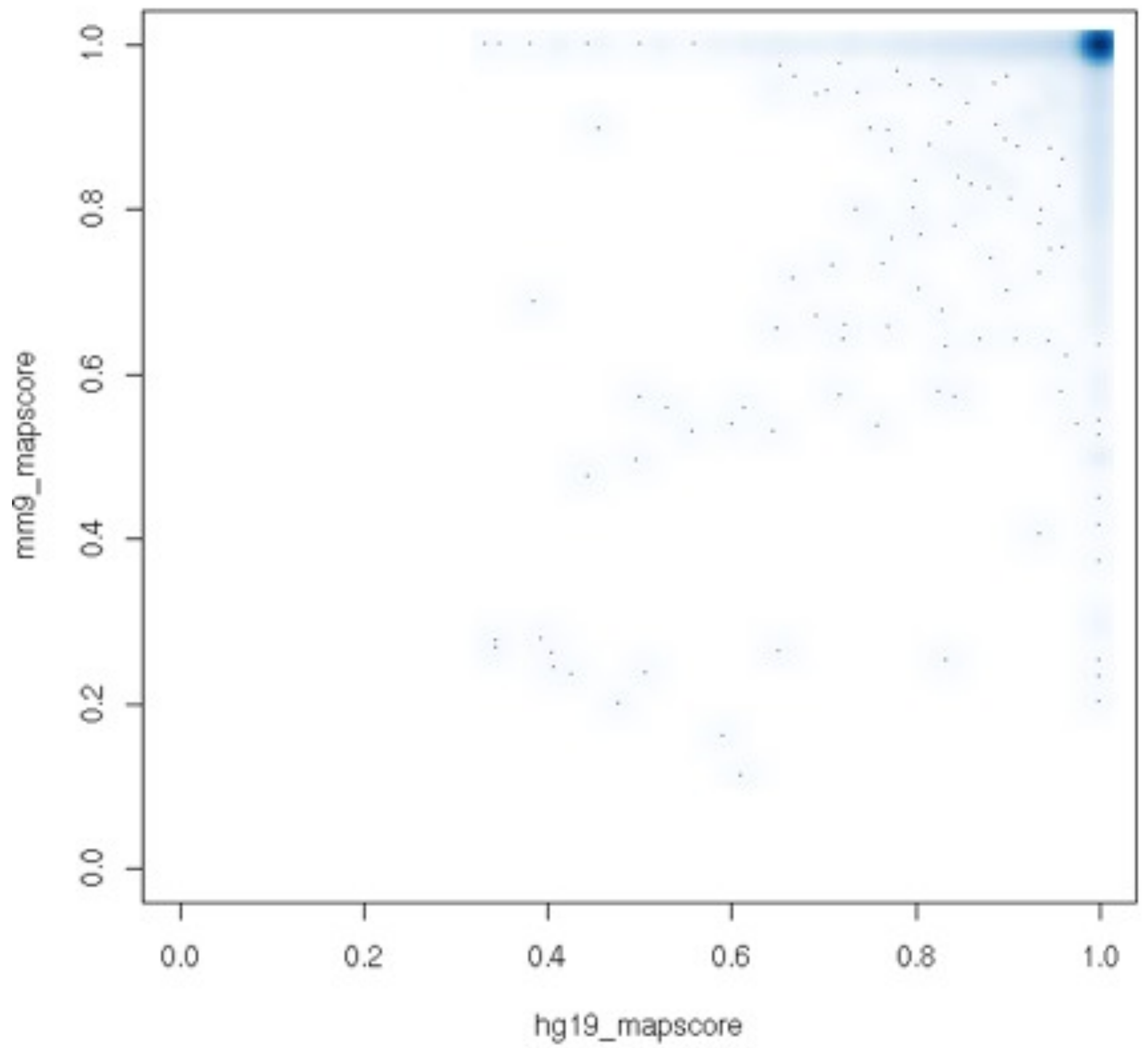






Cross-species alternative-splicing

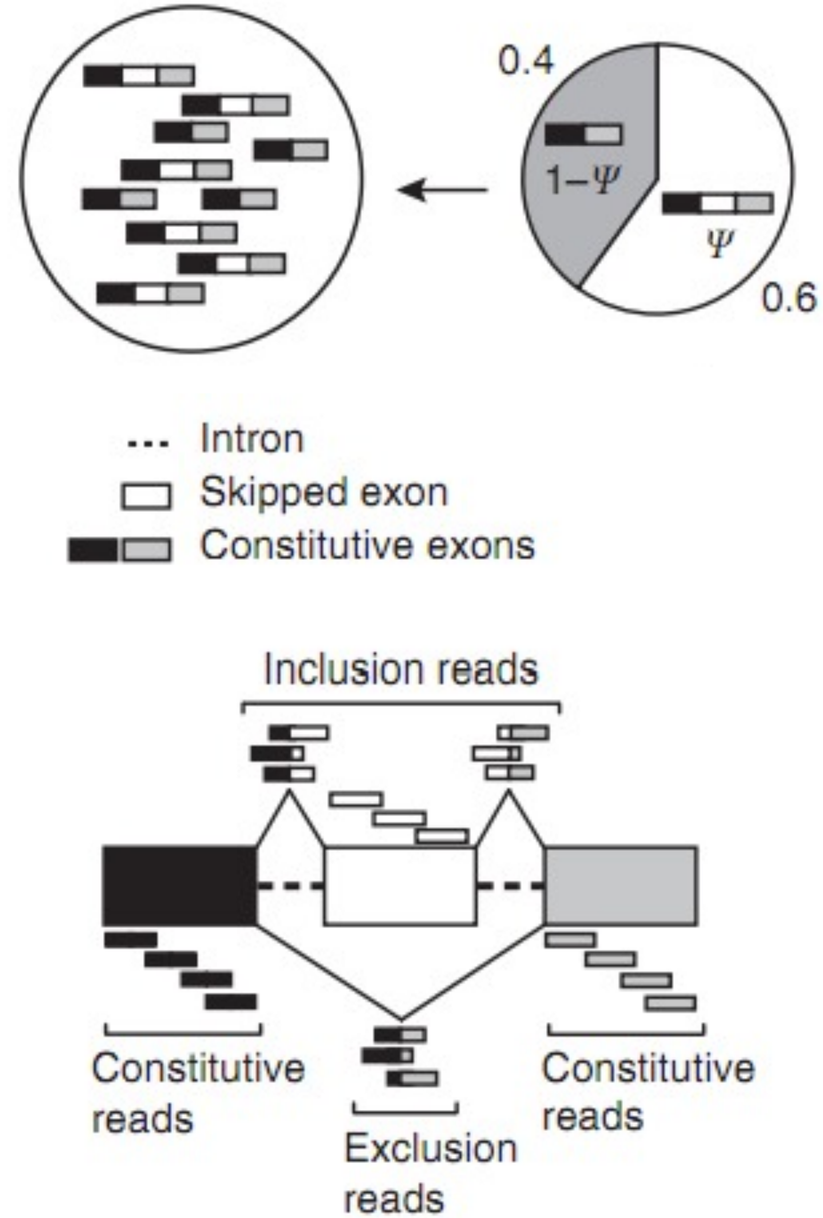
- Method to consider : mappability difference across species ...



- AS event-based analyses: 1. relative expression level-based 2. read-based

AS event-based

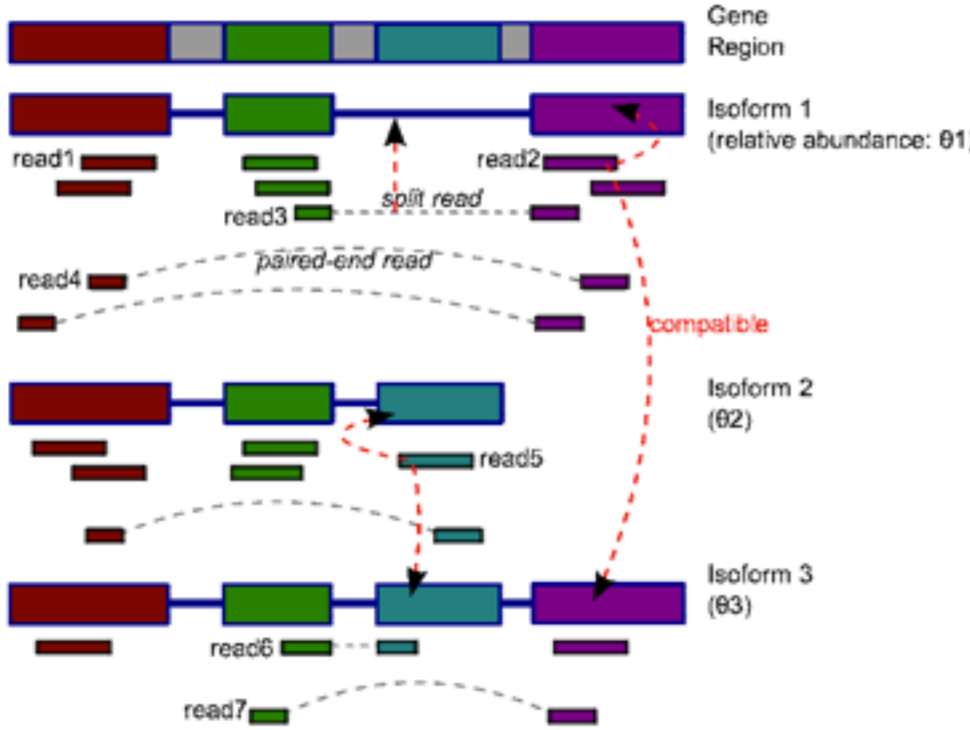
Alternative transcript events	Total events (×10³)	Number detected (×10³)	Both isoforms detected	Number tissue-regulated	% Tissue-regulated (observed)	% Tissue-regulated (estimated)
Skipped exon	37	35	10,436	6,822	65	72
Retained intron	1	1	167	96	57	71
Alternative 5′ splice site (A5SS)	15	15	2,168	1,386	64	72
Alternative 3′ splice site (A3SS)	17	16	4,181	2,655	64	74
Mutually exclusive exon (MXE)	4	4	167	95	57	66
Alternative first exon (AFE)	14	13	10,281	5,311	52	63
Alternative last exon (ALE)	9	8	5,246	2,491	47	52
Tandem 3′ UTRs	7	7	5,136	3,801	74	80
Total	105	100	37,782	22,657	60	68

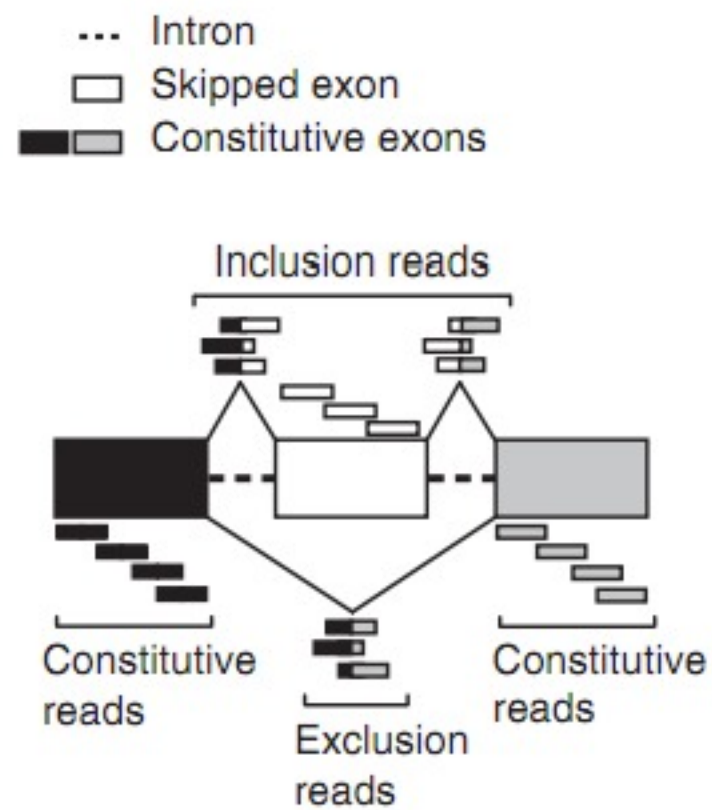


C. B. Burge group

1. relative expression level-based

IQSeq





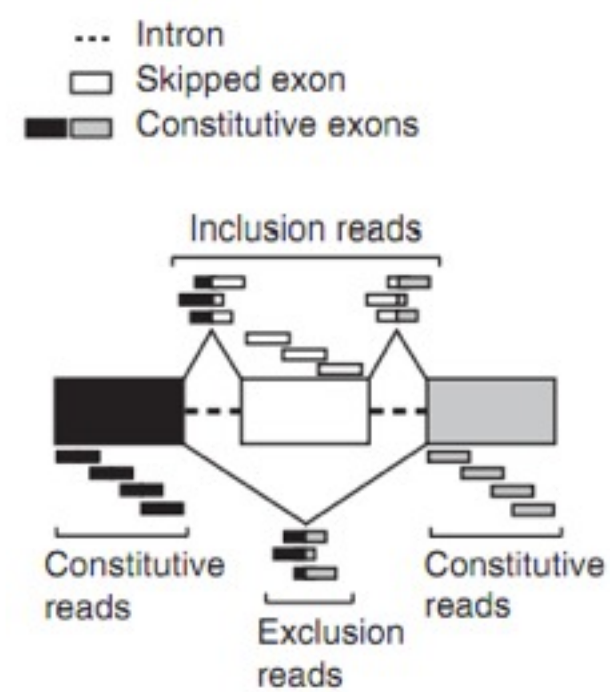
- Construct all "possible" exon-skipping events
- Run IQSeq to compute local exon inclusion/exclusion rate
- T-test for inclusion/exclusion rate in pair-wise species comparison

- Fisher exact test of reads supporting two exclusive events, Chi-square test of reads distribution in three exons
- How to treat replicates: Log-linear model from Bullard et al. (total reads -> total reads mapped to two exclusive events; reads in a gene -> reads supporting either event)

Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments

James H Bullard[1*†], Elizabeth Purdom[2†], Kasper D Hansen[1], Sandrine Dudoit[1,2]

$$\log(\mathrm{E}[X_{i,j} \mid d_i]) = \log d_i + \lambda_{a(i),j} + \theta_{i,j},$$



Cross-species alternative-splicing : dups?