#### ENGINE: an Enhancer-Gene Interaction dEtection method using robust feature extraction. Part2: Tuning and feature selection

Lou Shaoke

Department of Molecular Biophysics and Biochemistry

loushaoke@gmail.com

November 18, 2015

Yale







408 negative set: MCF7 specific ChIA-PET interactions

Data transformation



![](_page_3_Figure_1.jpeg)

![](_page_4_Figure_0.jpeg)

![](_page_4_Figure_1.jpeg)

![](_page_5_Figure_0.jpeg)

![](_page_5_Figure_1.jpeg)

![](_page_6_Figure_1.jpeg)

![](_page_7_Figure_1.jpeg)

![](_page_8_Figure_1.jpeg)

![](_page_8_Picture_2.jpeg)

![](_page_9_Figure_1.jpeg)

The range of signal is in [min(A)\*min(B), max(A)\*max(B)], then convert to grayscale psuedo image: integer in [0, 255].

![](_page_10_Figure_0.jpeg)

![](_page_11_Figure_0.jpeg)

![](_page_12_Figure_0.jpeg)

![](_page_13_Figure_0.jpeg)

Heterogeneity; saturation affect feature detection; positive set have relative high signal

# Additional negative dataset test


Orignal	negative dataset	Random shift region	Random signal		
AUC	0.92	0.93	0.93		

### Feature selection

![](_page_15_Figure_1.jpeg)

P(pos|neg) conditional density

Feature distribution

#### Feature selection

![](_page_16_Figure_1.jpeg)

 $\label{eq:pvalue} pvalue(= \sum (dhyper(pos\_hit : total\_hit, \#pos\_sample, \#neg\_sample, total\_hit))) < 0.05 \mbox{ and } FC>2, \ \#pos\_features in each marker:$ 

H3k27acH3k4me1H3k4me2H3k4me3H3k9acH3k9me1H3k9me3P300nCpG395835742462400142721106721228More #sig\_features  $\neq$  high importance;

![](_page_17_Picture_1.jpeg)

Example for top H3K27ac features

![](_page_18_Figure_1.jpeg)

		een7	Ŧ
		111 9112 9114 911 911 911 911 911 911 911 911 91	1
	ANN TYPE MEA THE		1
K\$62H3k8me15tdSig			t
K562H3kRec5tdSig			
K562H3Mme3StdSig		2.00	1
K542H3k4me2S3dSig		1.0	
K562H3k4me15tdSig			
K562H3k27ac5id5ig			
k\$42.e123.gpromoter.reg.bed			
genoade.v19.promotected			
H3k0me3.all.bad		2545 J 6005 J	
H3k4me1.all.bed			
H3k9me1.all.bed		9351 9751	
H3k4me2.al.bed		110 9171.3 1554 4007 1736	
Hikkac.all.bed		1233	
H3k27ac.al.bed			
H3Mme3.all.bed			
P300 all bed			
k542.o123.enh.reg.bed			
E123_15_enh.bed		6,600 7,600 7,600 7,600 7,600 7,600 7,600 7,600 7,600	1
wgEncodeOpenChromDnaseK56 bioWg			1
nCpG.al.bed			1
H3Mme2.A.bed	T		Ĩ
H3k4me2.A0.bed		dent Abat	1
H3Mme2.B.bed		1910 9P13 9824	1

		etr/22	p12	p11.2 p11.1	9114	911.21	a11.22 a11.23	9124	9122	et2.3	a13.1 a	13.2 013.31	et 332 et 33	-
														-
	A Tree			22,220 ka		22,238 kb	60 kb	22,340 kb		22,259 kb		22,360 kb		•
	100	The second second	101. A& & &		an 1 - 1	6. 5.	-			1			-	-
K\$62H3k9me1StdSig		No. of the local sector					Addisonal		يطلبه بالت	and a state of the second	add a sheep	All March and Area	he and the sta	4
K542H3k8ac5td5ig		Landa alla an	<u> </u>	JV V		Ach	A have been a first of		mar hell.	والمعد والعالم مراجع الم	un ale	all and a		
K542H3k4me3StdSig		11.201												
K542H3k4me2SteSig		1-10	and and the	ALL ALL I	MA									
K562H3k4me15tdSig		190			1111	4.1	Andrew		HALL NO.		and a shift of	dia se a sa		
K562H0k27ac5h55ig	Т	1998 Lable and Andrews	al Alda	And	11 A.	w.h.h.						And the second	hat	
k\$42.o123.gpromater.reg.bed				_										
genoode.v18.promotat.bad				MAPK1 MAPK1	MAPIKI									
H3k0me3.all.bed				4530.7 12004 2200.2	5720.2							2929.5	1523.3 3160.3	,
H3k4me1.all.bed														
H3k9me1.al.bed					•									
Withdow? all had				6455.4 2143.7 2038	1.2 2005.4							1713 3687	17776	
				1181.1 21863								1723.3	20064	-
Hikhacalized				9423.1								10655	16336	
H3k27as.al.bed														
H3k4me3.al.bed				1120 14993										
P300.al.bed														
k\$42.o123.enh.mg.bed	П			2190.9								10167 1942.2 19		
						-								
E123_15_enh.bed		Er/AG 6	_EebG	6_Eviv0	7_Enh	7_Enh	12_ErABV					12_ErABV	12_EnhBiv	
wgEncodeOpenOncenDeaseK56	11	H-000		-	12_EnhB	~	7_64						7_bn	
bigWg				Anna	-				-					4
nCpG.al.bed				1673.3 12805 332	9 1897.8 1264.4							8240.0 3326	7854.5	
H3k4me1.80.bed				19242 7215.4									12691	
Hikkme2.A.bed				Mant										
H3Mme2.8 ked				Blast								Blast		-
H3Mme3.A.bed														_

![](_page_21_Figure_0.jpeg)