

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



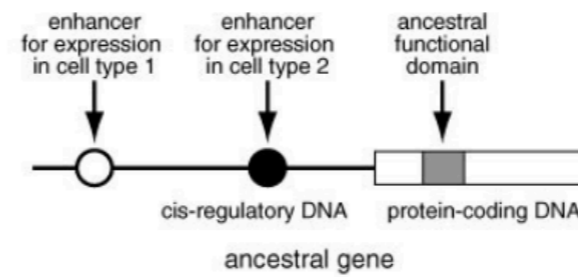
Human



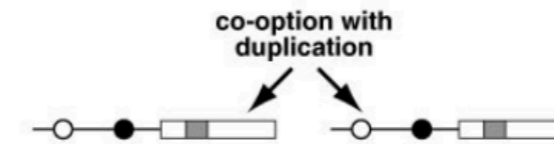
Rhesus



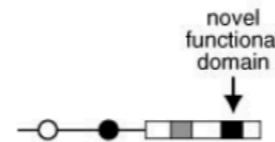
Mouse



**co-option without duplication**  
(ancestral function maintained)



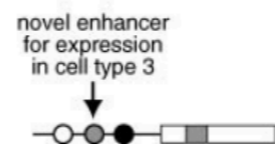
**A. Evolution of novel protein function**



**C. Daughter paralogs "split up" ancestral expression**



**B. Evolution of novel expression pattern by cis-regulatory changes**



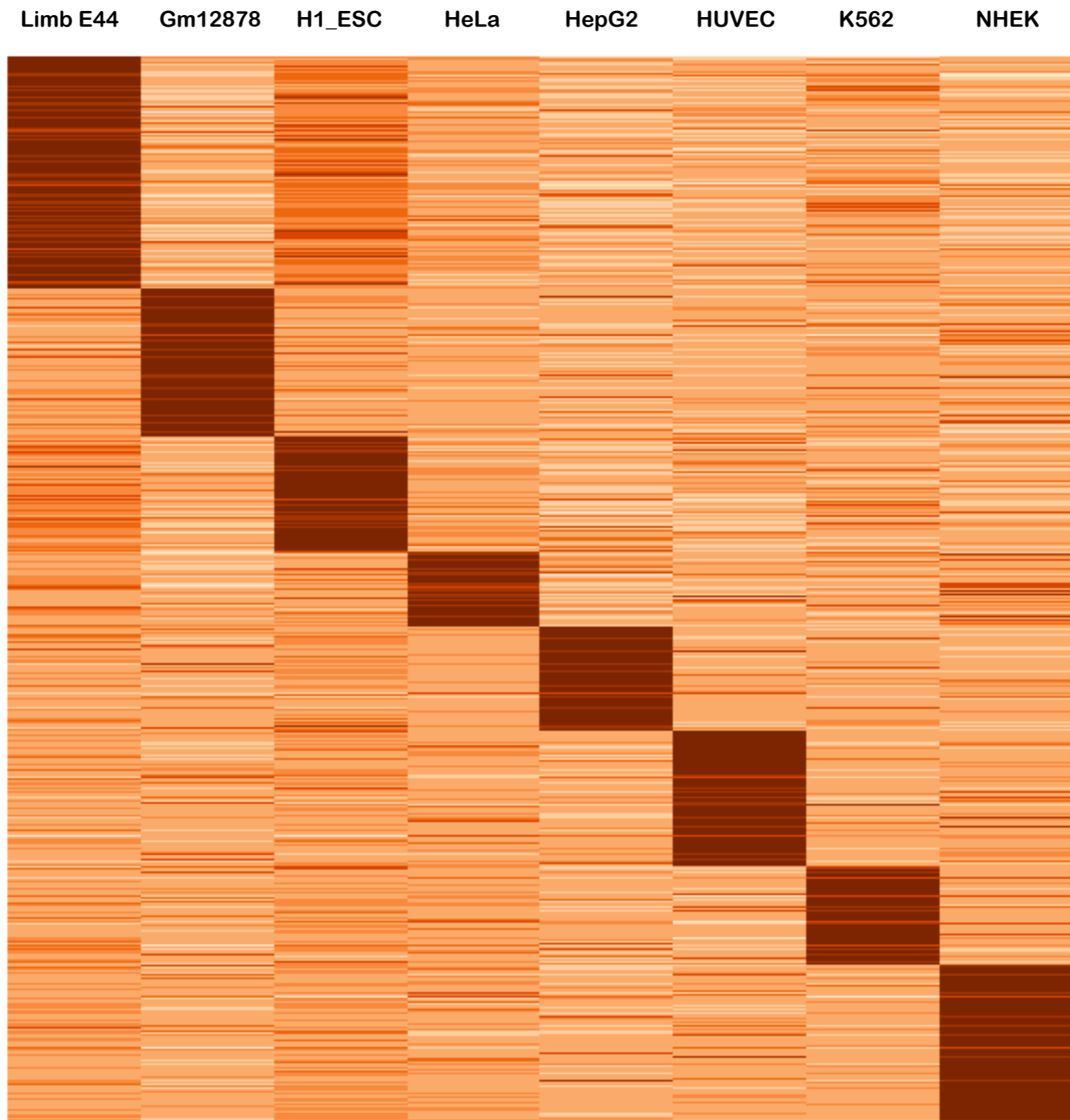
**D. Daughter paralogs evolve independent novelties**



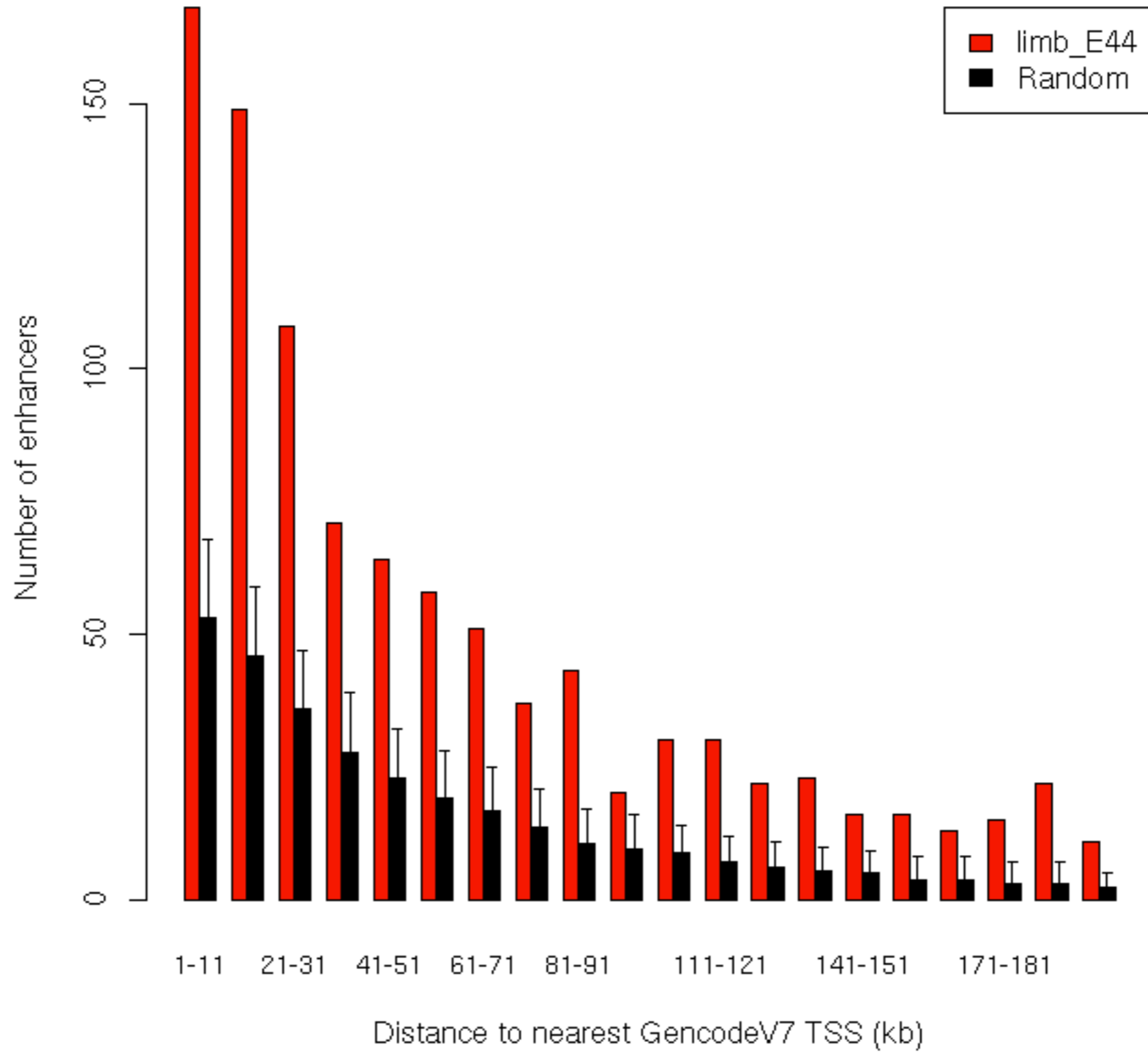
## H3K27ac marking and gene expression ?

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

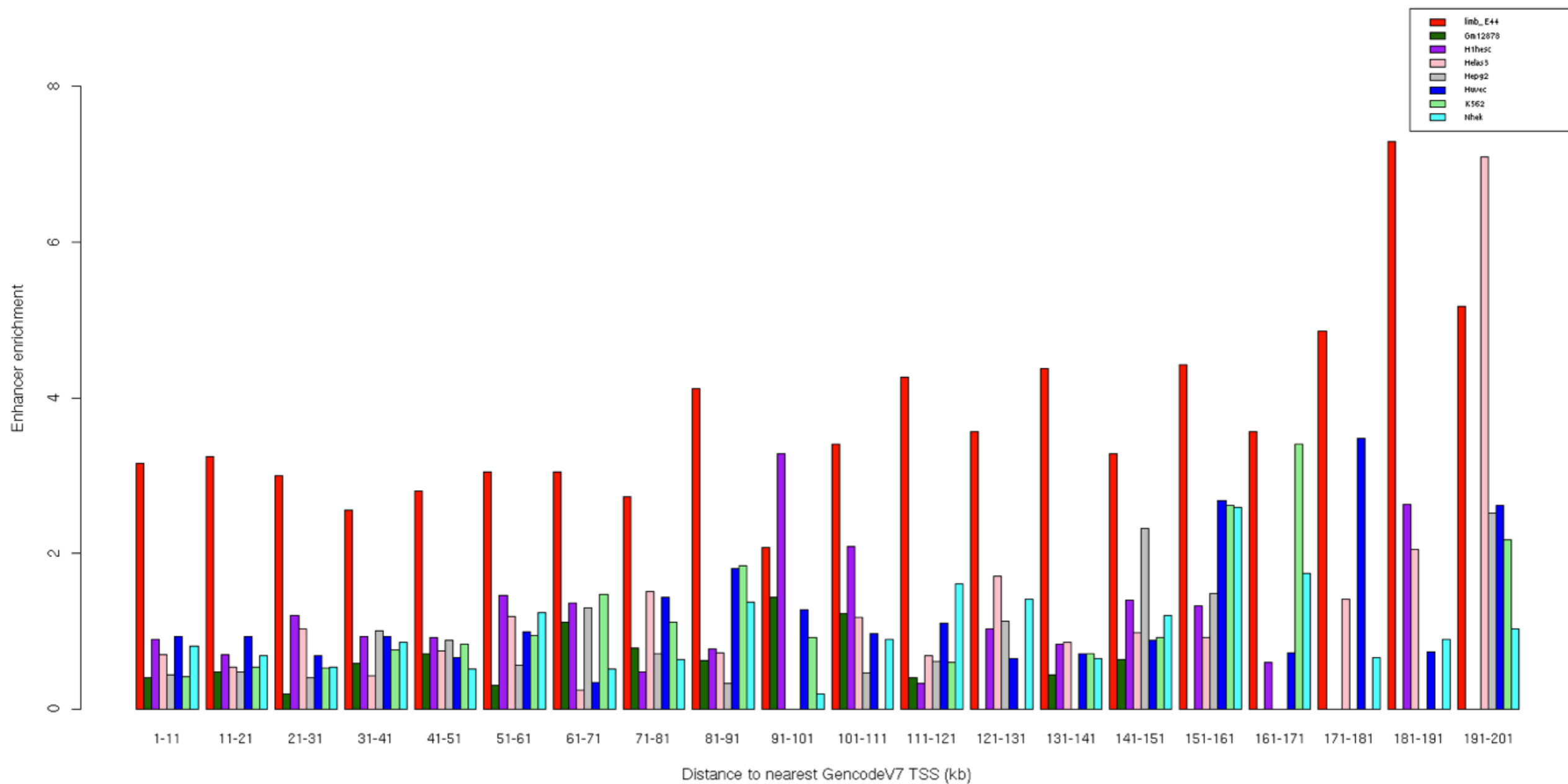
# enhancer



# limb\_E44\_414top

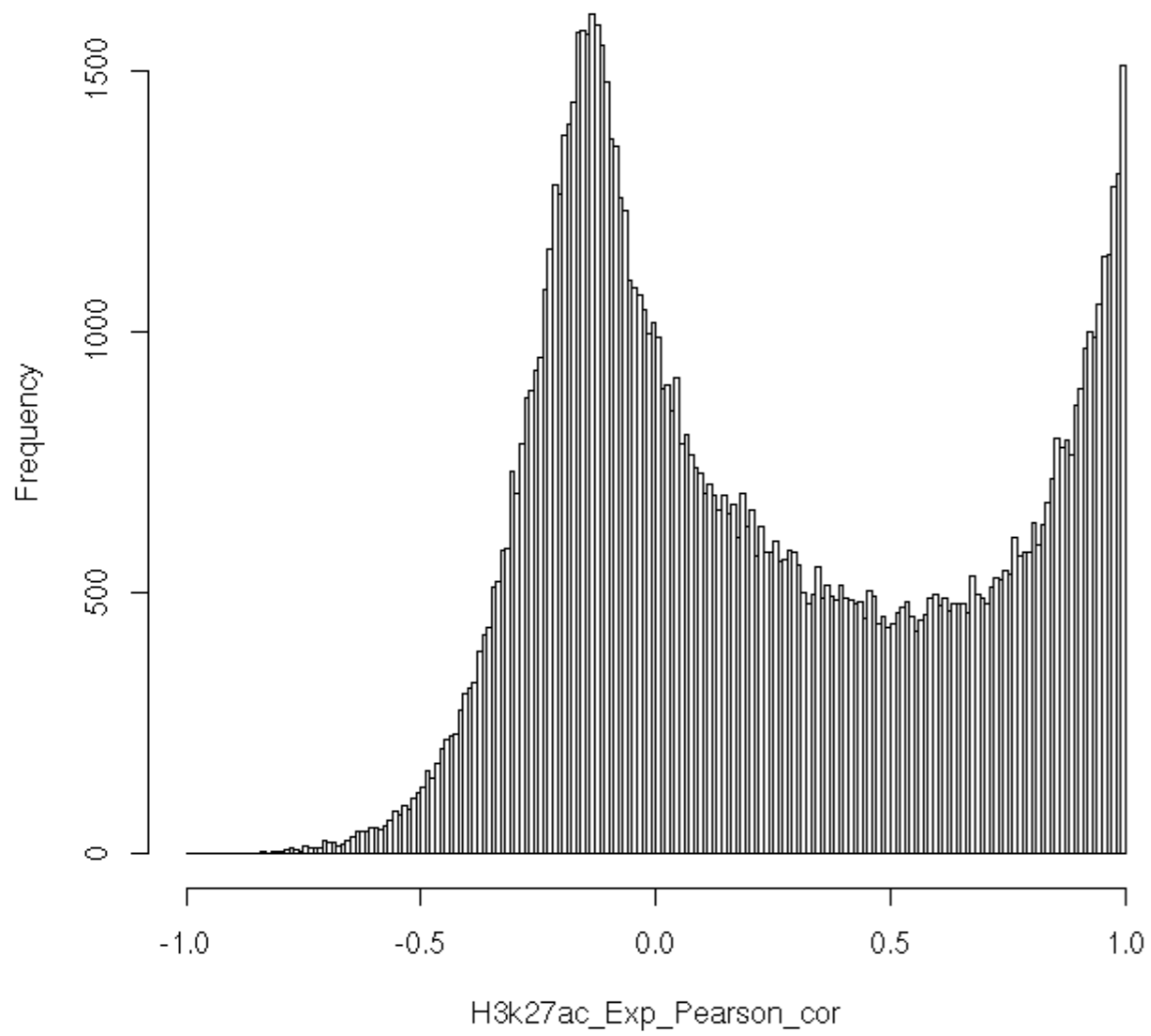


limb\_E44\_414top

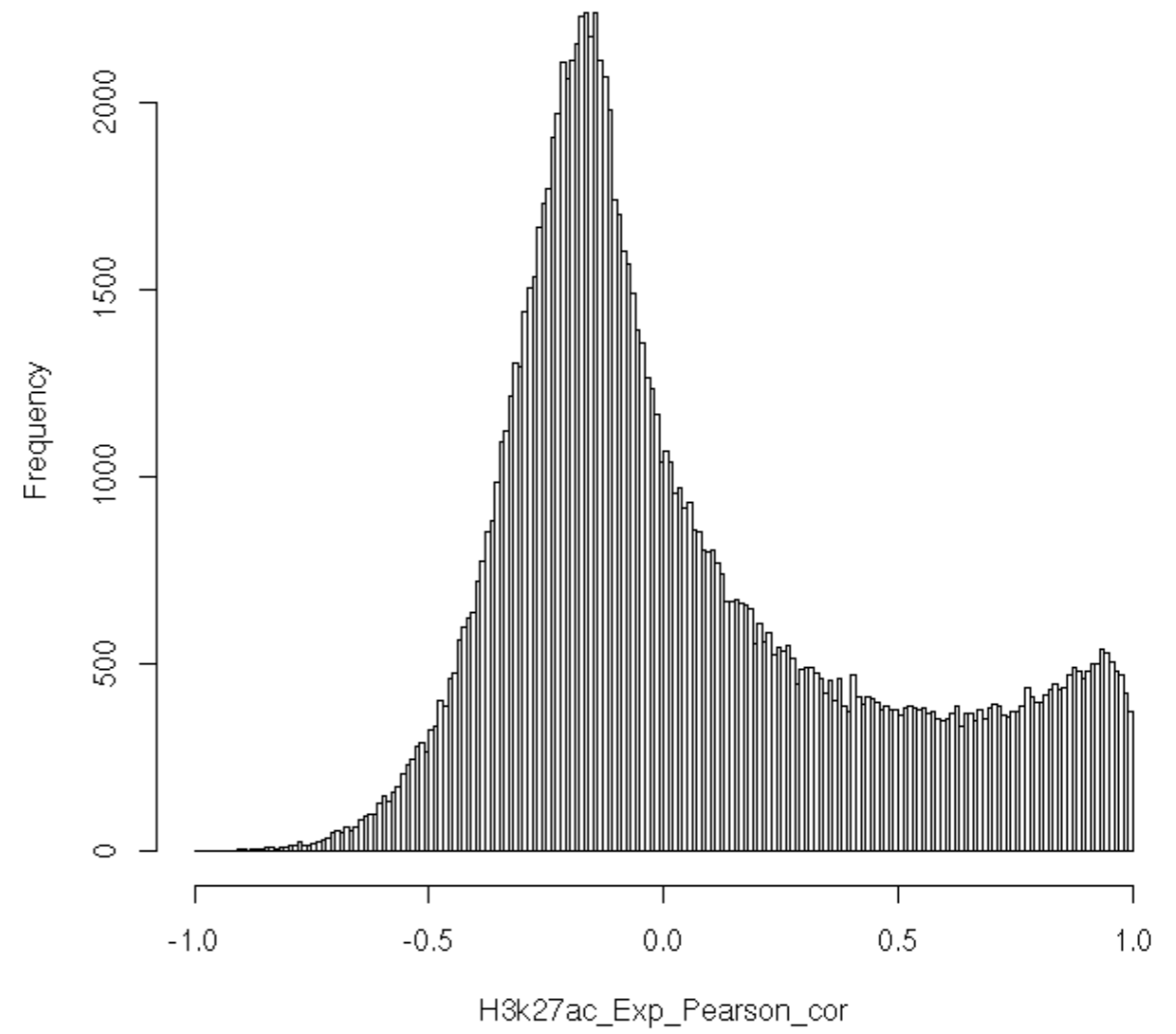


# Signal correlation in closest pairs: enhancer marking and TSS expression

**hs\_masterH3k27ac\_0.00001\_enhancer\_closest\_gencodev7\_rna\_pair**

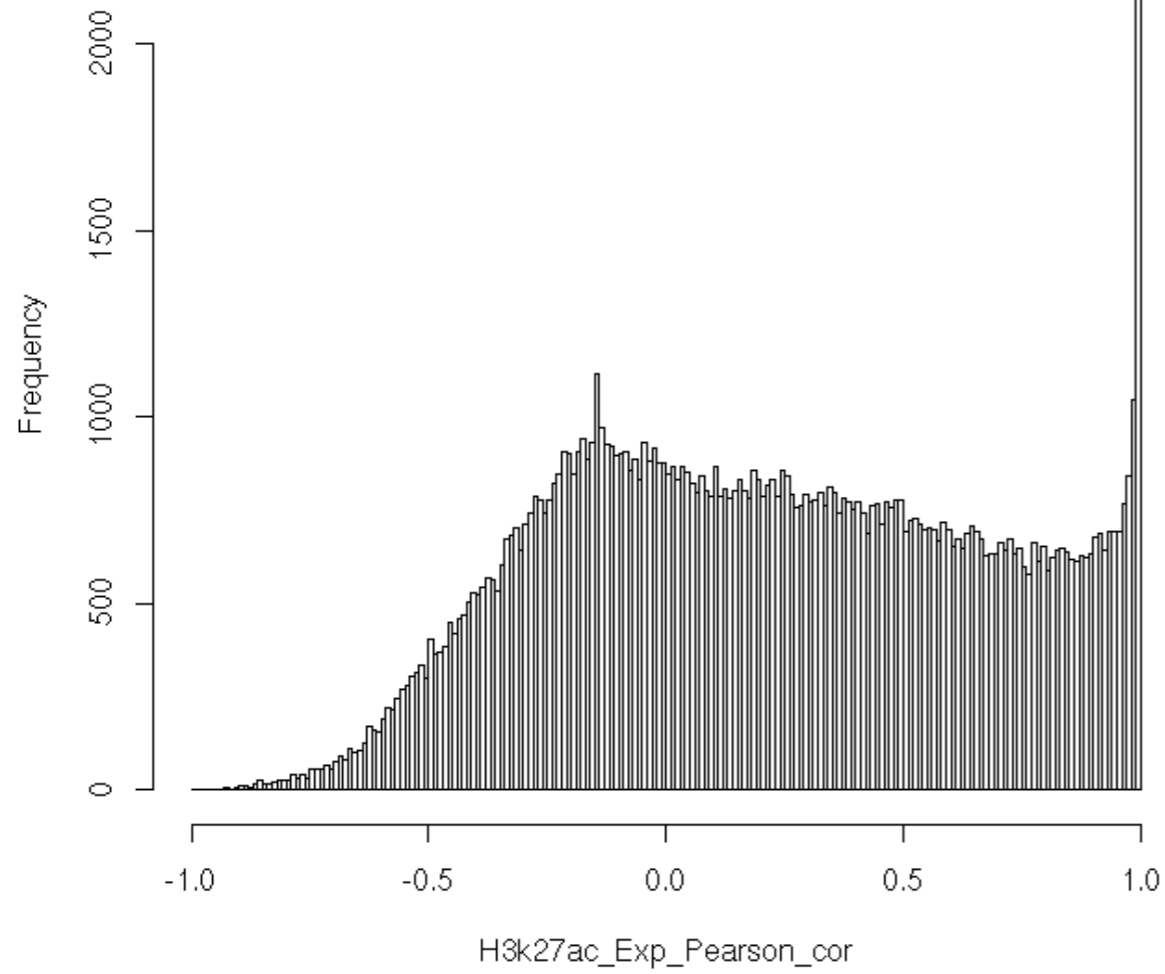


**hs\_masterH3k27ac\_0.00001\_enhancer\_random\_gencodev7\_rna\_pair\_H3k27ac\_Pearson\_cor**

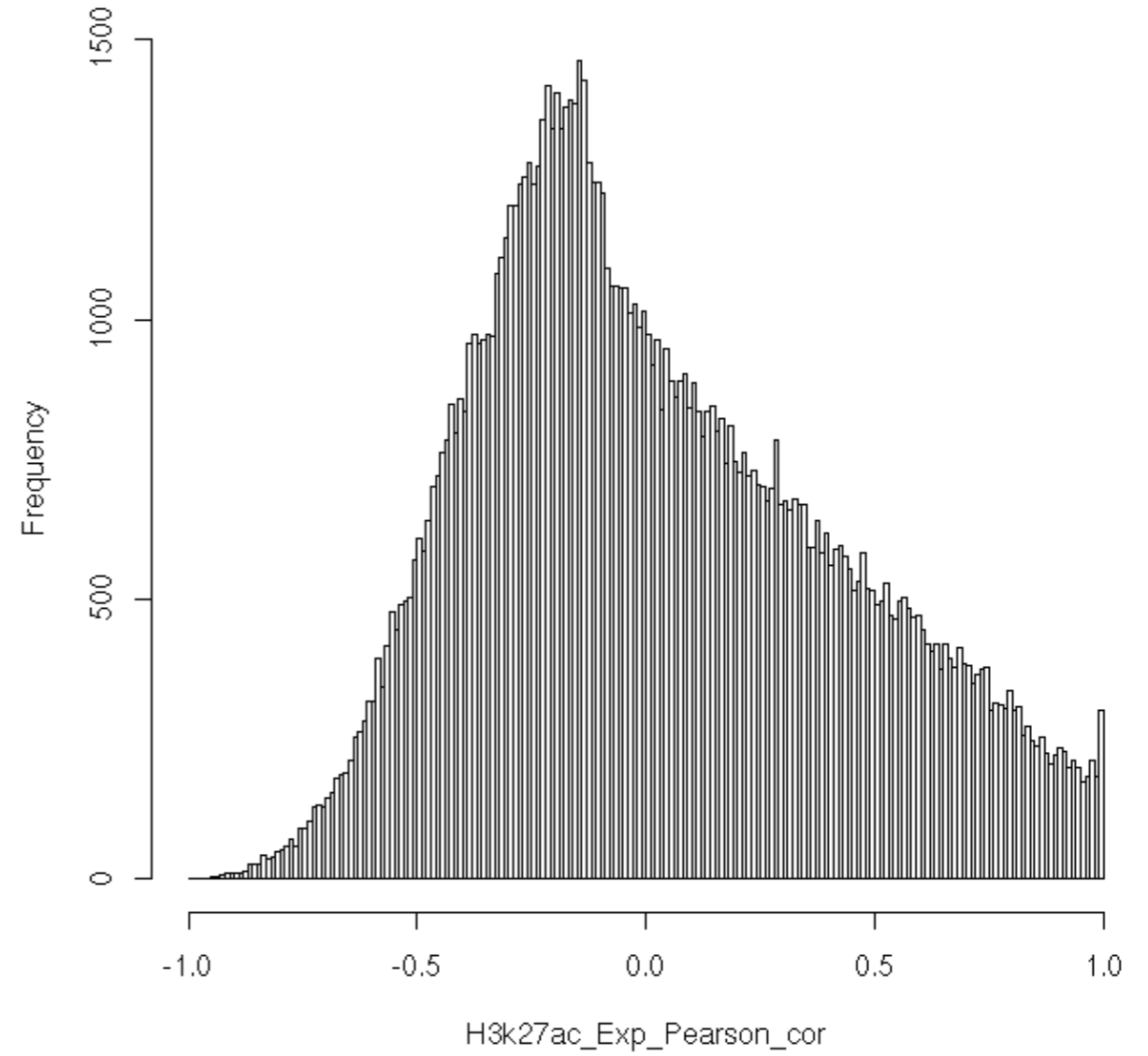


# Signal correlation: promoter marking and TSS expression

**hs\_gencodev7\_promoter\_rna**

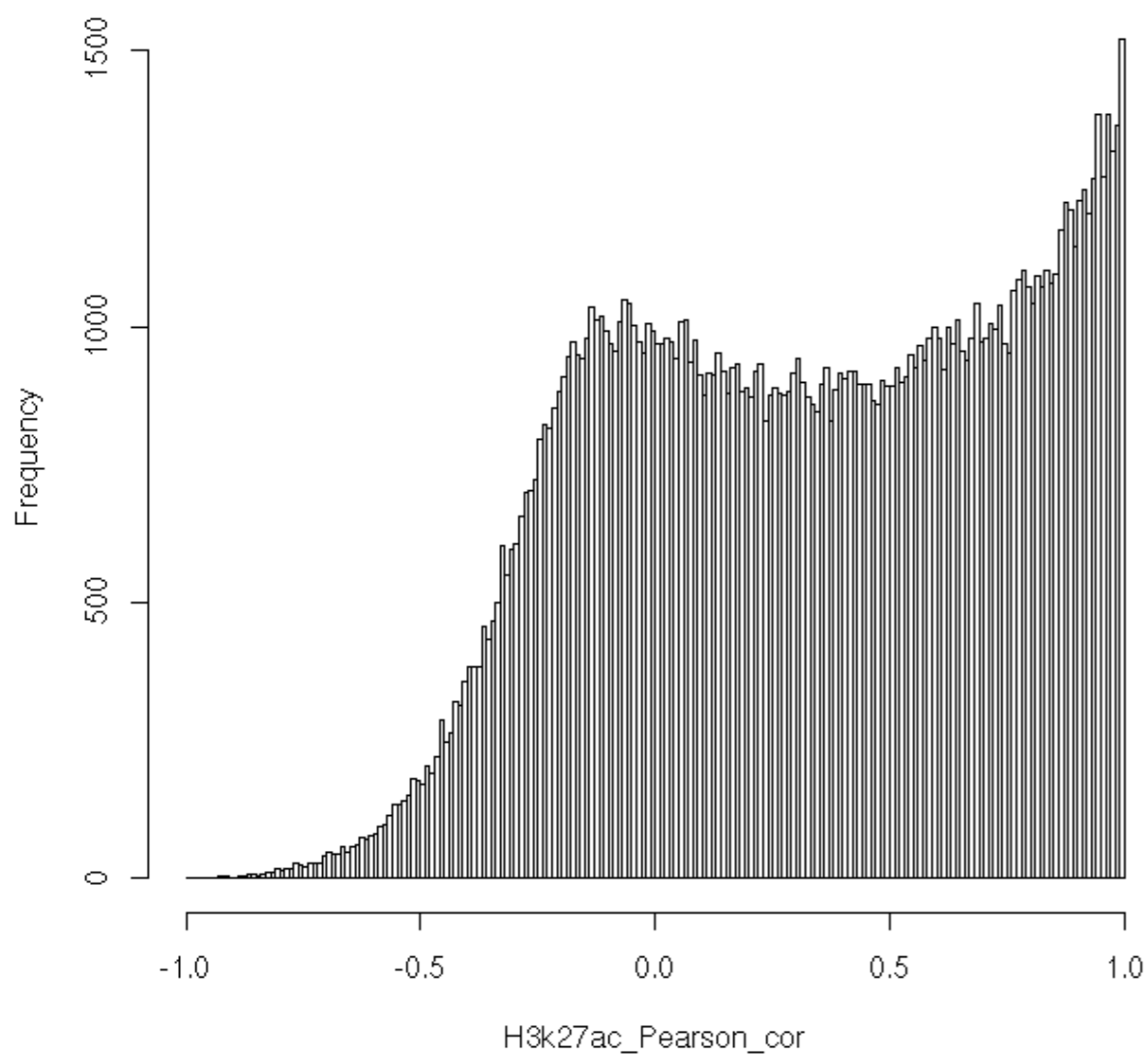


**hs\_gencodev7\_promoter\_random\_rna**

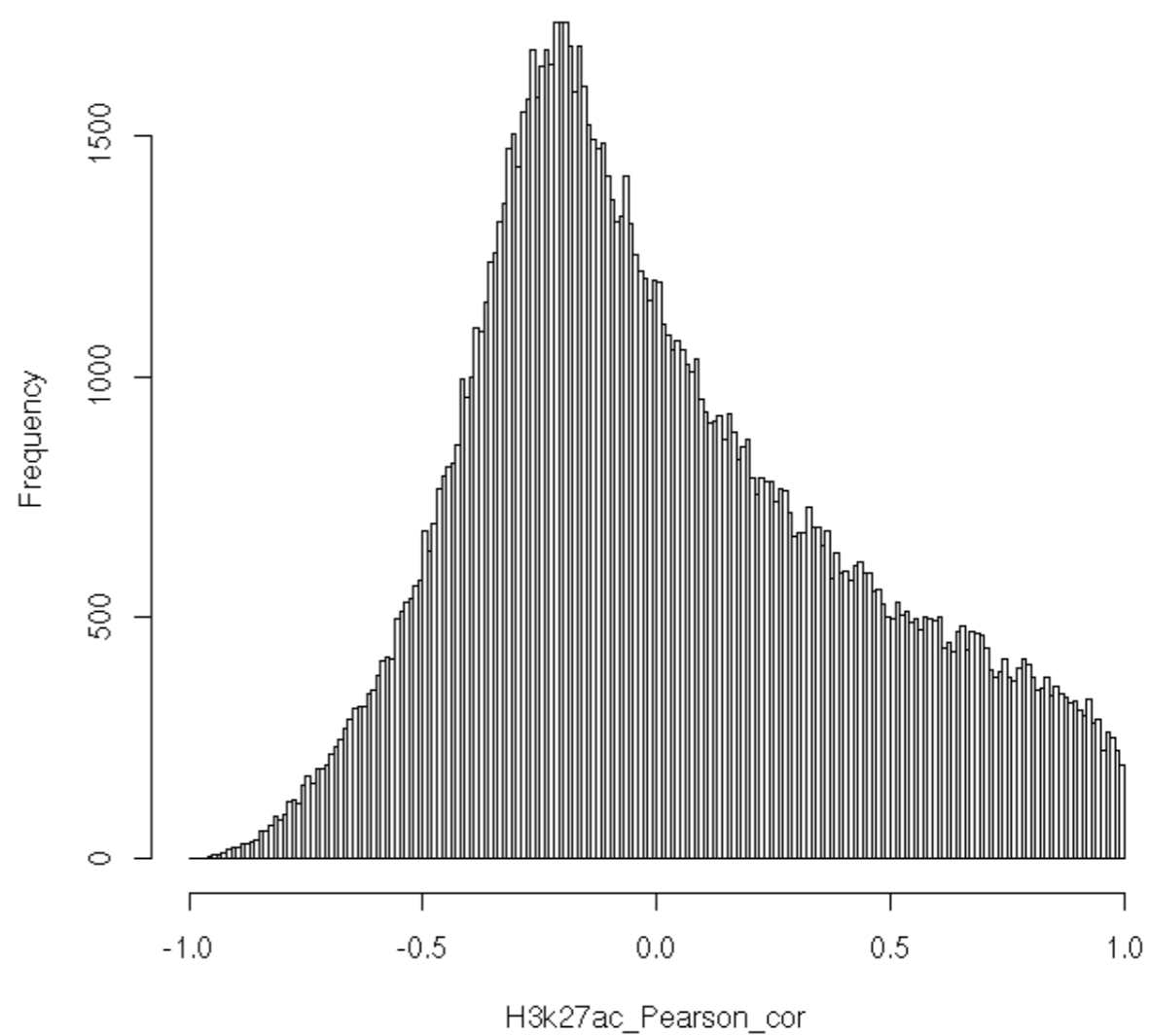


# Signal correlation in closest pairs: enhancer and promoter marking

**hs\_masterH3k27ac\_0.00001\_enhancer\_closest\_gencodev7\_promoter\_pair**

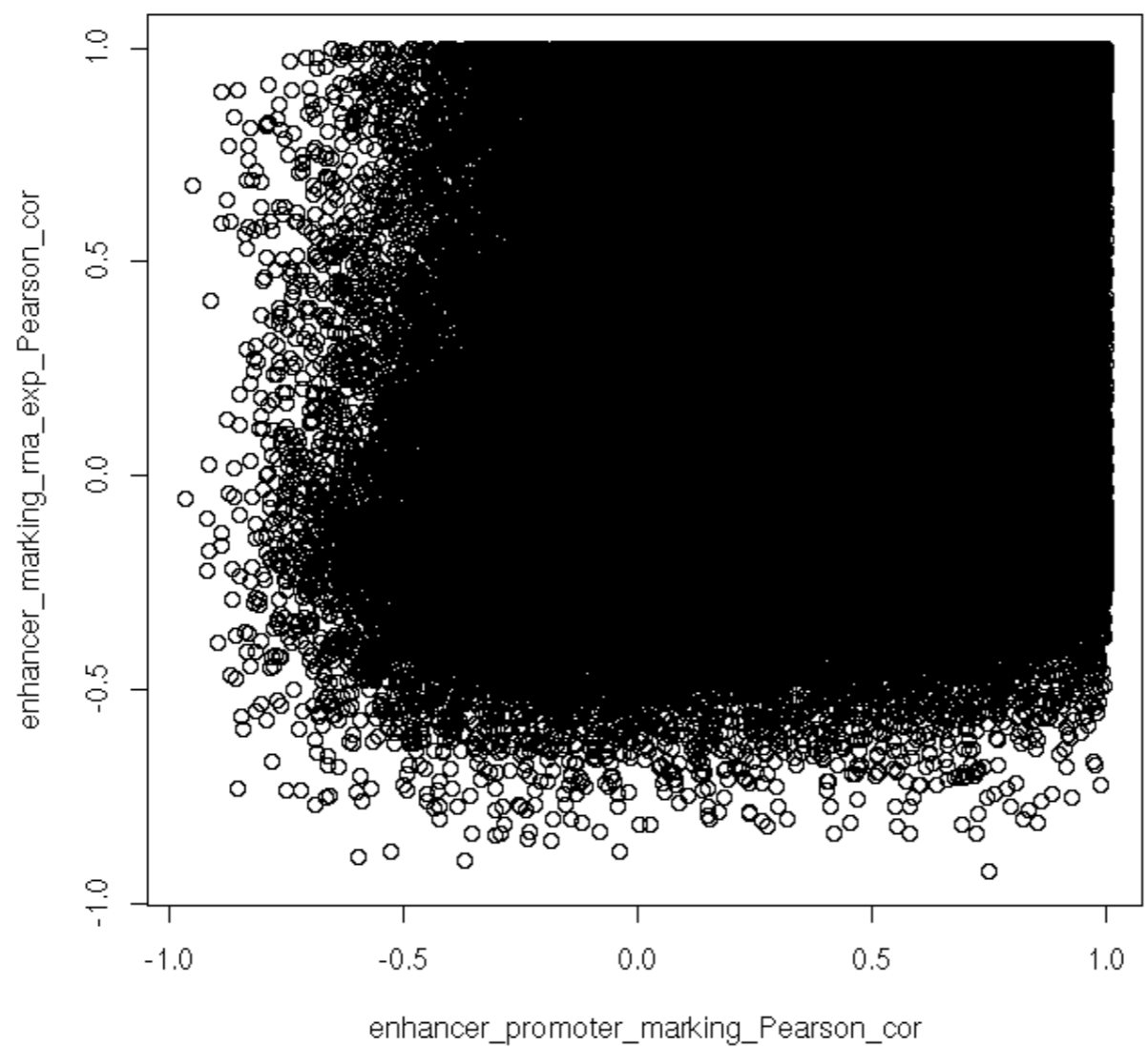


**hs\_masterH3k27ac\_0.00001\_enhancer\_random\_gencodev7\_promoter\_pair\_H3k27ac\_Pearson\_cor**





hs\_masterH3k27ac\_0.00001\_enhancer\_closest\_promoter



Pearson\_cor: 0.3332615

# CRE cross-species dynamics

## I. 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

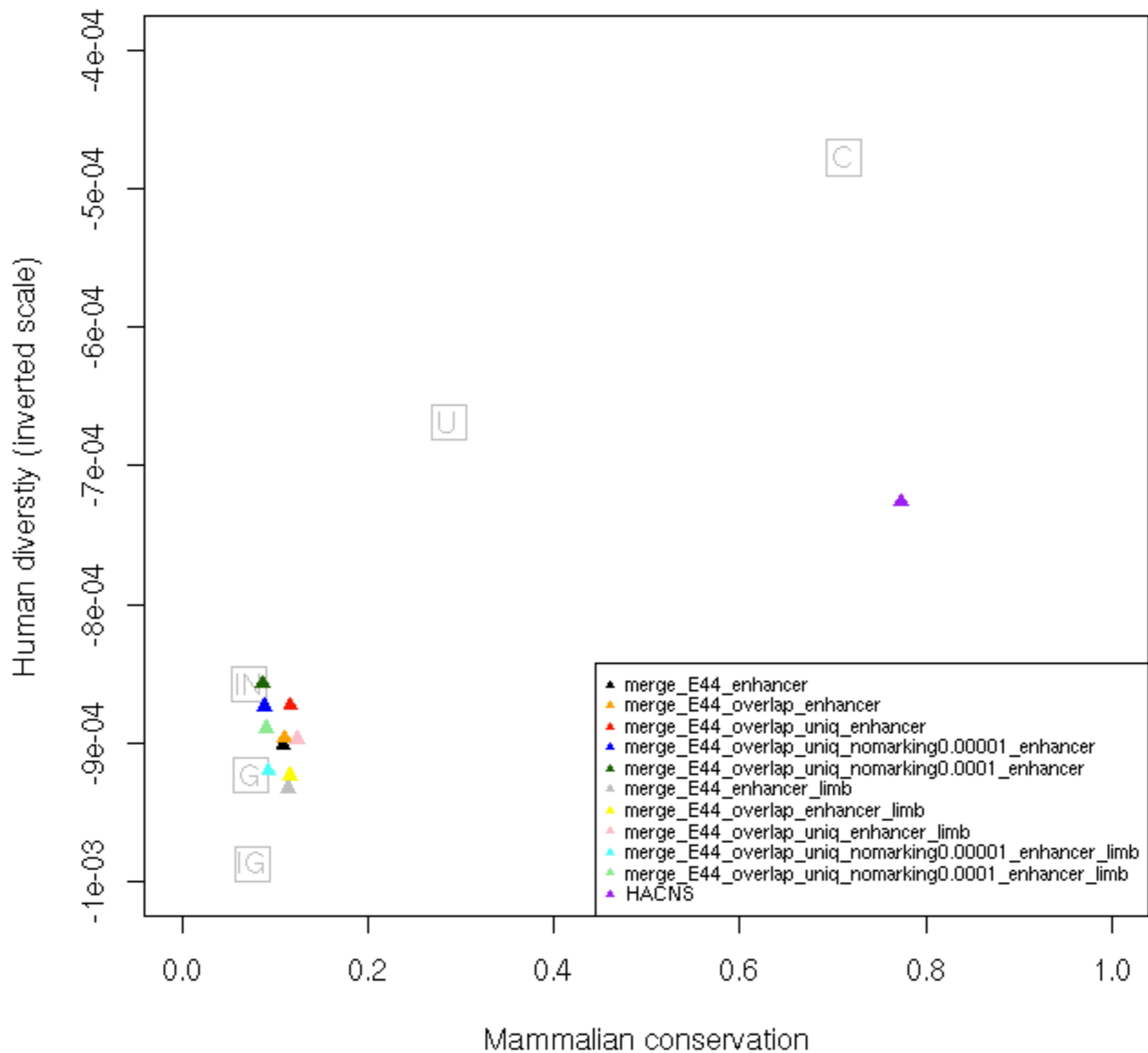
- E44 hs-specific enhancer gain: FIMO\_hs\_gain\_hs\_mm & FIMO\_hs\_gain\_hs\_rm
- mm\_e12.5 hs-specific enhancer loss: FIMO\_hs\_loss\_hs\_mm

A handwritten diagram on a piece of paper. At the top, three labels are written: "per TF", "hs-spec", and "constitutive". Below these labels is a 2x2 grid. The top-left cell of the grid contains the text "# elements with TF motif hit". To the left of the grid, the label "hs" is written vertically. Below the grid, the label "rm / mm" is written.

	hs-spec	constitutive
hs	# elements with TF motif hit	
rm / mm		

### 3. CRE age and constraint

#### Human and mammalian constraint



## E44 hs-specific enhancer & IKG\_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=-0.0029 | 1497, 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=0.0049 | 1789, 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

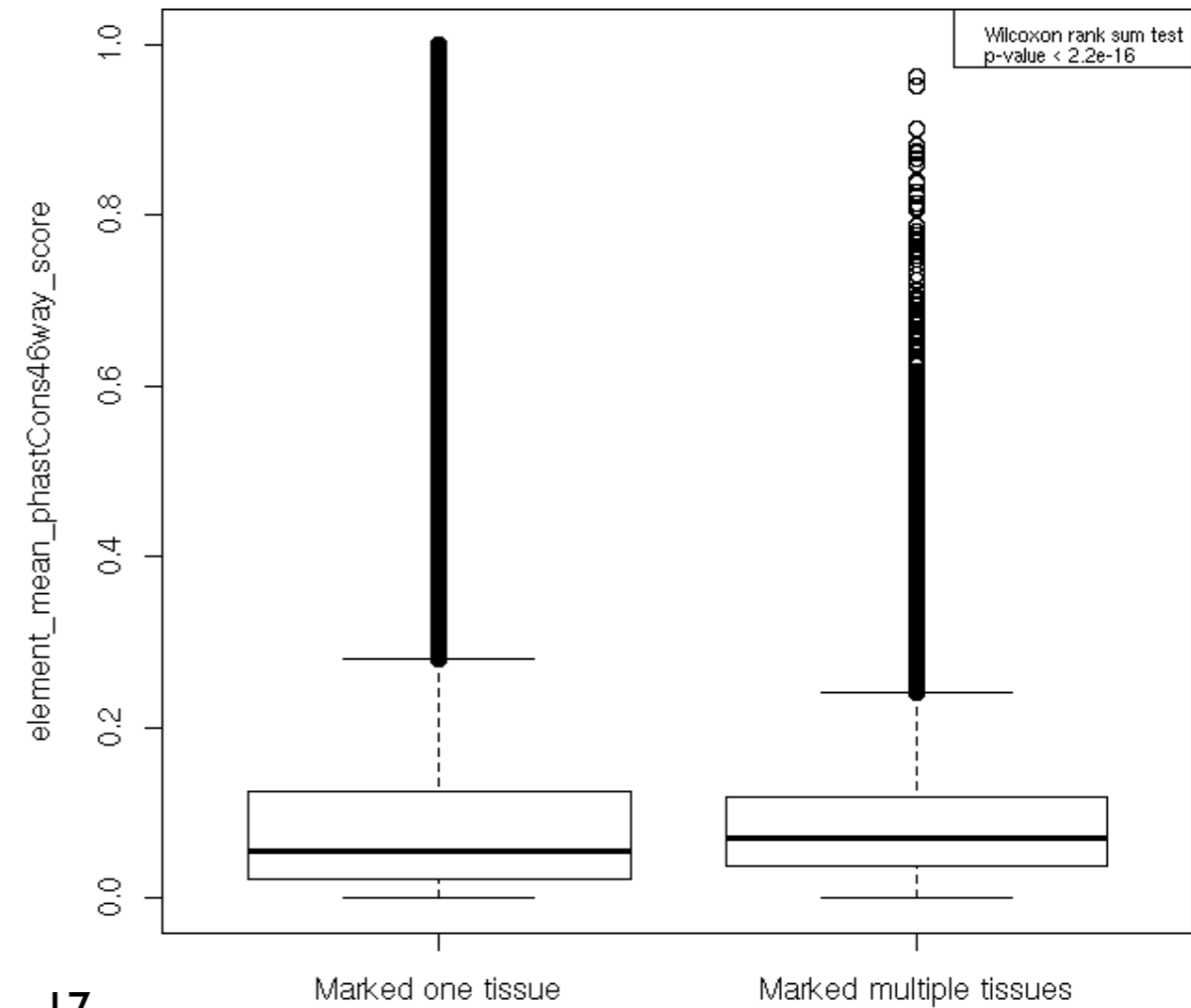
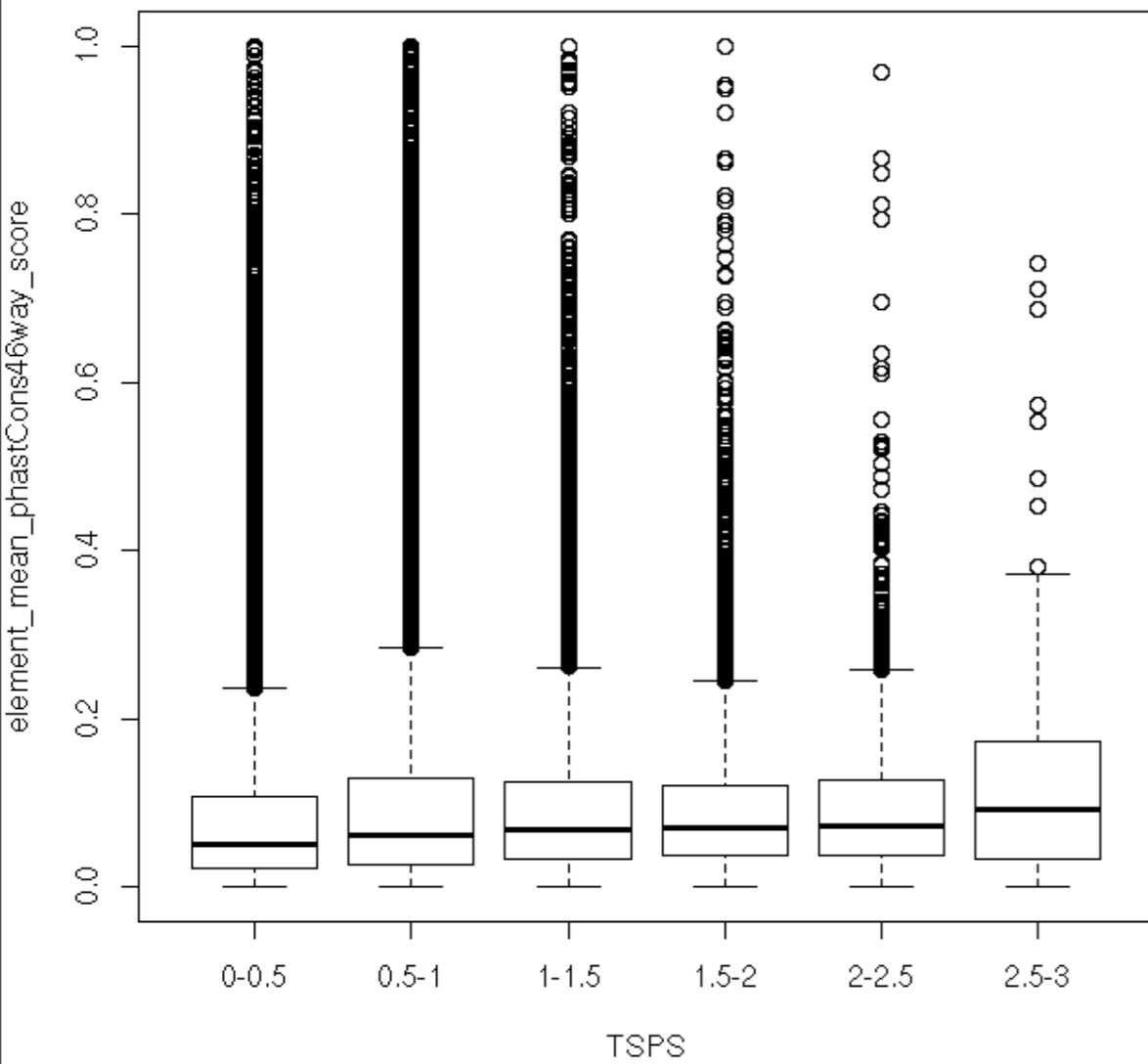
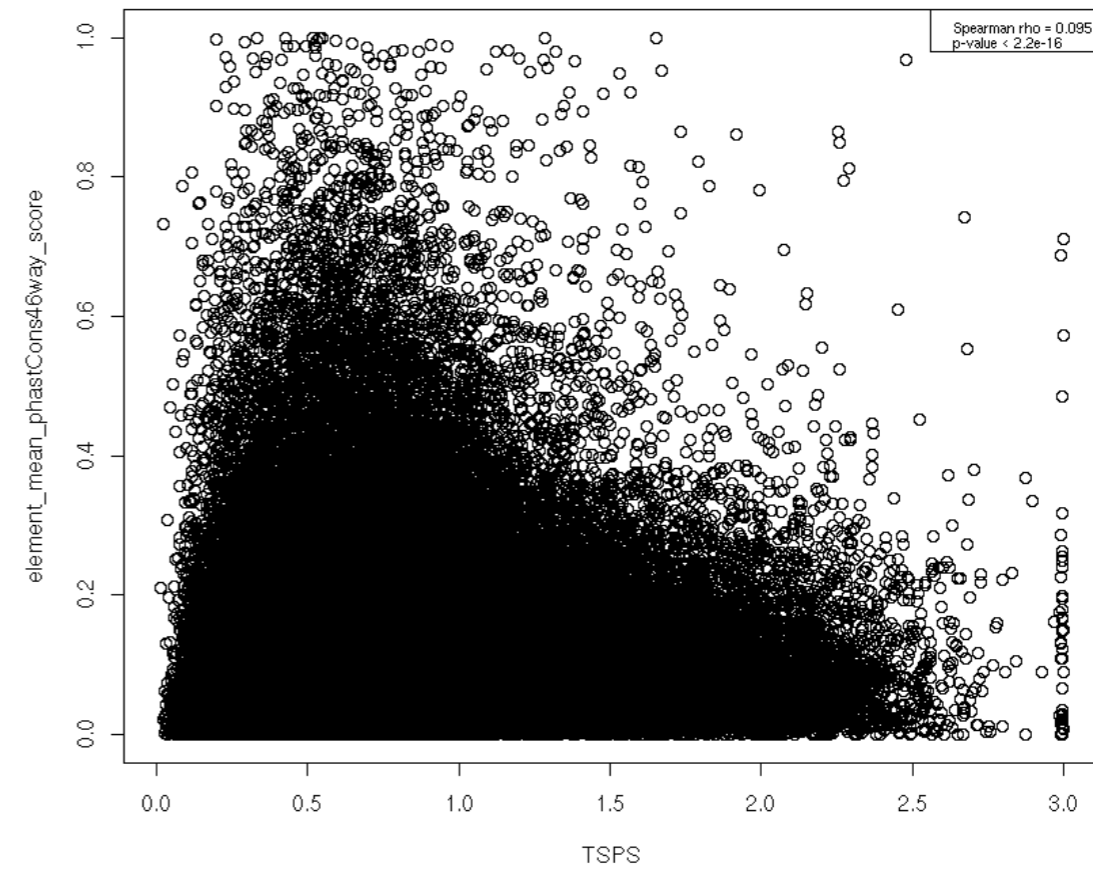
C4\_nonlimb VS C4\_limb: diff=-0.003348254, p=0.635 | p=0.304

C5\_nonlimb VS C5\_limb: diff=-0.002775451, p=0.78 | p=0.364



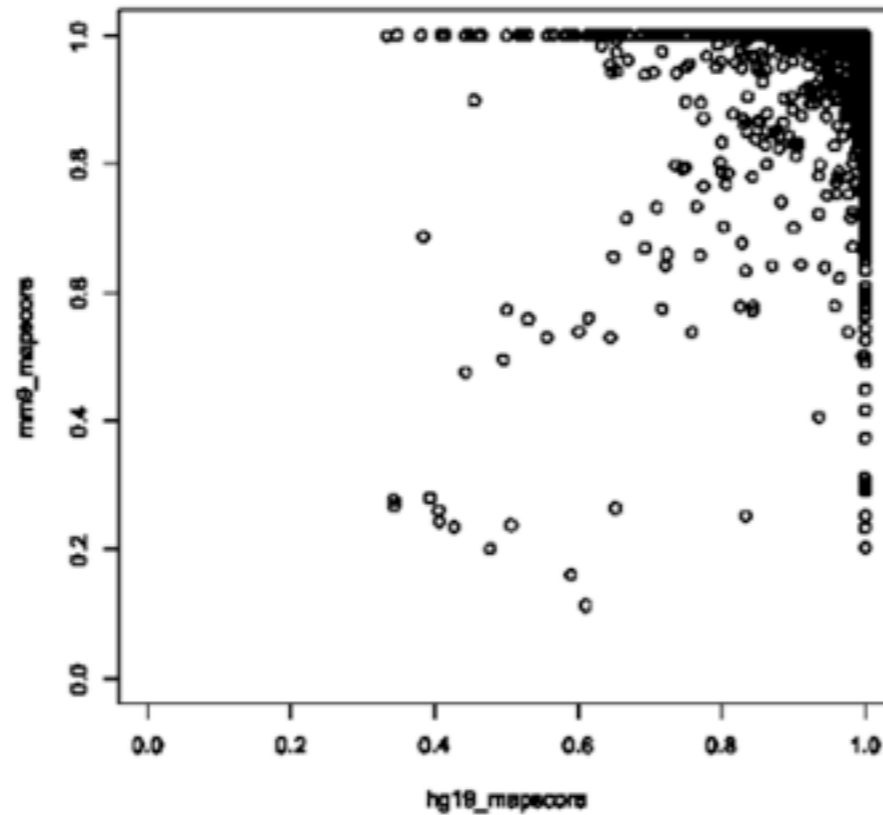
# TSPS & phastCons

hs\_masterH3k27ac\_0.00001\_enhancer



# Cross-species alternative-splicing

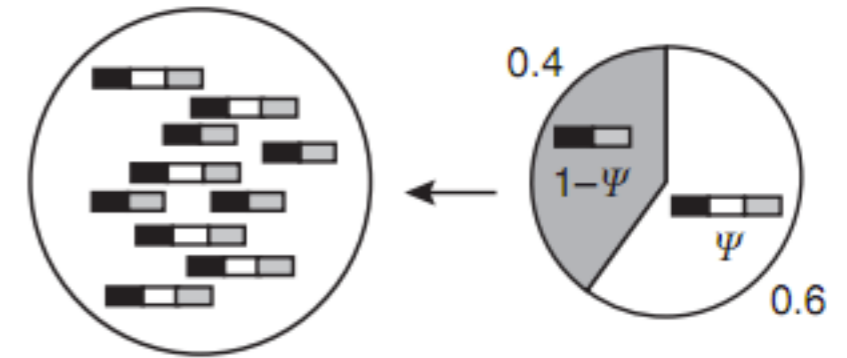
- Method to consider : mappability difference across species ...



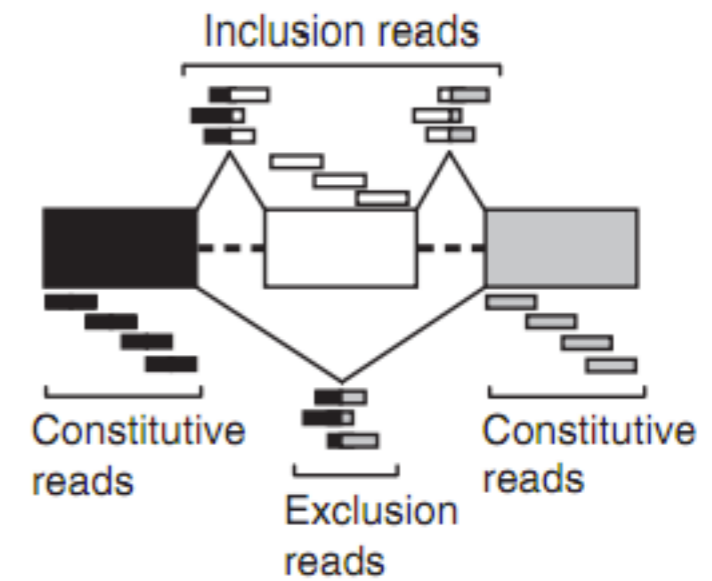
- AS event-based analyses: 1. relative expression level-based 2. read-based

# AS event-based

Alternative transcript events		Total events ( $\times 10^3$ )	Number detected ( $\times 10^3$ )	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
<b>Total</b>		<b>105</b>	<b>100</b>	<b>37,782</b>	<b>22,657</b>	<b>60</b>	<b>68</b>

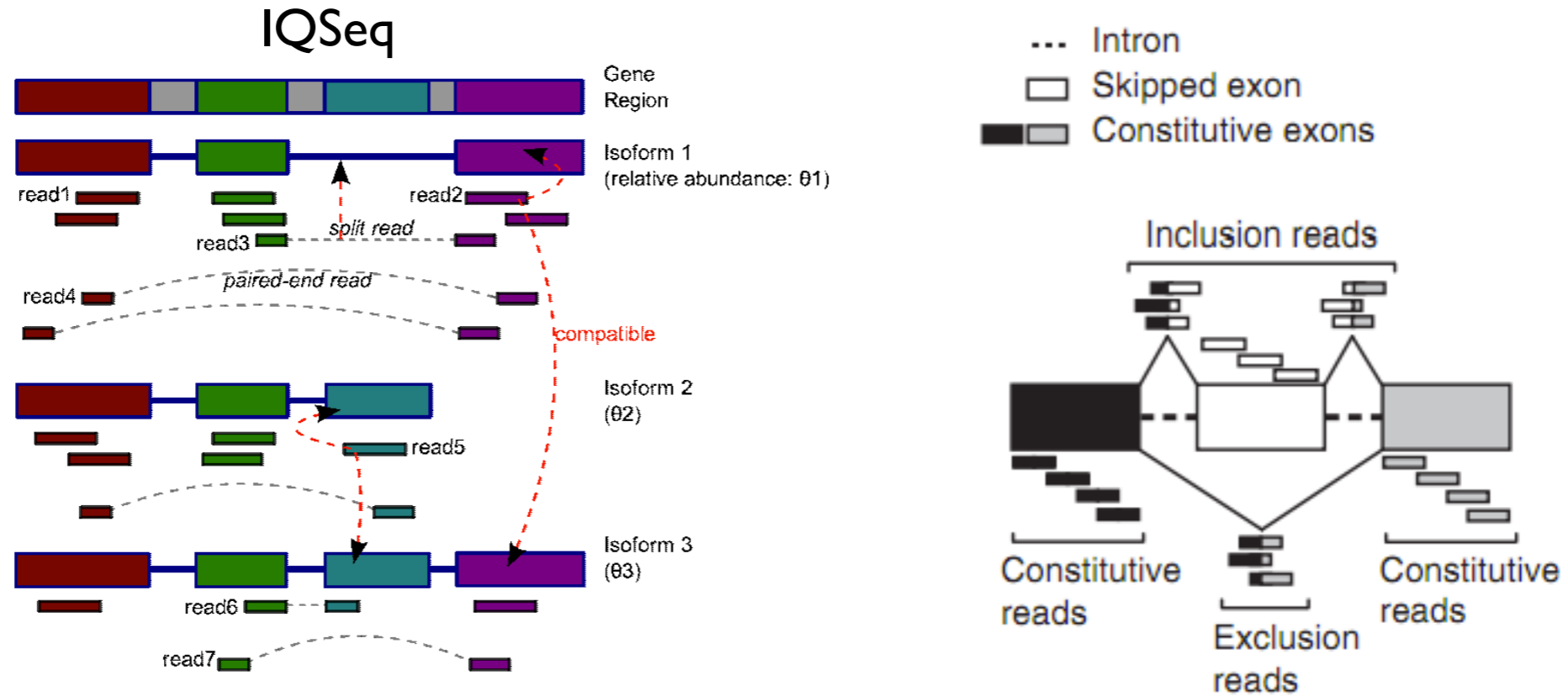


... Intron  
 □ Skipped exon  
 ■ Constitutive exons



C. B. Burge group

# I. relative expression level-based



- 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

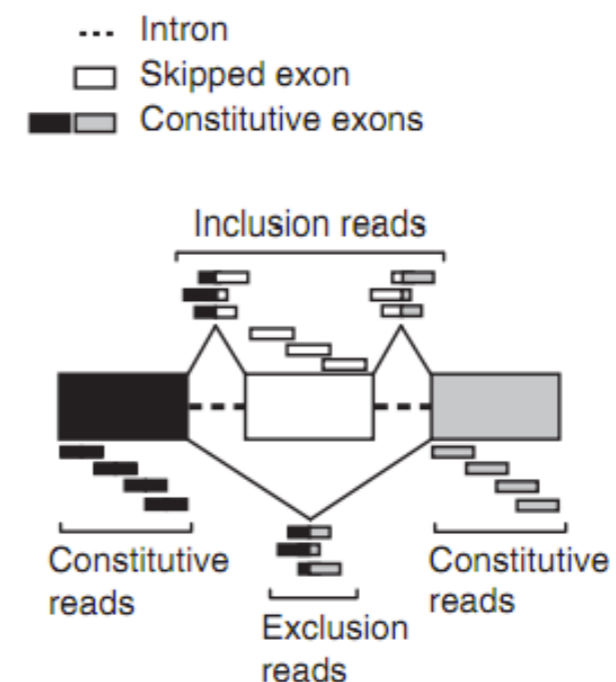
## 2. read-based

- 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<sup>1\*</sup>, Elizabeth Purdom<sup>2†</sup>, Kasper D Hansen<sup>1</sup>, Sandrine Dudoit<sup>1,2</sup>

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



# Cross-species alternative-splicing : dups?

