

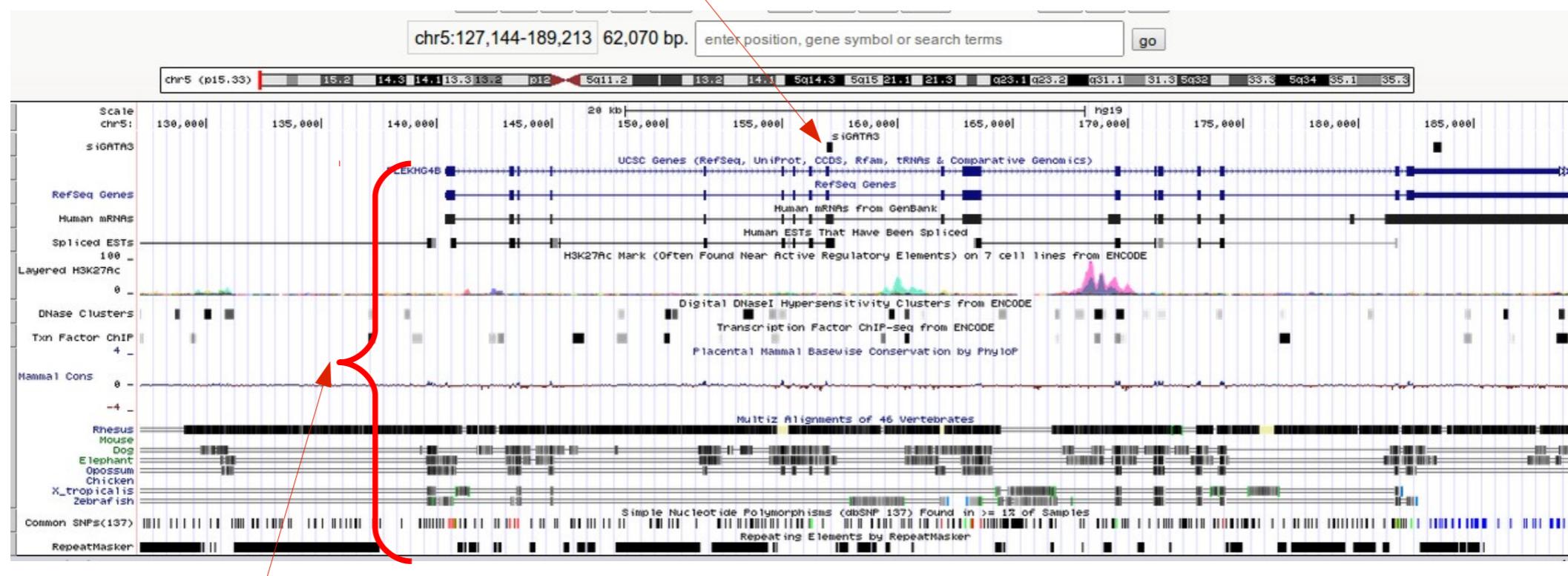
ChIP-seq

Annotation and Visualization

How to add biological meaning to peaks

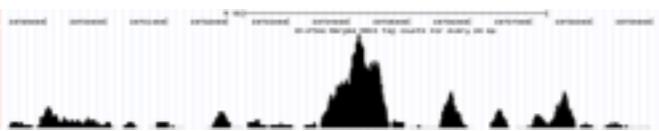
M. Defrance, C. Herrmann, D. Puthier, M. Thomas-Chollier,
S Le Gras, J van Helden

Custom track uploaded by the user (here ESR1 peaks in siGATA3 context)



public UCSC annotation/data tracks

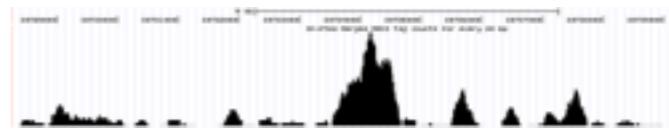
ChIP-seq peaks



Typical questions

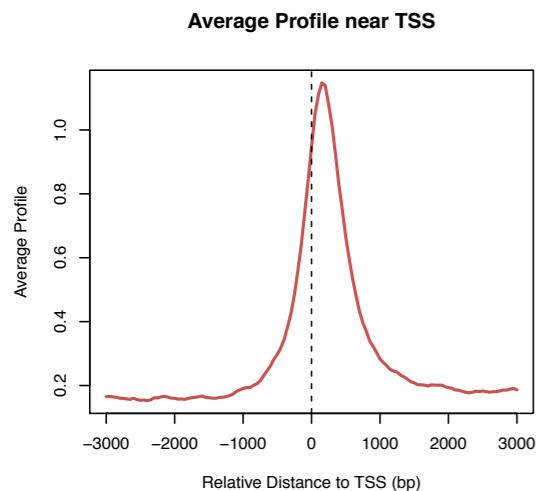
- What are the genes associated to the peaks?
- Are some genomic categories over-represented?
- Are some functional categories over-represented?
- Are the peaks close to the TSS, ...?

ChIP-seq peaks



Annotation Visualisation

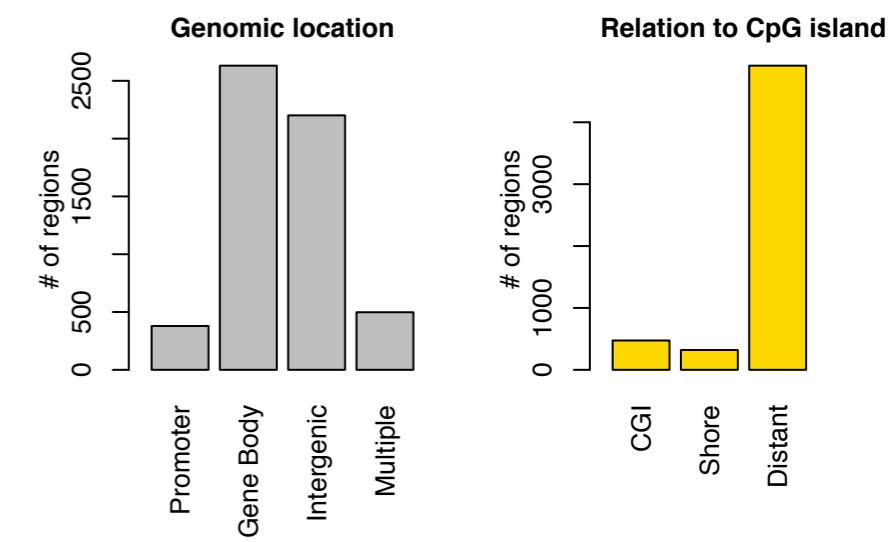
Enrichment profiles



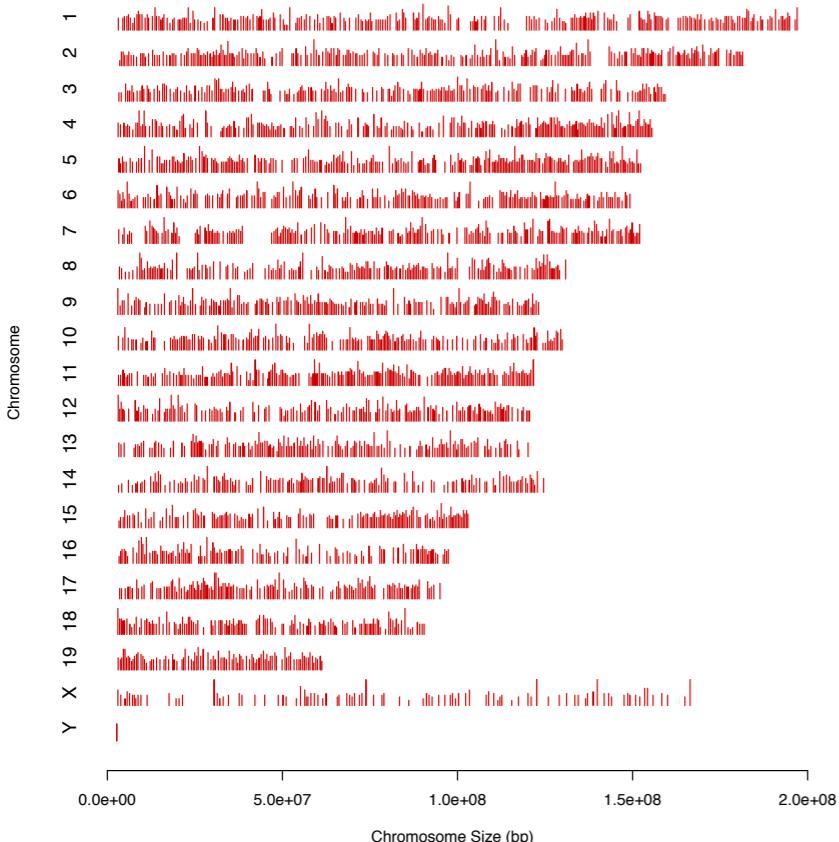
Annotated peaks

chr	start	end	Gene
chr15	65294195	65295186	
chrX	19635923	19638359	Chst7
chr8	33993863	33995559	
chr10	114236977	114239326	Trhde
chrX	69515082	69516482	Gabre
chr4	49857142	49858913	Grin3a
chr16	7352861	7353410	Rbfox1
chr7	64764156	64765421	Gabra5
chrX	83436881	83438330	Nr0b1
chr10	120288598	120289143	Msrb3
chr5	67446361	67446855	Limch1

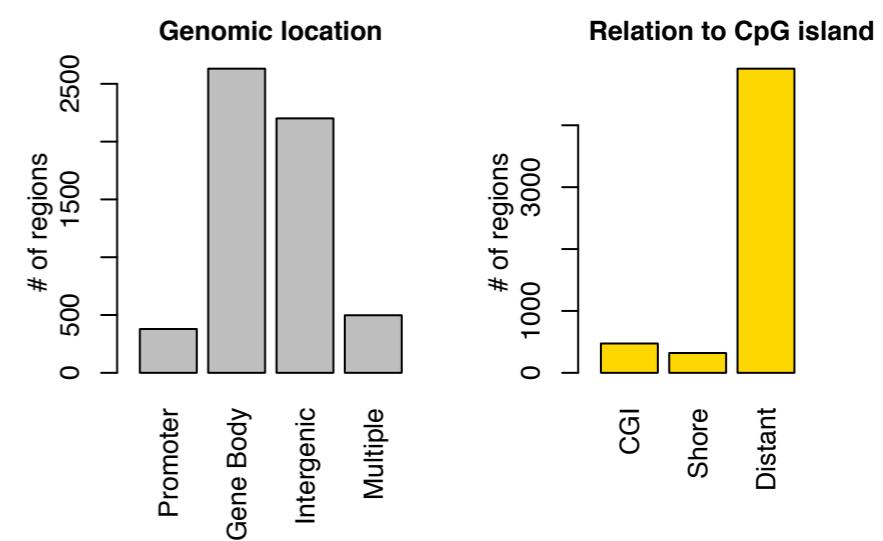
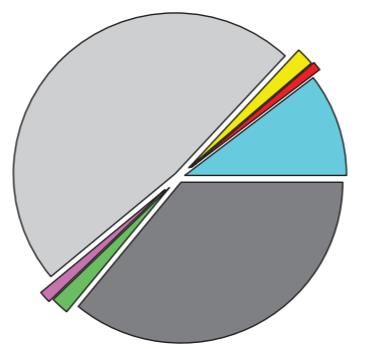
Genomic & functional Annotation



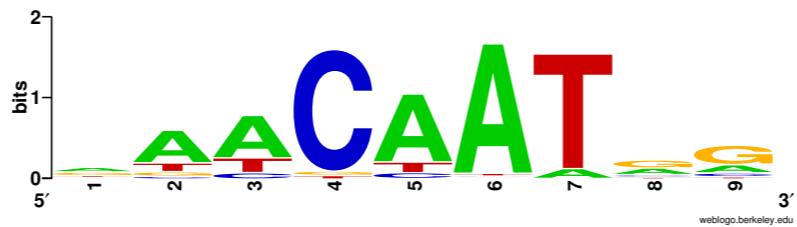
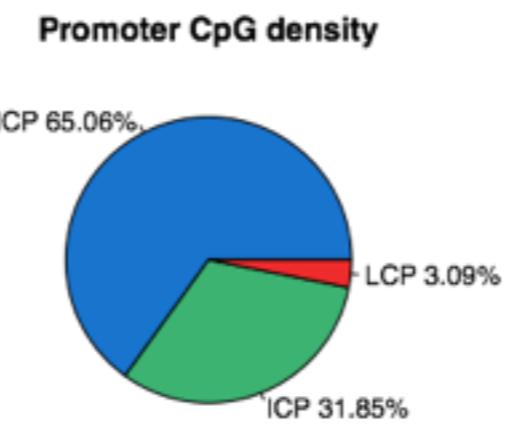
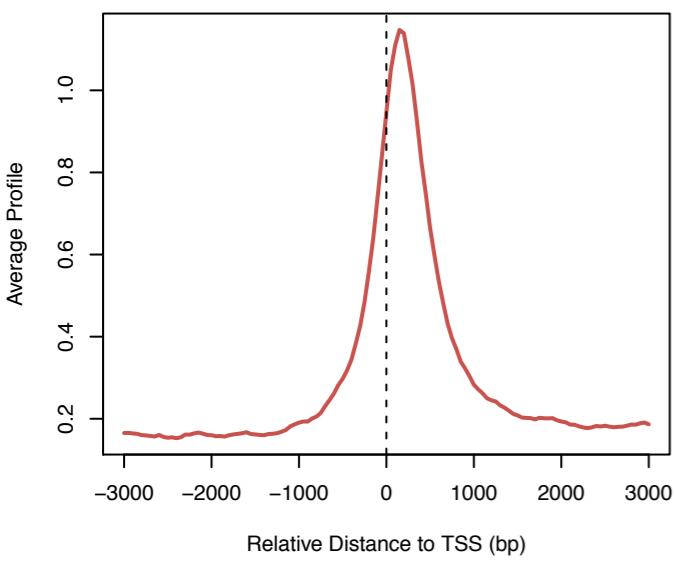
ChIP Regions (Peaks) over Chromosomes



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R		
1	PeakID	Chr	Start	End	Strand	Peak	Sco	Focus	R ₀	Annotation	Detailed Anno	Distance to T	Nearest Pror	PromoterID	Nearest Uniq	Nearest Refs	Nearest Ense	Gene Name	Gene Alias	Gene Descrip
2	chr18-1	chr18	69007968	69008268	+	593	0.939	intron (NR_03> intron (NR_03>	74595	NR_034133	400655	Hs.579378	NR_034133	LOC400655	-					hypothetical
3	chr9-1	chr9	88209966	88210266	+	531.9	0.946	Intergenic	50894	NM_0011851	79670	Hs.597057	NM_0011851	ENSG000000000000	ZCCHC6	DKFZp66B1				zinc finger, C
4	chr14-1	chr14	62337073	62337373	+	505.4	0.918	intron (NM_15> intron (NM_15>	244485	NM_172375	27133	Hs.27043	NM_139318	ENSG000001 KCNHS	EAG2 H-EAG	potassium va				
5	chr17-1	chr17	5076243	5076543	+	492.1	0.936	intron (NR_03> intron (NR_03>	2414	NM_207103	388325	Hs.462080	NM_207103	ENSG000001 C17orf187	FLJ32580 M-chromosome					
6	chr17-2	chr17	47851714	47852014	+	476.2	0.824	Intergenic	259488	NM_001082	56934	Hs.463466	NM_001082	ENSG000001 CA10	CA-RPX CAR	carbonic anh				
7	chr10-1	chr10	98420680	98420980	+	474.9	0.967	intron (NM_15> intron (NM_15>	49439	NM_152309	118788	Hs.310456	NM_152309	ENSG000001 PIK3AP1	BCAP RP11-	phospholinos				
8	chr9-2	chr9	81294389	81294689	+	456.3	0.957	Intergenic	82159	NM_007005	7091	Hs.444213	NM_007005	ENSG000001 TLE4	BCE-1 BCE1					transducin-β
9	chr14-2	chr14	36817736	36818036	+	452.3	0.757	intron (NM_15> intron (NM_15>	81017	NM_001195	145282	Hs.660396	NM_001195	ENSG000001 MIPO1	DKFZp313M mirror-image					
10	chr18-2	chr18	20049825	20050125	+	449.7	0.853	intron (NM_06> intron (NM_06>	56219	NM_018030	114876	Hs.370725	NM_018030	ENSG000001 OSBP1A	FLJ10217 Oxysterol bin					
11	chr7-1	chr7	12226829	12227129	+	445.7	0.901	intron (NM_01> intron (NM_01>	9606	NM_001134	54664	Hs.396358	NM_001134	ENSG000001 TMEM106B	FLJ11273 M-transmembr					
12	chr14-3	chr14	88712188	88712488	+	443.1	0.844	intron (NM_0C> intron (NM_0C>	240869	NM_005197	1112	Hs.621371	NM_001085	ENSG000000 FOXN3	C14orf116 C	forkhead box				
13	chr18-3	chr18	62951924	62952224	+	443.1	0.947	Intergenic	382689	NR_033921	643542	Hs.652901	NR_033921	LOC643542	-				hypothetical	
14	chr3-1	chr3	32196769	32197069	+	443.1	0.87	Intergenic	58256	NM_178868	152189	Hs.154986	NM_178868	ENSG000001 CMTM8	CKLFSF8 CKL	CKLF-like MA				
15	chr11-1	chr11	110685448	110685748	+	425.8	0.907	Intergenic	9849	NR_034154	399948	Hs.729225	NR_034154	C11orf92	DKFZp781P1	chromosome				
16	chr4-1	chr4	81755366	81755666	+	423.2	0.908	intron (NM_15> intron (NM_15>	279618	NM_152770	255119	Hs.527104	NM_152770	ENSG000001 C4orf22	MGC35043	chromosome				



Average Profile near TSS



CAAT

Mouse Phenotype

Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment
abnormal limbs/digits/tail morphology	2	2.0559e-91	6.6837e-88	2.1465
abnormal craniofacial morphology	3	9.3822e-91	2.0334e-87	2.0082
abnormal limb morphology	5	2.4990e-80	3.2497e-77	2.3077
abnormal appendicular skeleton morphology	10	3.0255e-70	1.9672e-67	2.3450
abnormal skeleton extremities morphology	12	3.2687e-69	1.7711e-66	2.3724
abnormal paw/hand/foot morphology	13	4.0300e-69	2.0156e-66	2.6813
abnormal head morphology	14	6.4657e-67	3.0029e-64	2.0134
abnormal digit morphology	18	1.0543e-61	3.8084e-59	2.6982
abnormal cartilage morphology	23	7.3728e-58	2.0843e-55	2.3432
abnormal skeleton development	24	3.5769e-56	9.6904e-54	2.0833
abnormal long bone morphology	25	4.6593e-56	1.2118e-53	2.3374

ChIP-seq peaks

(bed, xls, txt file)

MACS peaks in bed format

chr1 3001827	3002328	MACS_peak_1	55.28
chr1 3067471	3067948	MACS_peak_2	50.67
chr1 3660316	3662844	MACS_peak_3	352.43
chr1 3842462	3842994	MACS_peak_4	59.21
chr1 3877254	3877710	MACS_peak_5	52.72
chr1 3939314	3939679	MACS_peak_6	82.99

Statistical significance
-10 log(P-value)

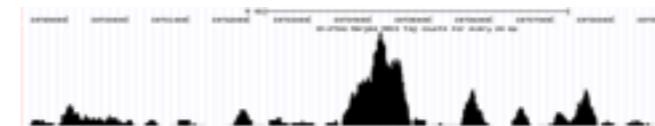


MACS peaks extented format

Chr	Start	End	W	Summit	Tags	Sig	Fold	FDR
chr16	35981451	35981951	321	35981701	24	1107.07	30.55	0.0
chr18	30784846	30785346	628	30785096	40	964.91	43.62	0.0
chr14	79381873	79382373	441	79382123	29	939.17	37.2	0.0
chr12	34467249	34467749	1160	34467499	53	928.38	19.93	0.0
chr8	90304944	90305444	1804	90305194	80	883.76	10.21	0.0
chr15	65294343	65294843	992	65294593	62	824.32	13.4	0.0
chr17	48499365	48499865	370	48499615	24	798.58	20.62	0.0
chr18	72429446	72429946	531	72429696	31	790.48	39.77	10.0
chr15	54579253	54579753	487	54579503	29	781.63	32.15	9.09
chr13	56988583	56989083	916	56988833	60	777.7	9.44	8.33

ChIP-seq profiles

(wig, wig.gz, bigWig)



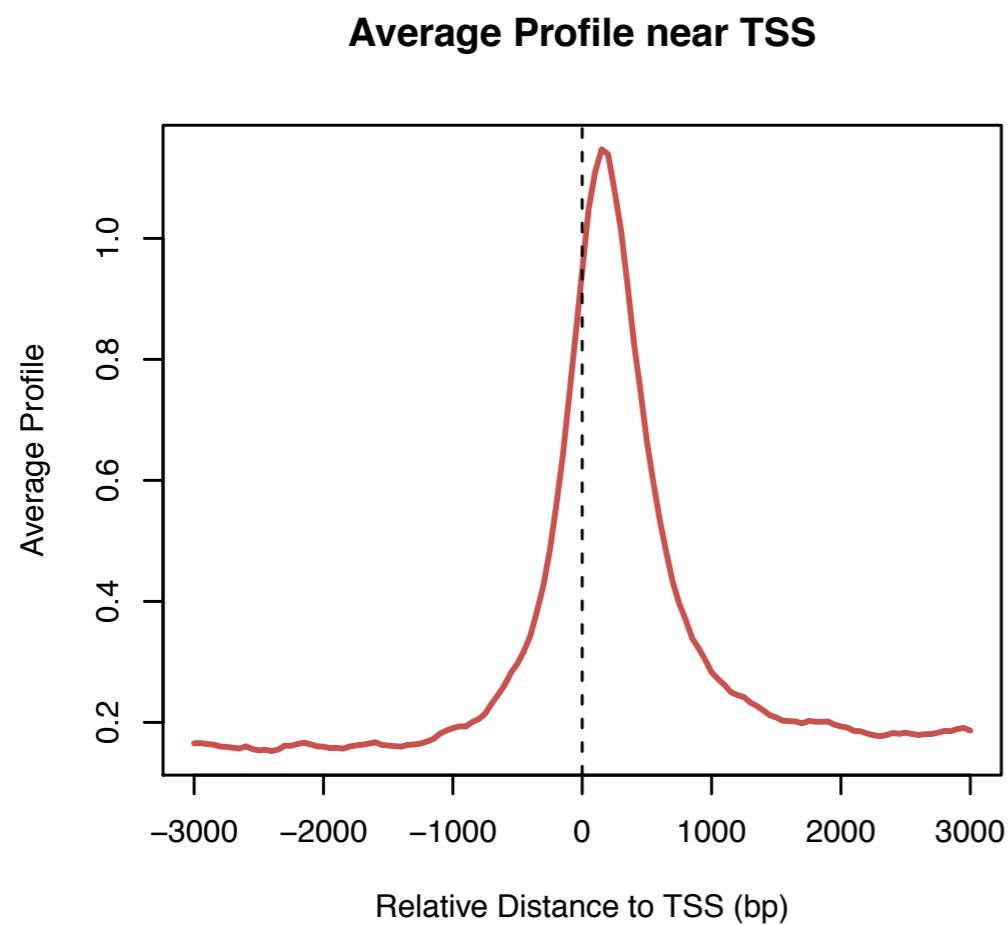
wig generated by MACS

```
track type=wiggle_0 name="ChIP-H3K4-1_treat_all" description="Extended tag pileup from MACS version  
1.4.1 for every 40 bp"  
variableStep chrom=chr1 span=40  
3000361 2  
3000401 2  
3000441 2  
3000481 4  
3000521 4  
3000561 2  
3000601 2  
3000641 2  
3001841 5  
3001881 5  
3001921 7  
3001961 9  
3002001 9  
3002041 6  
3002081 6  
3002121 4
```

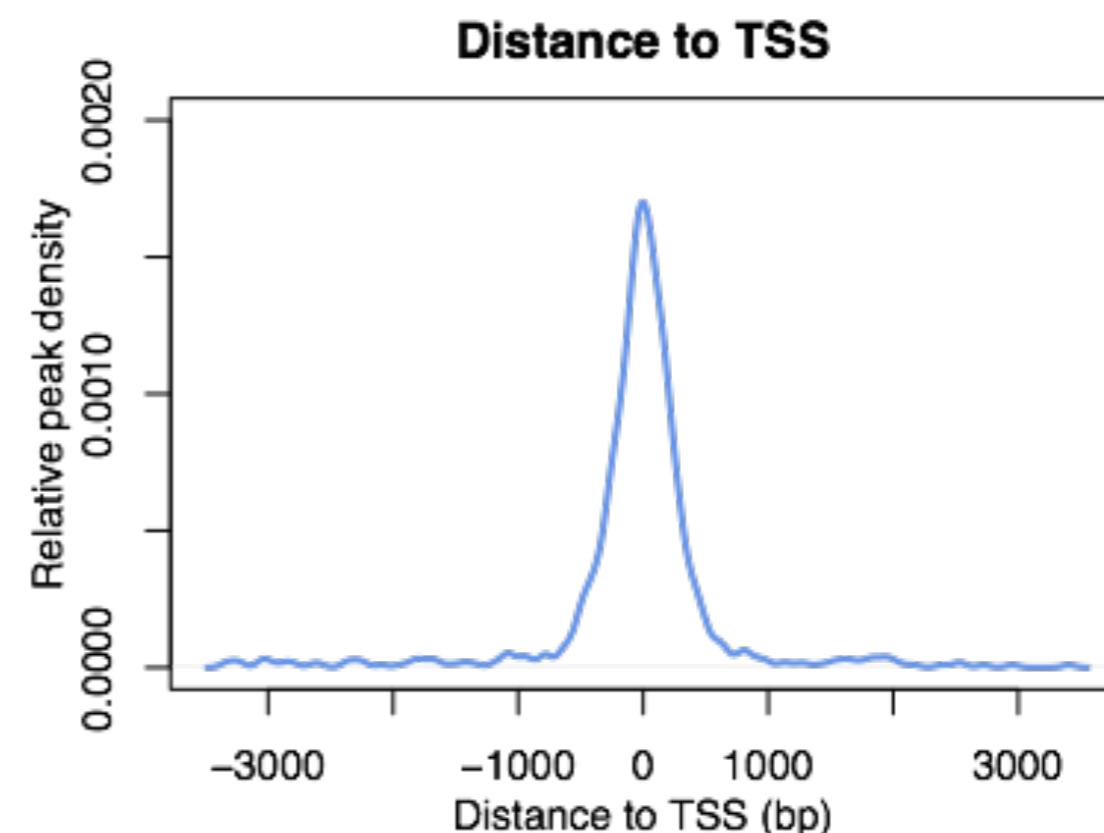
bigWig (converted from wig or bam)

indexed binary format

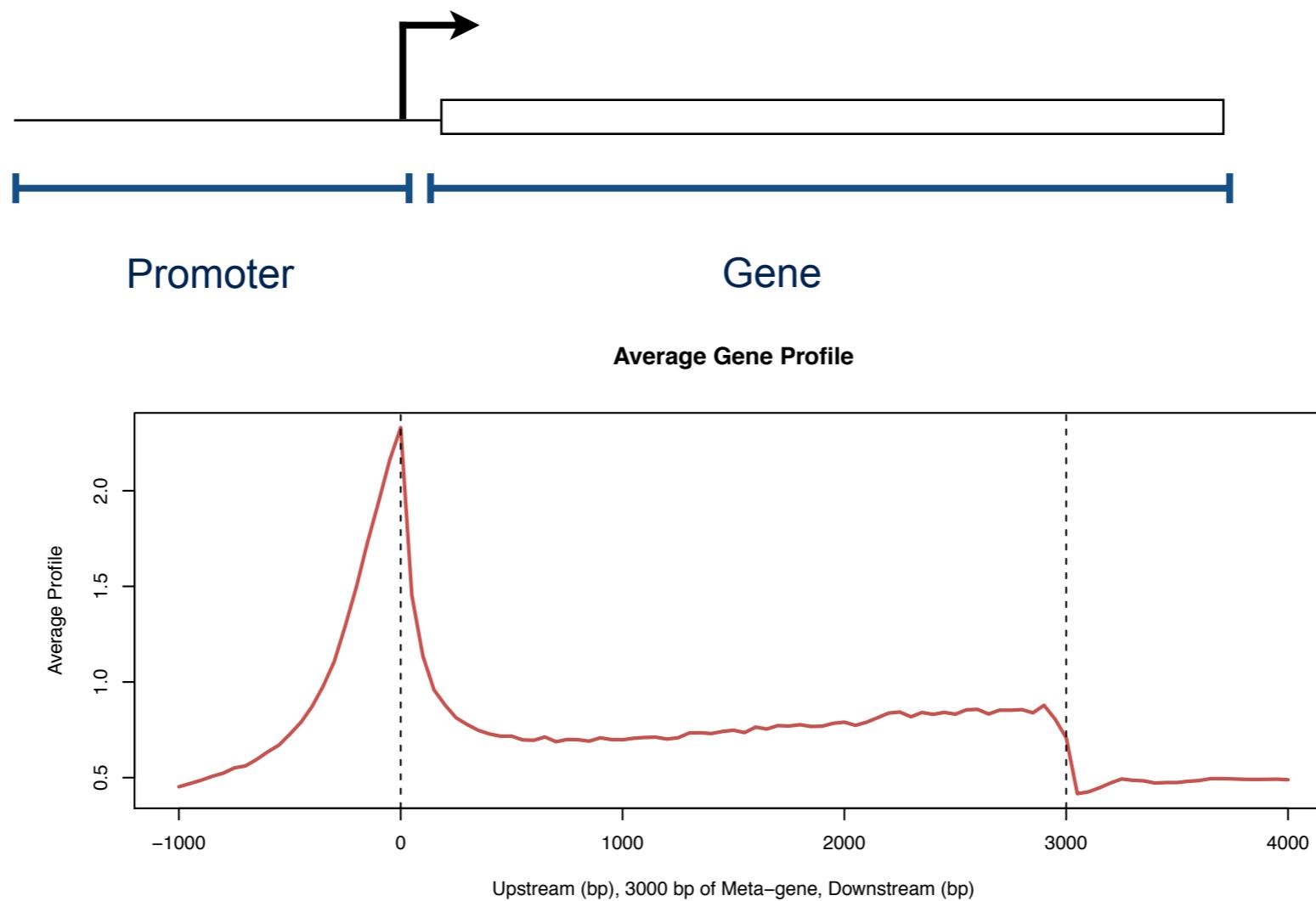
Profile around the TSS
using profile in wig



Peak distance to TSS distribution
using peaks in bed



Profile upstream and downstream TSS

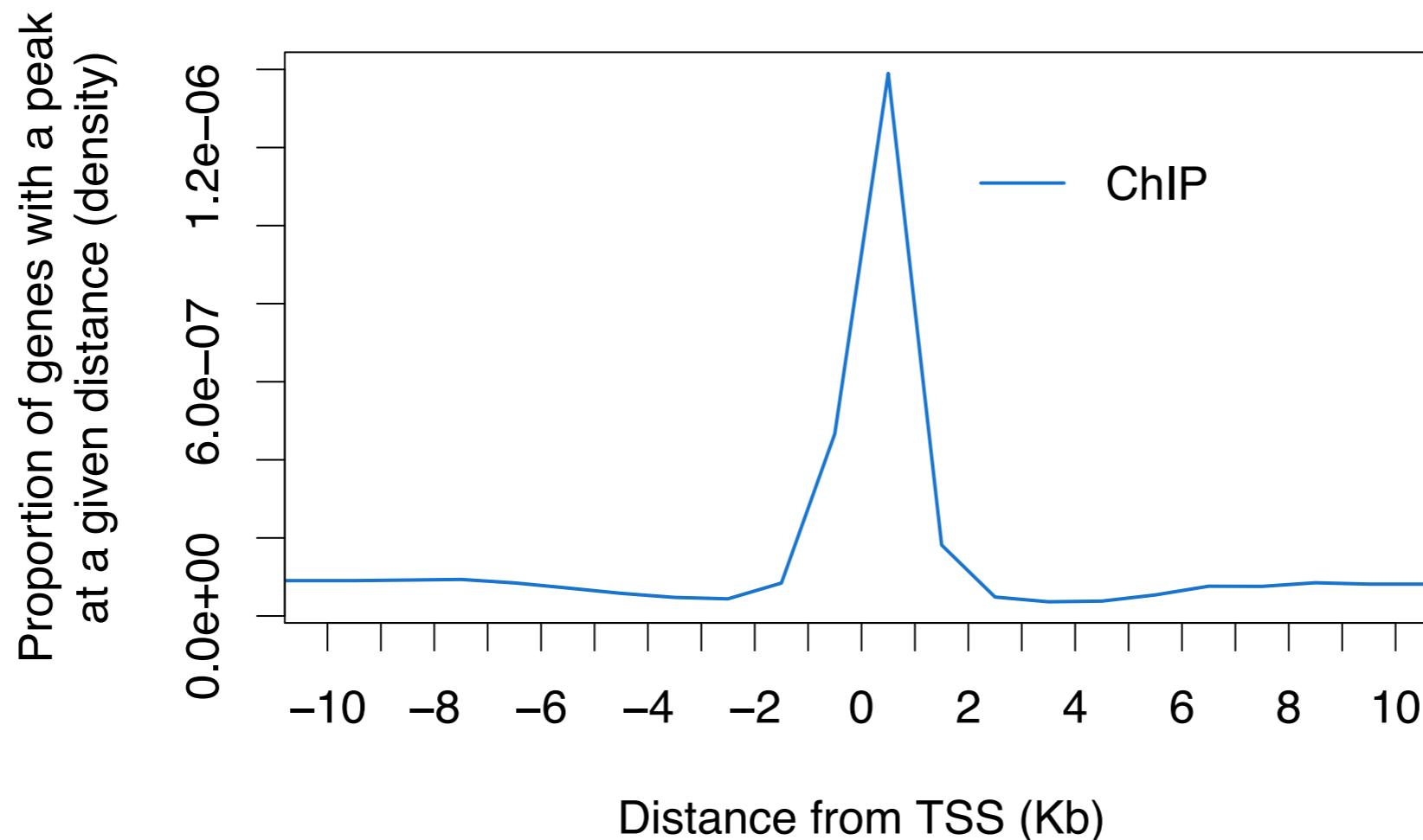


Galaxy: MakeTSSdist

Practice

INPUT: bed file with peaks

OUTPUT: peak distance to TSS distribution (density plot)

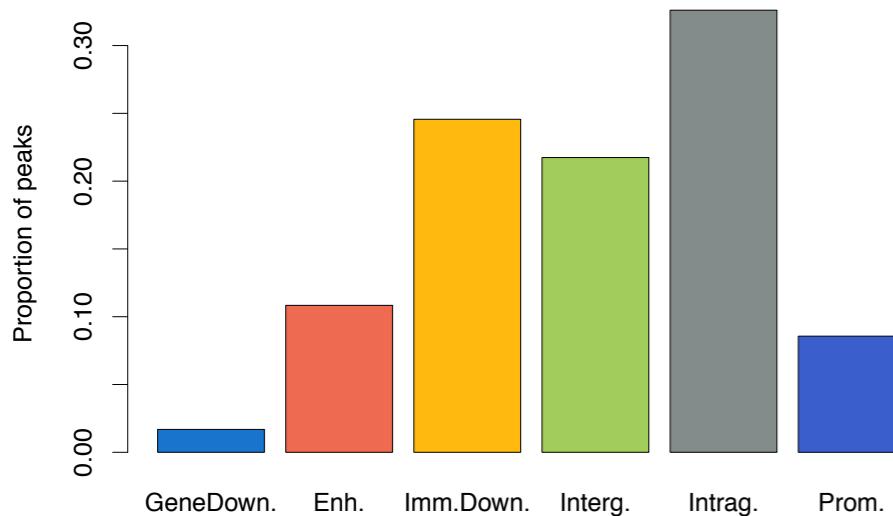


Galaxy: AnnotatePeaks

Practice

INPUT: bed file with peaks

OUTPUT: annotated peaks + distribution per category



Chromosome	Start	End	Max	Score	DistTSS	Type
chr1	3001827	3002328	3002077	55.28	659502	intergenic
chr1	3067471	3067948	3067709	50.67	593870	intergenic
chr1	3660316	3662844	3661580	352.43	-1	promoter
chr1	3842462	3842994	3842728	59.21	-181149	intergenic
chr1	3877254	3877710	3877482	52.72	-215903	intergenic
chr1	3939314	3939679	3939496	82.99	-277917	intergenic
chr1	4206037	4206512	4206274	50.86	144121	intergenic
chr1	4481463	4484213	4482838	268.57	3656	intragenic
chr1	4486799	4487684	4487241	88.18	-747	promoter
chr1	4561258	4562489	4561873	236.23	-75379	intergenic
chr1	4635092	4635552	4635322	52.32	140485	intergenic
chr1	4760253	4761284	4760768	111.13	15039	5kbDownstream
chr1	4773759	4776746	4775252	540.12	555	immediateDownstream
chr1	4797157	4800182	4798669	249.77	696	immediateDownstream
chr1	4841219	4842788	4842003	156.84	-6405	enhancer
chr1	4846807	4849844	4848325	377.92	-83	promoter
chr1	4873314	4873950	4873632	66.94	25224	intragenic
chr1	4885079	4885564	4885321	64.12	36913	intragenic

Annotation and visualisation

Species/Genome Assembly/Gene Set: Human Ensembl_CRCh38/hg38

Upstream Length: 5000

Downstream Length: 1000

The query peak file to be annotated: no file selected

File format: UCSC BED GFF3 EpiCenter Report Other text file

If other, please specify the delimiter and column numbers:
field delimiter: tab whitespace comma semicolon pipe
column number: chromosome: , start position: , end position:

The optional comparison peak files: set1 no file selected

set2 no file selected set3 no file selected

set4 no file selected set5 no file selected

File format: UCSC BED GFF3 EpiCenter Report Other text file

If other, please specify the delimiter and column numbers:
field delimiter: tab whitespace comma semicolon pipe
column number: chromosome: , start position: , end position:

Search distance to query peaks: 2000

SUBMIT **RESET**

<http://manticore.niehs.nih.gov:8080/pavis/>

Output Example

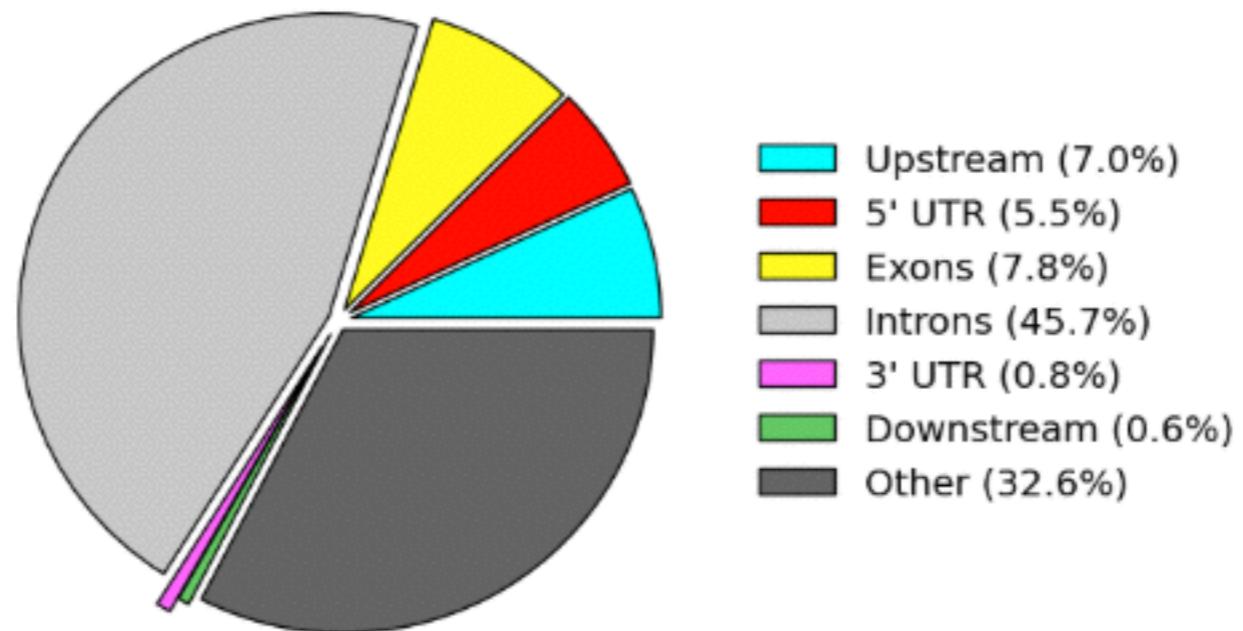
Peak Location Annotation ([Annotation File 119.01 kB](#))

Location	Query Peak	Number	Proportion	EnrichTest1	EnrichTest2	Comparison Peak
Upstream	Q-Upstream	210	7.0%	1.00e+00	4.35e-01	C-Upstream
5' UTR	Q-5UTR	166	5.5%	2.82e-100	1.36e-138	C-5UTR
Exons/CDS	Q-Exon	235	7.8%	3.25e-22	4.18e-59	C-Exon
Introns	Q-Intron	1371	45.7%	NA	1.00e+00	C-Intron
3' UTR	Q-3UTR	23	0.8%	1.00e+00	1.00e+00	C-3UTR
Downstream	Q-Downstream	18	0.6%	1.00e+00	1.00e+00	C-Downstream
Unclassified	NA	977	32.6%	NA	NA	NA

[The tab delimited form of the table](#)

Note: Upstream length was set to 5000 and Downstream length was set to 1000 (0=no limit).

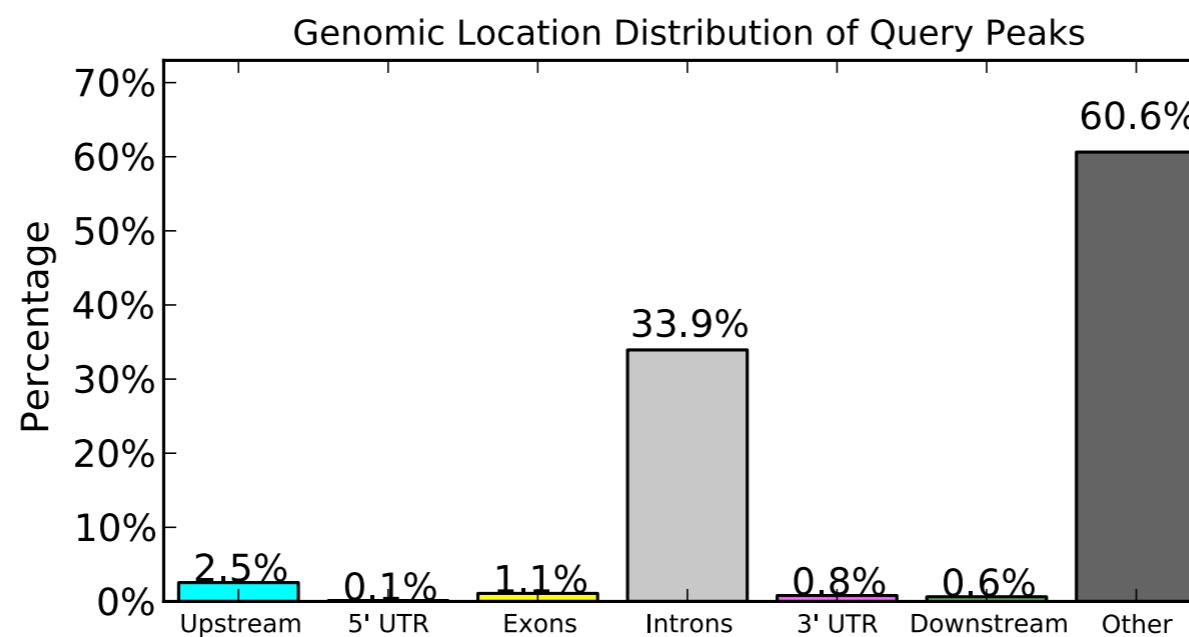
Distribution of Peaks in Relation to Genes



[Click here to download the Pie-Chart in the PDF format](#)

Detailed view

Chromosome	Loci	Start Loci	End	Gene ID	Gene Symbol	Strand	Distance to TSS
chr13		022690027	022690527	NM_000231	SGCG	+	+37218
chr13		023047991	023048491	NM_148957	TNFRSF19	+	+5733
chr13		023359572	023360072	NM_005932	MIPEP	-	+1765
chr13		023634753	023635253	NR_031753	MIR2276	+	+0449
chr13		024956993	024957493	NM_016529	ATP8A2	+	+113035
chr13		025197768	025198268	NM_016529	ATP8A2	+	+353810
chr13		025317576	025318076	NM_016529	ATP8A2	+	+473618



PAVIS

Optional practice

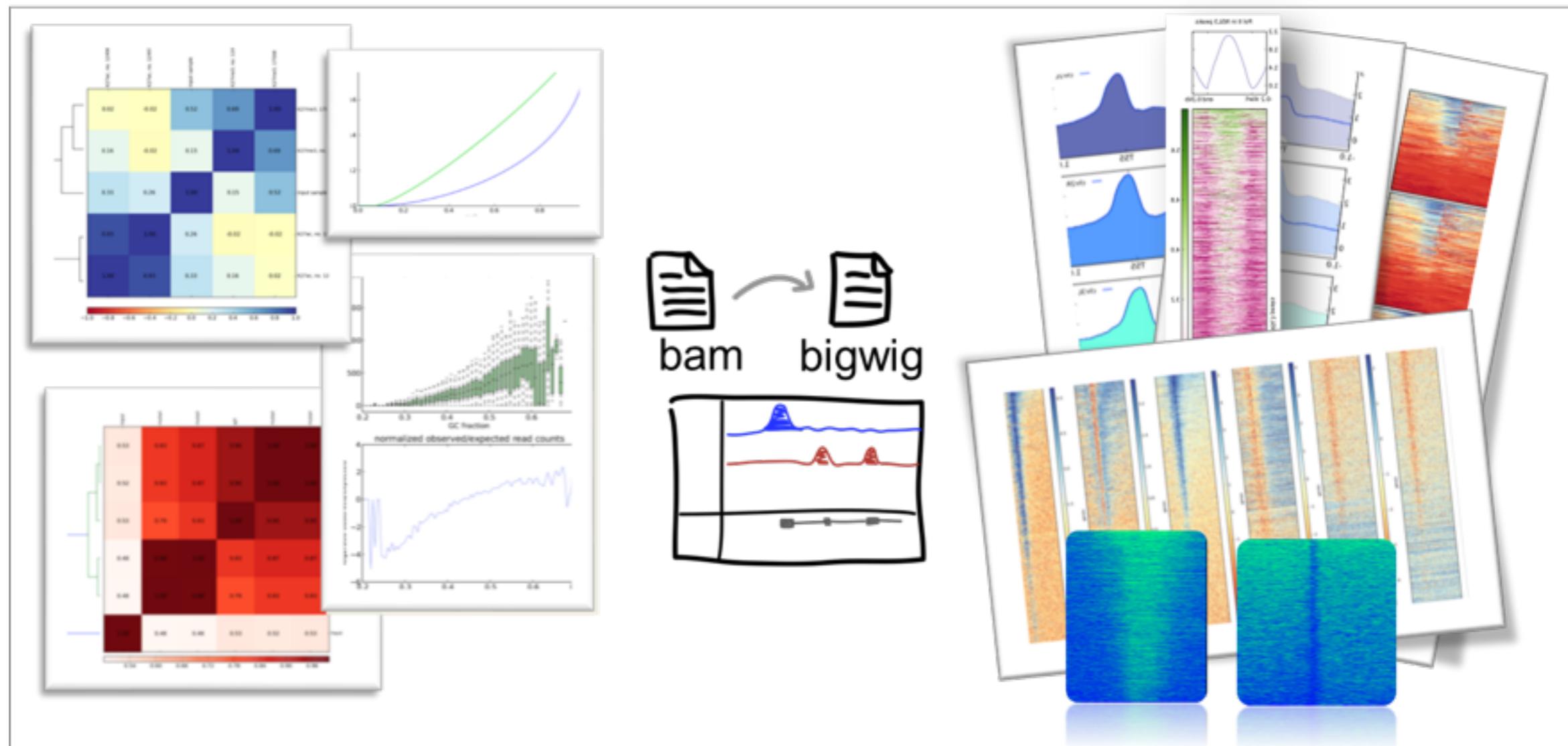
INPUT: peaks

OUTPUT: annotated peaks + figures

Chromosome	Loci	Start Loci	End	Gene ID	Gene Symbol	Strand	Distance to TSS
chr13		022690027	022690527	NM_000231	SGCG	+	+37218
chr13		023047991	023048491	NM_148957	TNFRSF19	+	+5733
chr13		023359572	023360072	NM_005932	MIPEP	-	+1765
chr13		023634753	023635253	NR_031753	MIR2276	+	+0449
chr13		024956993	024957493	NM_016529	ATP8A2	+	+113035
chr13		025197768	025198268	NM_016529	ATP8A2	+	+353810
chr13		025317576	025318076	NM_016529	ATP8A2	+	+473618

Fidel Ramírez^{1,†}, Friederike Dündar^{1,2,†}, Sarah Diehl¹, Björn A. Grüning³ and Thomas Manke^{1,*}

QUALITY CHECKS – FORMAT CONVERSION & NORMALIZATION – PLOTTING

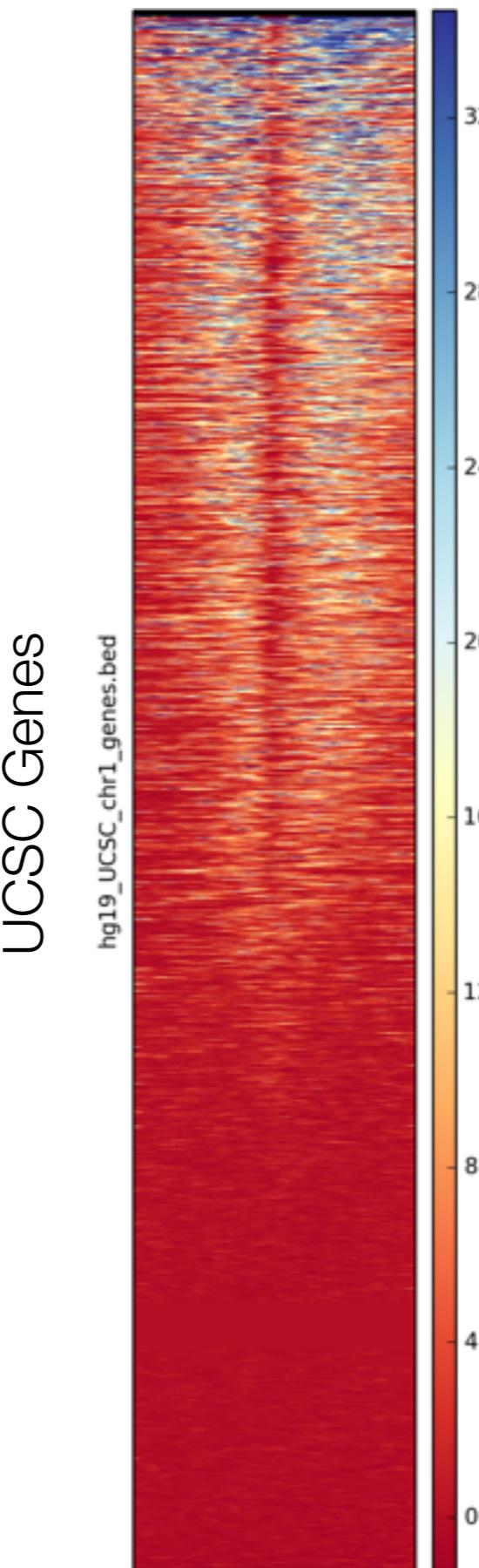


deepTools: heatmap

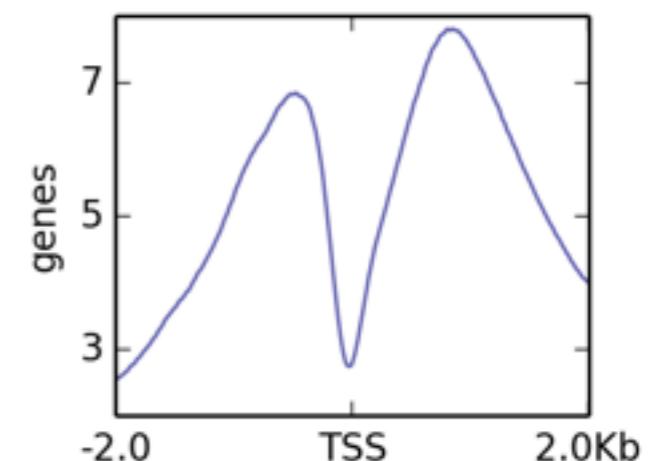
INPUT: ChIP bigWig + bed of feature

OUTPUT: heatmap

TSS



Practice

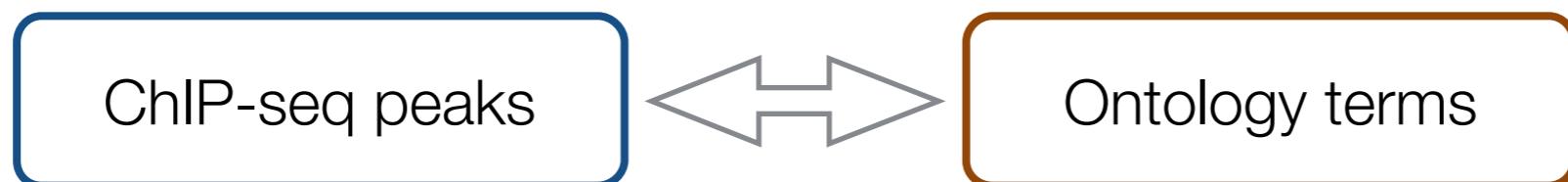


GREAT

GREAT improves functional interpretation of
cis-regulatory regions

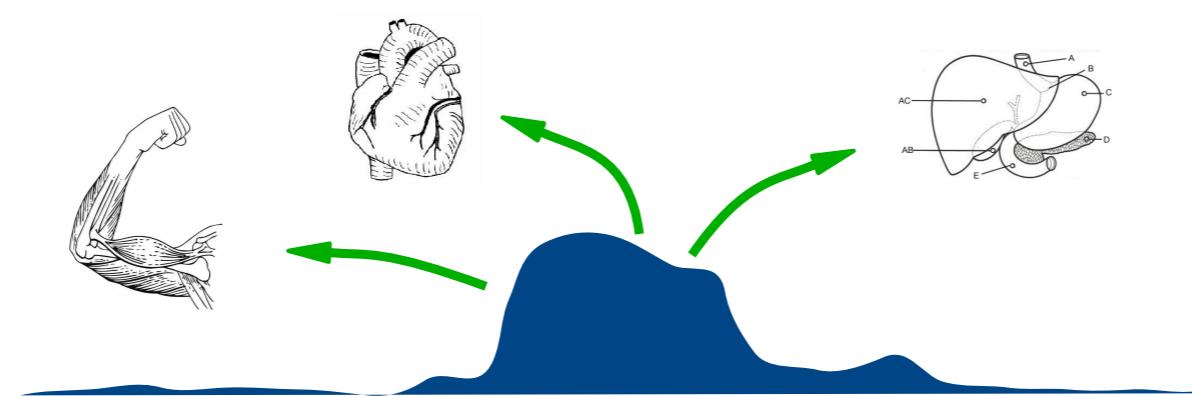
Cory Y McLean¹, Dave Bristor^{1,2}, Michael Hiller², Shoa L Clarke³, Bruce T Schaar², Craig B Lowe⁴,
Aaron M Wenger¹ & Gill Bejerano^{1,2}

Functional annotation of *cis*-regulatory regions



Ontology terms

GO Molecular Function
GO Biological Process
Disease Ontology
Pathways
...



Species Assembly

- Human: GRCh37 ([UCSC hg19, Feb/2009](#))
- Human: NCBI build 36.1 ([UCSC hg18, Mar/2006](#))
- Mouse: NCBI build 37 ([UCSC mm9, Jul/2007](#))
- Zebrafish: Wellcome Trust Zv9 ([danRer7, Jul/2010](#))



Zebrafish CNE set

*Can I use a different species or assembly?***Test regions**

- BED file: [Choose File](#) no file selected

- BED data:

*What should my test regions file contain?**How can I create a test set from a UCSC Genome Browser annotation track?***Background regions**

- Whole genome

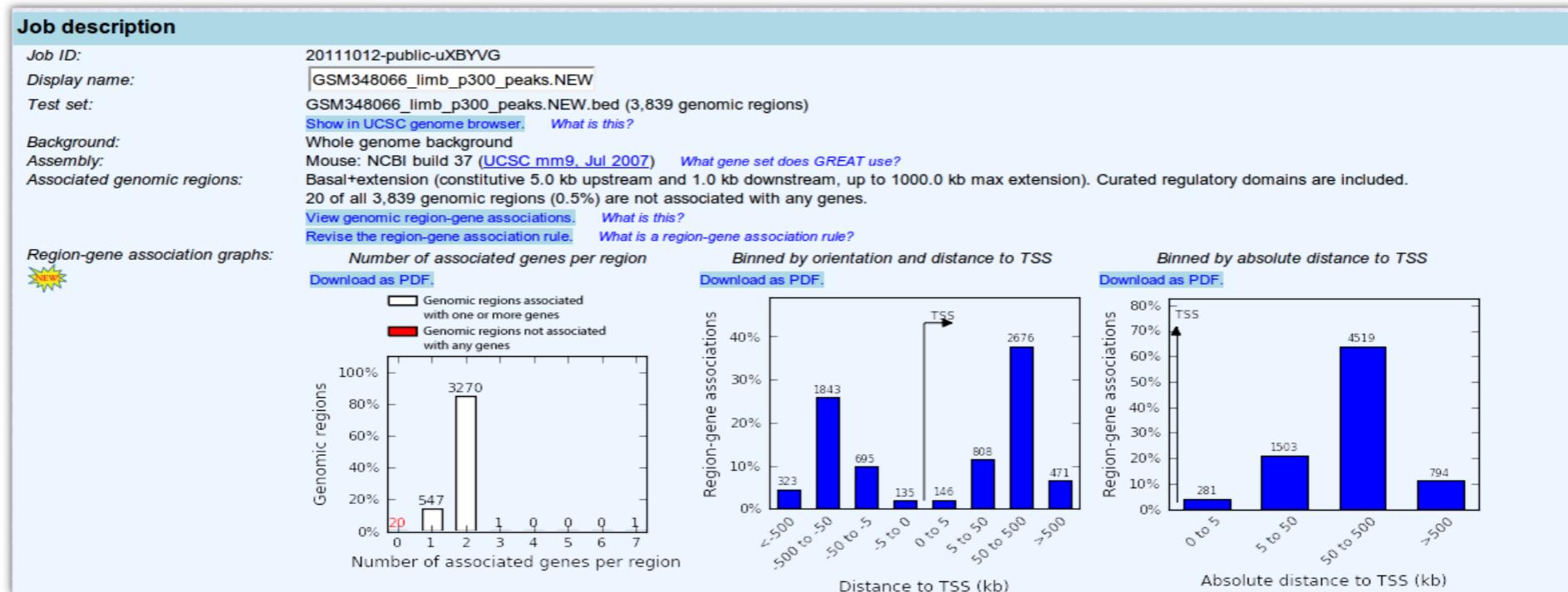
- BED file: [Choose File](#) no file selected

- BED data:

*When should I use a background set?**What should my background regions file contain?***Association rule settings**[Show settings »](#)[Submit](#)[Reset](#)

Note: Only human (hg19 and hg18), mouse (mm9) and zebrafish (danRer7) genomes are supported

GREAT

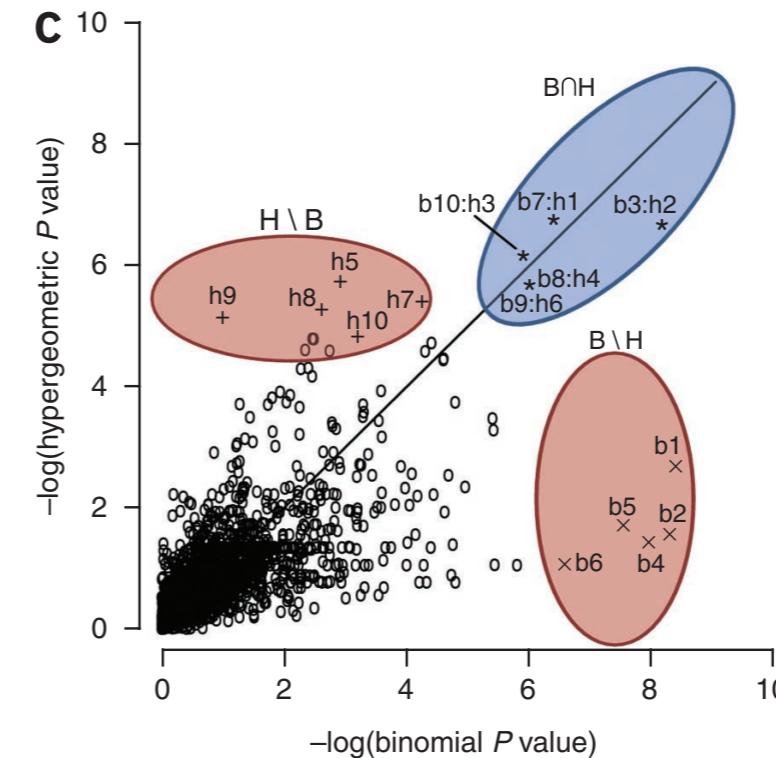
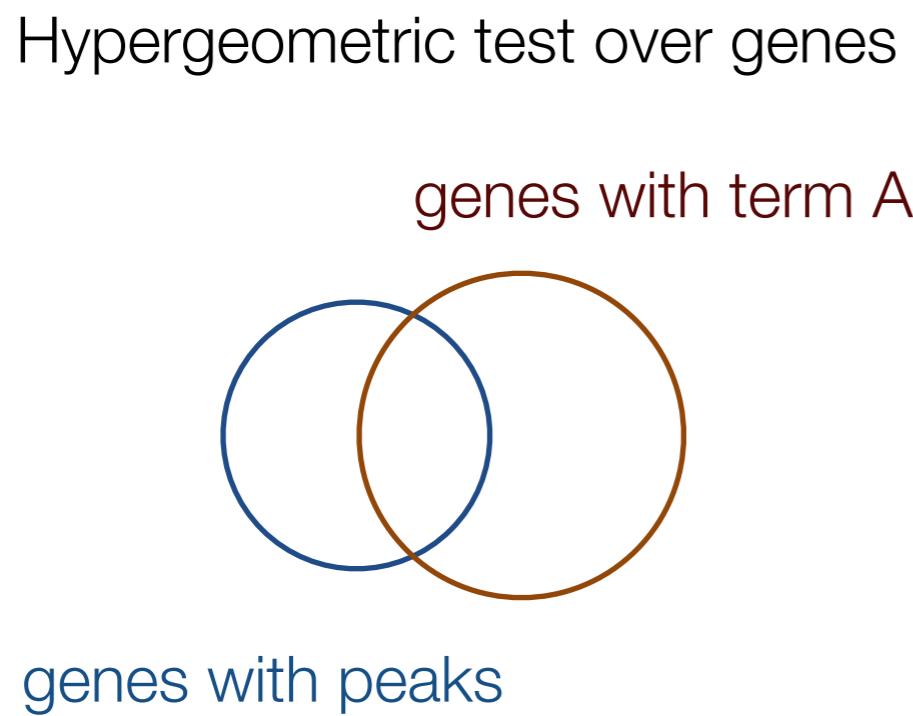


X Mouse Phenotype Global Controls

Table controls: Export ▾ Shown top rows in this table: 20 Set Term annotation count: Min: 1 Max: Inf Set

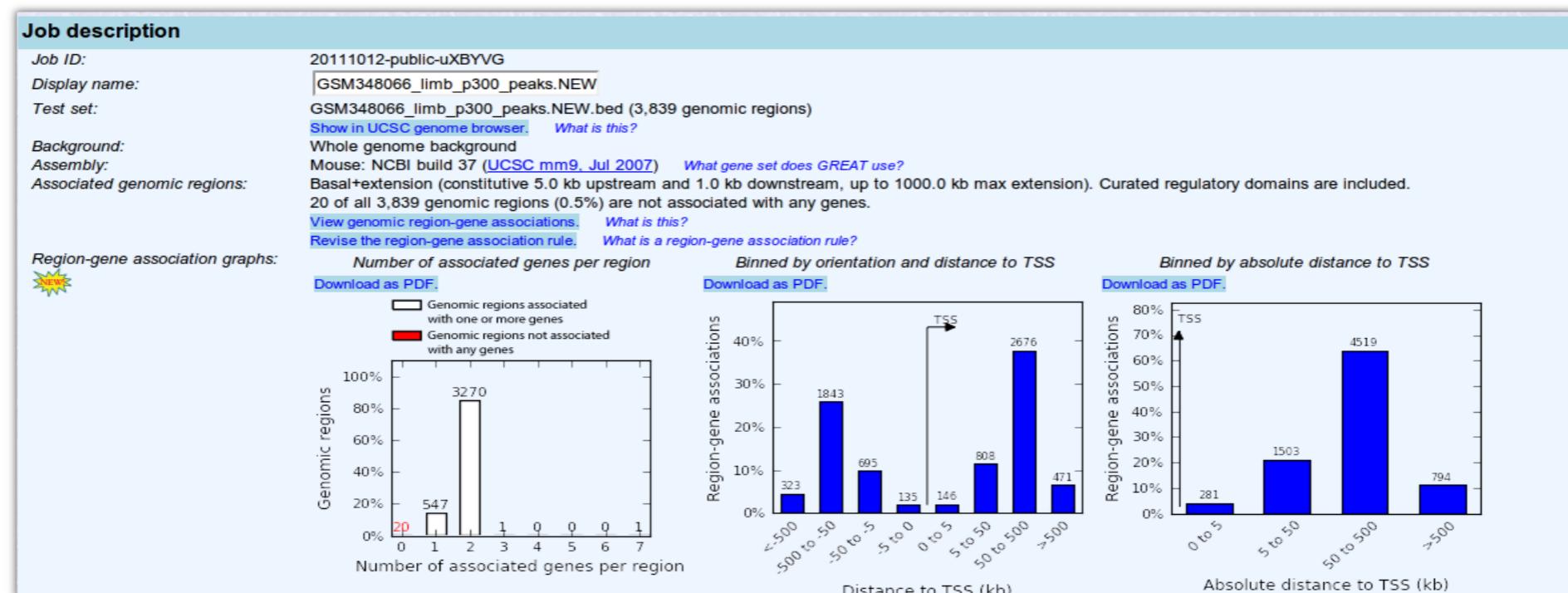
Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
abnormal limbs/digits/tail morphology	2	2.0559e-91	6.6837e-88	2.1465	780	20.32%	6	2.5295e-40	2.2020	278	681	8.31%
abnormal craniofacial morphology	3	9.3822e-91	2.0334e-87	2.0082	887	23.10%	10	8.9231e-36	2.0382	297	786	8.88%
abnormal limb morphology	5	2.4990e-80	3.2497e-77	2.3077	604	15.73%	9	7.4787e-37	2.4541	202	444	6.04%
abnormal appendicular skeleton morphology	10	3.0255e-70	1.9672e-67	2.3450	517	13.47%	17	3.9549e-30	2.4098	172	385	5.14%
abnormal skeleton extremities morphology	12	3.2687e-69	1.7711e-66	2.3724	499	13.00%	21	7.0557e-29	2.4222	163	363	4.87%
abnormal paw/hand/foot morphology	13	4.0300e-69	2.0156e-66	2.6813	404	10.52%	23	5.4918e-28	2.7186	126	250	3.77%
abnormal head morphology	14	6.4657e-67	3.0029e-64	2.0134	672	17.50%	25	2.9042e-27	2.0562	223	585	6.67%
abnormal digit morphology	18	1.0543e-61	3.8084e-59	2.6982	358	9.33%	36	1.2033e-25	2.7998	109	210	3.26%
abnormal cartilage morphology	23	7.3728e-58	2.0843e-55	2.3432	430	11.20%	29	1.1337e-26	2.5089	140	301	4.19%
abnormal skeleton development	24	3.5769e-56	9.6904e-54	2.0833	530	13.81%	38	5.2377e-25	2.1414	185	466	5.53%
abnormal long bone morphology	25	4.6593e-56	1.2118e-53	2.3374	419	10.91%	43	4.9983e-24	2.3823	140	317	4.19%

GREAT



INPUT: bed file with peaks

OUTPUT: Enriched GO terms and functions

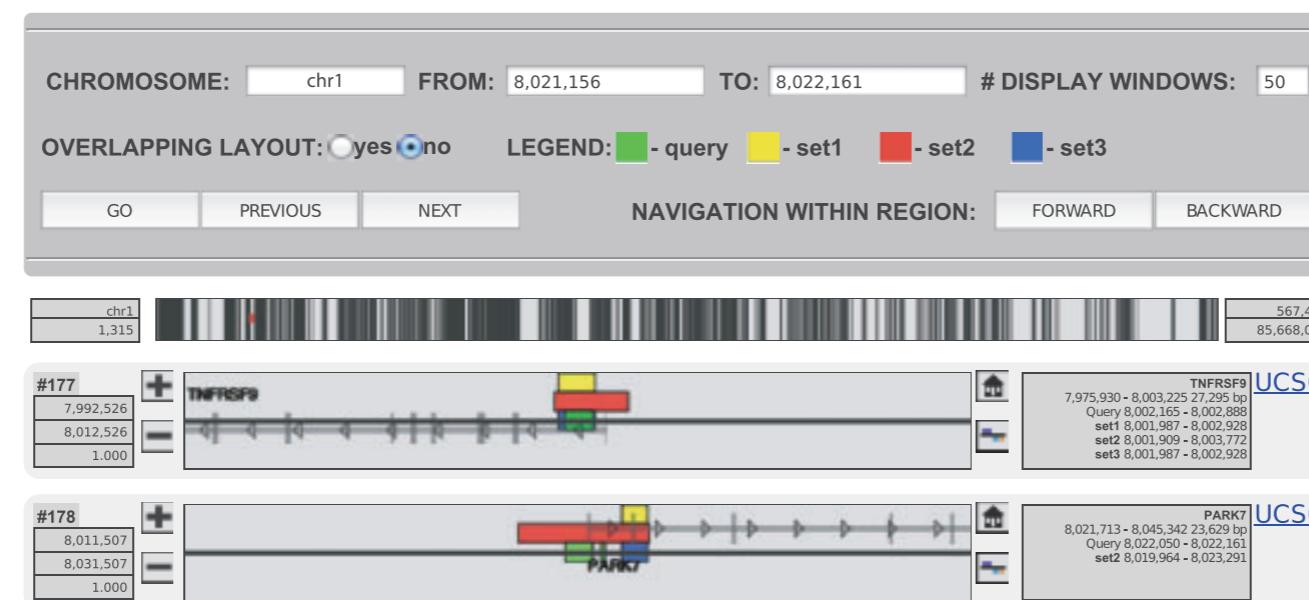
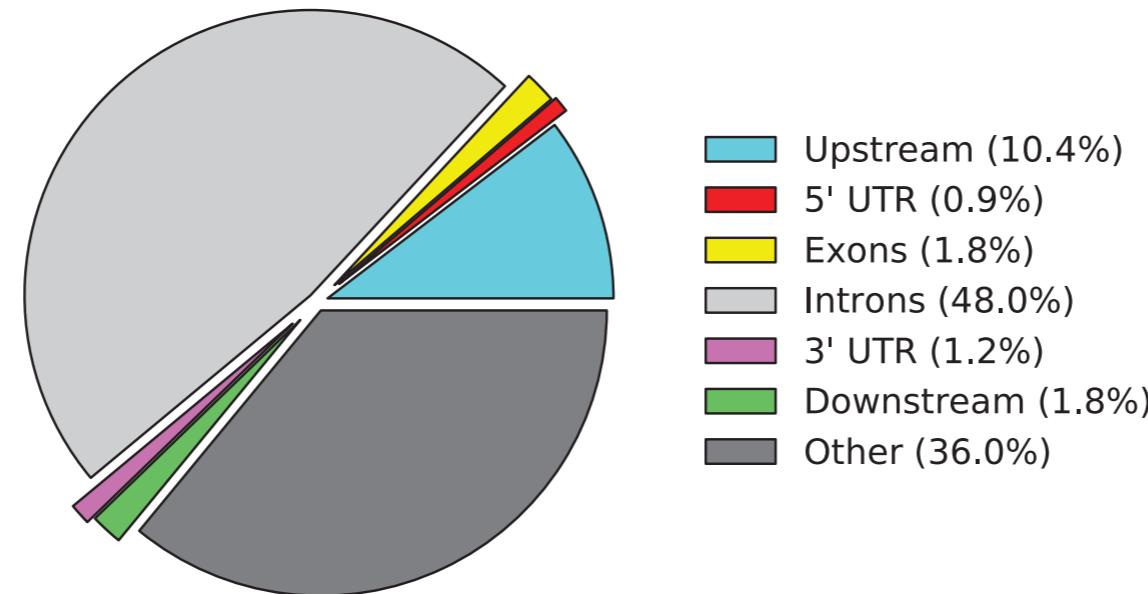


X Mouse Phenotype Global Controls

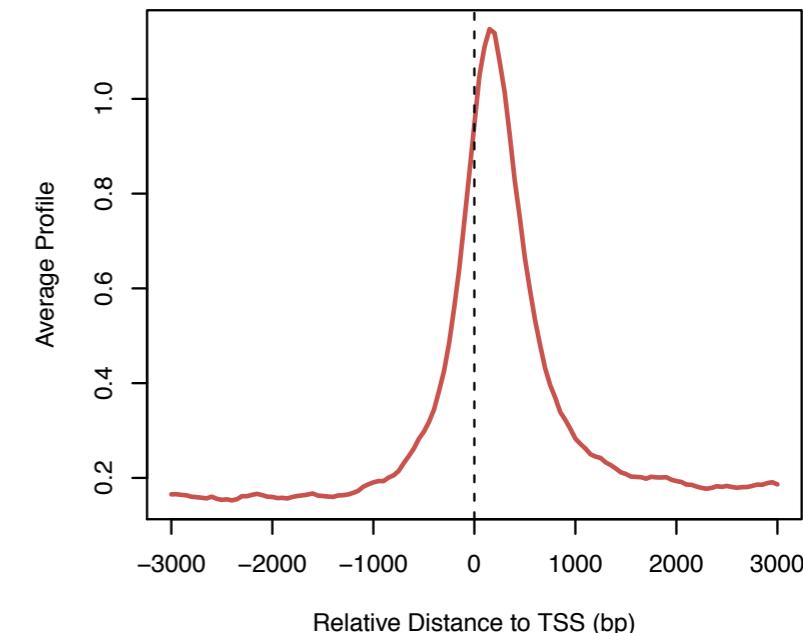
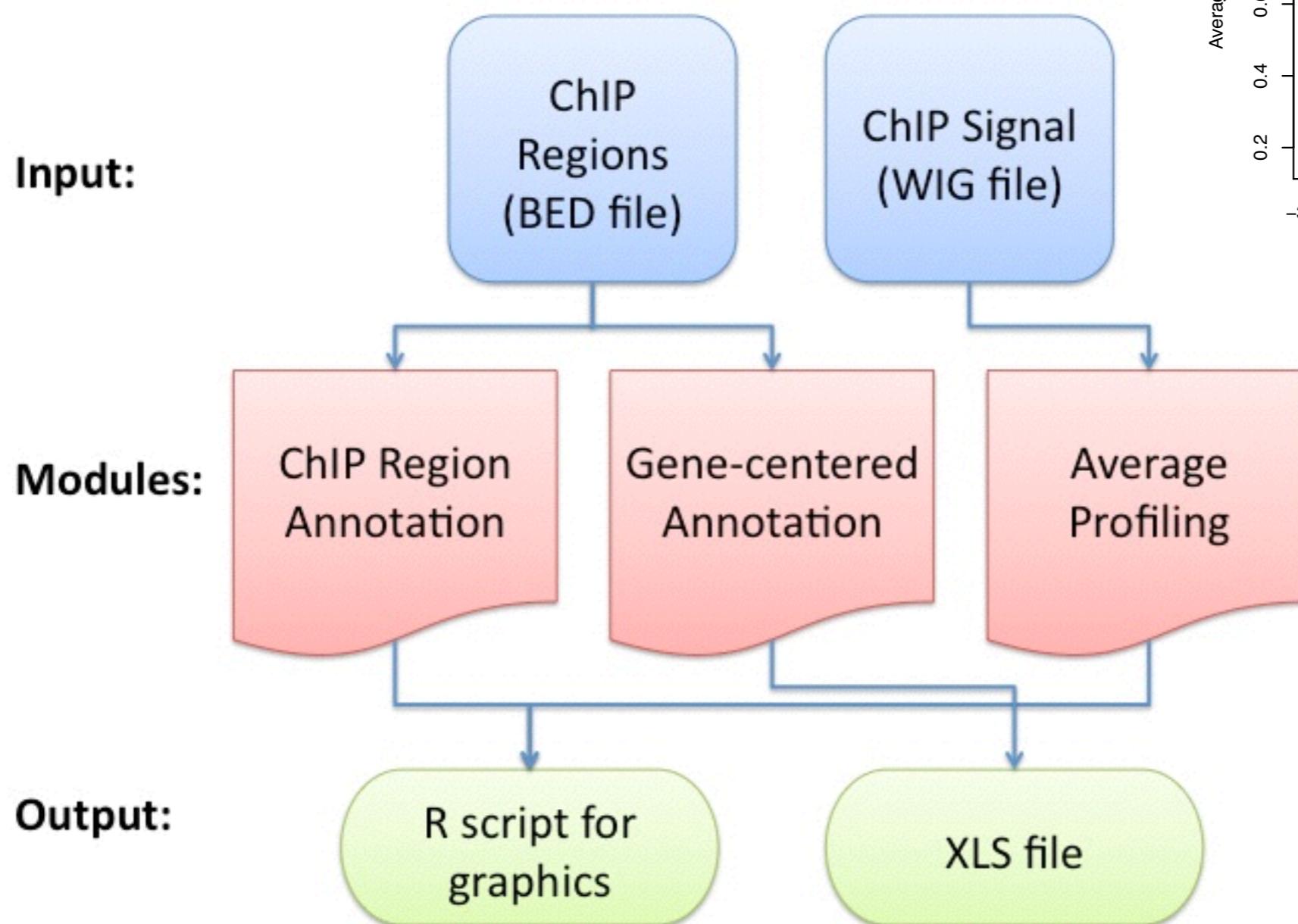
Table controls: Export ▾ Shown top rows in this table: 20 Set Term annotation count: Min: 1 Max: Inf Set

Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
abnormal limbs/digits/tail morphology	2	2.0559e-91	6.6837e-88	2.1465	780	20.32%	6	2.5295e-40	2.2020	278	681	8.31%
abnormal craniofacial morphology	3	9.3822e-91	2.0334e-87	2.0082	887	23.10%	10	8.9231e-36	2.0382	297	786	8.88%
abnormal limb morphology	5	2.4990e-80	3.2497e-77	2.3077	604	15.73%	9	7.4787e-37	2.4541	202	444	6.04%
abnormal appendicular skeleton morphology	10	3.0255e-70	1.9672e-67	2.3450	517	13.47%	17	3.9549e-30	2.4098	172	385	5.14%
abnormal skeleton extremities morphology	12	3.2687e-69	1.7711e-66	2.3724	499	13.00%	21	7.0557e-29	2.4222	163	363	4.87%
abnormal paw/hand/foot morphology	13	4.0300e-69	2.0156e-66	2.6813	404	10.52%	23	5.4918e-28	2.7186	126	250	3.77%
abnormal head morphology	14	6.4657e-67	3.0029e-64	2.0134	672	17.50%	25	2.9042e-27	2.0562	223	585	6.67%
abnormal digit morphology	18	1.0543e-61	3.8084e-59	2.6982	358	9.33%	36	1.2033e-25	2.7998	109	210	3.26%
abnormal cartilage morphology	23	7.3728e-58	2.0843e-55	2.3432	430	11.20%	29	1.1337e-26	2.5089	140	301	4.19%
abnormal skeleton development	24	3.5769e-56	9.6904e-54	2.0833	530	13.81%	38	5.2377e-25	2.1414	185	466	5.53%
abnormal long bone morphology	25	4.6593e-56	1.2118e-53	2.3374	419	10.91%	43	4.9983e-24	2.3823	140	317	4.19%

A comprehensive framework for the analysis of ChIP-seq data



CEAS (Cis-regulatory Element Annotation System)

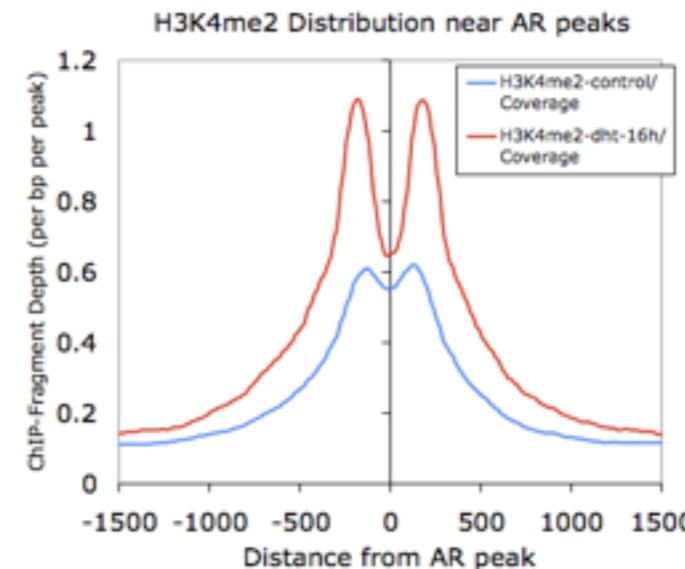
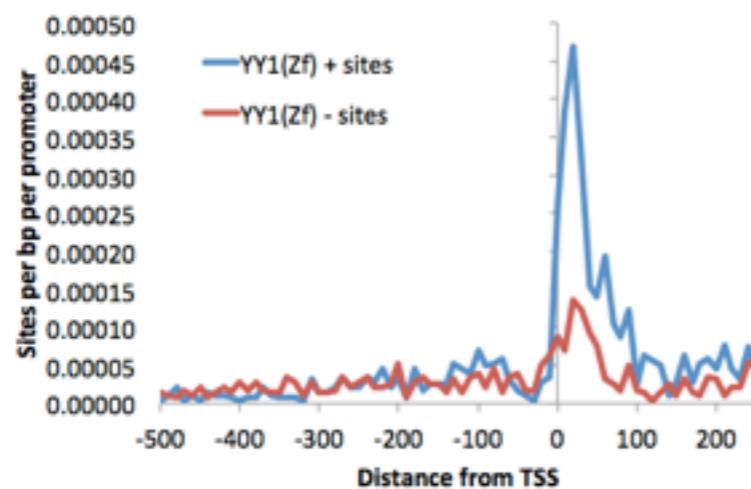


HOMER

Simple Combinations of Lineage-Determining Transcription Factors Prime *cis*-Regulatory Elements Required for Macrophage and B Cell Identities

Sven Heinz,^{1,7} Christopher Benner,^{1,7} Nathanael Spann,^{1,7} Eric Bertolino,⁴ Yin C. Lin,³ Peter Laslo,⁶ Jason X. Cheng,⁴ Cornelis Murre,³ Harinder Singh,^{4,5} and Christopher K. Glass^{1,2,*}

Motif discovery and NGS data analysis



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	PeakID	Chr	Start	End	Strand	Peak Sco	Focus Ra	Annotation	Detailed Anno	Distance to T	Nearest Pror	PromoterID	Nearest Unig	Nearest Refs	Nearest Ense	Gene Name	Gene Alias	Gene Descrip
2	chr18-1	chr18	69007968	69008268	+	593	0.939	intron (NR_03)	intron (NR_03)	74595	NR_034133	400655	Hs.579378	NR_034133		LOC400655	-	hypothetical
3	chr9-1	chr9	88209966	88210266	+	531.9	0.946	Intergenic	Intergenic	-50894	NM_001185	79670	Hs.597057	NM_001185	ENSG000000 ZCCHC6	DKFZp666B1	zinc finger, C	
4	chr14-1	chr14	62337073	62337373	+	505.4	0.918	intron (NM_17)	intron (NM_17)	244485	NM_172375	27133	Hs.27043	NM_139318	ENSG000001 KCNHS	EAG2 H-EAG	potassium vc	
5	chr17-1	chr17	5076243	5076543	+	492.1	0.936	intron (NR_03)	intron (NR_03)	2414	NM_207103	388325	Hs.462080	NM_207103	ENSG000001 C17orf87	FLJ32580 M	chromosome	
6	chr17-2	chr17	47851714	47852014	+	476.2	0.824	Intergenic	Intergenic	-259488	NM_001082	56934	Hs.463466	NM_001082	ENSG000001 CA10	CA-RPX CAR	carbonic anh	
7	chr10-1	chr10	98420680	98420980	+	474.9	0.967	intron (NM_15)	intron (NM_15)	49439	NM_152309	118788	Hs.310456	NM_152309	ENSG000001 PIK3AP1	BCAP RP11-	phosphoinos	
8	chr9-2	chr9	81294389	81294689	+	456.3	0.957	Intergenic	Intergenic	-82159	NM_007005	7091	Hs.444213	NM_007005	ENSG000001 TLE4	BCE-1 BCE1	transducin-lil	
9	chr14-2	chr14	36817736	36818036	+	452.3	0.757	intron (NM_13)	intron (NM_13)	81017	NM_001195	145282	Hs.660396	NM_001195	ENSG000001 MIPOL1	DKFZp313M	mirror-image	
10	chr18-2	chr18	20049825	20050125	+	449.7	0.853	intron (NM_08)	intron (NM_08)	56219	NM_018030	114876	Hs.370725	NM_018030	ENSG000001 OSBPL1A	FLJ10217 OF	oxysterol bin	
11	chr7-1	chr7	12226829	12227129	+	445.7	0.901	intron (NM_01)	intron (NM_01)	9606	NM_001134	54664	Hs.396358	NM_001134	ENSG000001 TMEM106B	FLJ11273 M	transmembr	
12	chr14-3	chr14	88712188	88712488	+	443.1	0.844	intron (NM_00)	intron (NM_00)	240869	NM_005197	1112	Hs.621371	NM_001085	ENSG000000 FOXN3	C14orf116 C	forkhead box	
13	chr18-3	chr18	62951924	62952224	+	443.1	0.947	Intergenic	Intergenic	-382689	NR_033921	643542	Hs.652901	NR_033921		LOC643542	-	hypothetical
14	chr3-1	chr3	32196769	32197069	+	443.1	0.87	Intergenic	Intergenic	-58256	NM_178868	152189	Hs.154986	NM_178868	ENSG000001 CMTM8	CKLFSF8 CKL	CKLF-like MA	
15	chr11-1	chr11	110685448	110685748	+	425.8	0.907	Intergenic	Intergenic	-9849	NR_034154	399948	Hs.729225	NR_034154		C11orf92	DKFZp781P1	chromosome
16	chr4-1	chr4	81755366	81755666	+	423.2	0.908	intron (NM_15)	intron (NM_15)	279618	NM_152770	255119	Hs.527104	NM_152770	ENSG000001 C4orf22	MGC35043	chromosome	

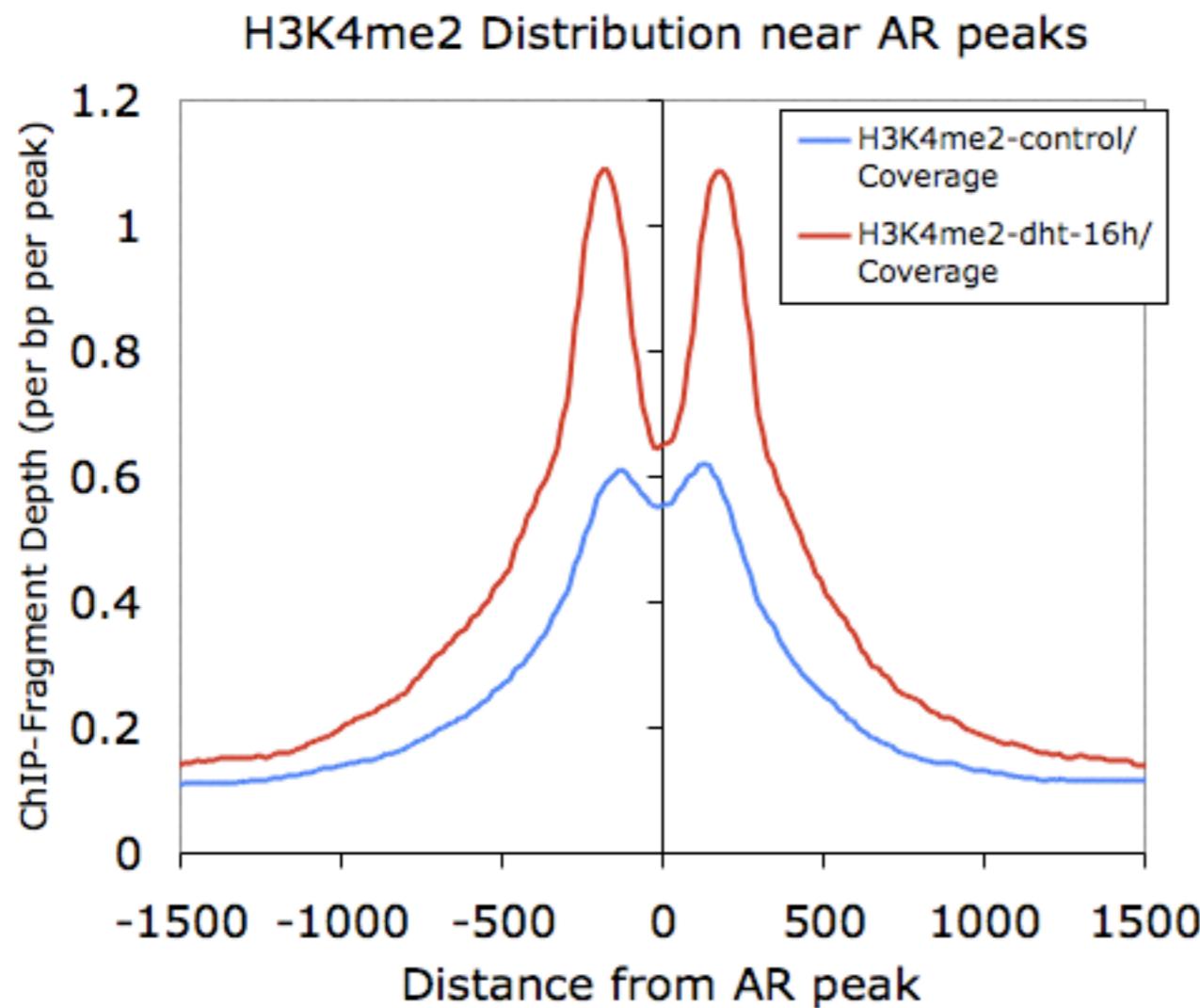
<http://homer.salk.edu/homer/>

HOMER: annotate peaks

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	PeakID	Chr	Start	End	Strand	Peak Sco	Focus Rz	Annotation	Detailed Anno	Distance to T	Nearest Prox	PromoterID	Nearest Unig	Nearest Refs	Ense	Gene Name	Gene Alias	Gene Descrip
2	chr18-1	chr18	69007968	69008268	+	593	0.939	intron (NR_03-	intron (NR_03-	74595	NR_034133	400655	Hs.579378	NR_034133		LOC400655	-	hypothetical
3	chr9-1	chr9	88209966	88210266	+	531.9	0.946	Intergenic	Intergenic	-50894	NM_0011854	79670	Hs.597057	NM_0011854	ENSG000000000000	ZCCHC6	DKFZp666B1	zinc finger, C
4	chr14-1	chr14	62337073	62337373	+	505.4	0.918	intron (NM_17	intron (NM_17	244485	NM_172375	27133	Hs.27043	NM_139318	ENSG0000001	KCNH5	EAG2 H-EAG	potassium va
5	chr17-1	chr17	5076243	5076543	+	492.1	0.936	intron (NR_03-	intron (NR_03-	2414	NM_207103	388325	Hs.462080	NM_207103	ENSG0000001	C17orf87	FLJ32580 M	chromosome
6	chr17-2	chr17	47851714	47852014	+	476.2	0.824	Intergenic	Intergenic	-259488	NM_0010821	56934	Hs.463466	NM_0010821	ENSG0000001	CA10	CA-RPX CAR	carbonic anh
7	chr10-1	chr10	98420680	98420980	+	474.9	0.967	intron (NM_15	intron (NM_15	49439	NM_152309	118788	Hs.310456	NM_152309	ENSG0000001	PIK3AP1	BCAP RP11-	phosphoinos
8	chr9-2	chr9	81294389	81294689	+	456.3	0.957	Intergenic	Intergenic	-82159	NM_007005	7091	Hs.444213	NM_007005	ENSG0000001	TLE4	BCE-1 BCE1	transducin-lil
9	chr14-2	chr14	36817736	36818036	+	452.3	0.757	intron (NM_13	intron (NM_13	81017	NM_001195	145282	Hs.660396	NM_001195	ENSG0000001	MIPOL1	DKFZp313M	mirror-image
10	chr18-2	chr18	20049825	20050125	+	449.7	0.853	intron (NM_06	intron (NM_06	56219	NM_018030	114876	Hs.370725	NM_018030	ENSG0000001	OSBPL1A	FLJ10217 O	oxysterol bin
11	chr7-1	chr7	12226829	12227129	+	445.7	0.901	intron (NM_01	intron (NM_01	9606	NM_001134	54664	Hs.396358	NM_001134	ENSG0000001	TMEM106B	FLJ11273 M	transmembr
12	chr14-3	chr14	88712188	88712488	+	443.1	0.844	intron (NM_00	intron (NM_00	240869	NM_005197	1112	Hs.621371	NM_001085	ENSG0000000	FOXN3	C14orf116 C	forkhead box
13	chr18-3	chr18	62951924	62952224	+	443.1	0.947	Intergenic	Intergenic	-382689	NR_033921	643542	Hs.652901	NR_033921		LOC643542	-	hypothetical
14	chr3-1	chr3	32196769	32197069	+	443.1	0.87	Intergenic	Intergenic	-58256	NM_178868	152189	Hs.154986	NM_178868	ENSG0000001	CMTM8	CKLFSF8 CKL	CKLF-like MA
15	chr11-1	chr11	110685448	110685748	+	425.8	0.907	Intergenic	Intergenic	-9849	NR_034154	399948	Hs.729225	NR_034154		C11orf92	DKFZp781P1	chromosome
16	chr4-1	chr4	81755366	81755666	+	423.2	0.908	intron (NM_15	intron (NM_15	279618	NM_152770	255119	Hs.527104	NM_152770	ENSG0000001	C4orf22	MGC35043	chromosome

- 1 Peak ID
- 2 Chromosome
- 3 Peak start position
- 4 Peak end position
- 5 Strand
- 6 Peak Score
- 7 FDR/Peak Focus Ratio/Region Size
- 8 Annotation (i.e. Exon, Intron, ...)
- 9 Detailed Annotation (Exon, Intron etc. + CpG Islands, repeats, etc.)
- 10 Distance to nearest RefSeq TSS
- 11 Nearest TSS: Native ID of annotation file
- 12 Nearest TSS: Entrez Gene ID
- 13 Nearest TSS: Unigene ID
- 14 Nearest TSS: RefSeq ID
- 15 Nearest TSS: Ensembl ID
- 16 Nearest TSS: Gene Symbol
- 17 Nearest TSS: Gene Aliases
- 18 Nearest TSS: Gene description
- 19 Additional columns depend on options selected when running the program.

HOMER: compare peaks



Peak Co-Occurrence Statistics
Co-Bound Peaks
Differentially Bound Peaks

[Home](#) [About](#) [Contact](#)

[Home](#)

ReMap

An integrative ChIP-seq analysis of regulatory elements

Transcription Factors

Cell Types

Annotation Tool

Downloads

Welcome to ReMap an integrative analysis of transcription factor ChIP-seq experiments publicly available merged with the Encode dataset. Here we propose an extensive regulatory catalogue of 8 million transcription factor binding sites from 237 transcription factors (TFs). Among those factors 50 TFs are common with Encode, 82 TFs are Public specific and 105 Encode specific. The results of this analysis is available to browse or download either for a given transcription factor or for the entire dataset. Interactive [UCSC Genome Browser tracks](#) are also available.

A donut chart illustrating the distribution of transcription factors. The segments represent different categories: Public TFs (82), Common TFs (50), and Encode TFs (105).

Category	Count
Public TFs	82
Common TFs	50
Encode TFs	105

A donut chart showing the percentage of Public sites that overlap with the Encode dataset. The inner ring is grey and the outer ring is green.

57% Public sites overlap Encode

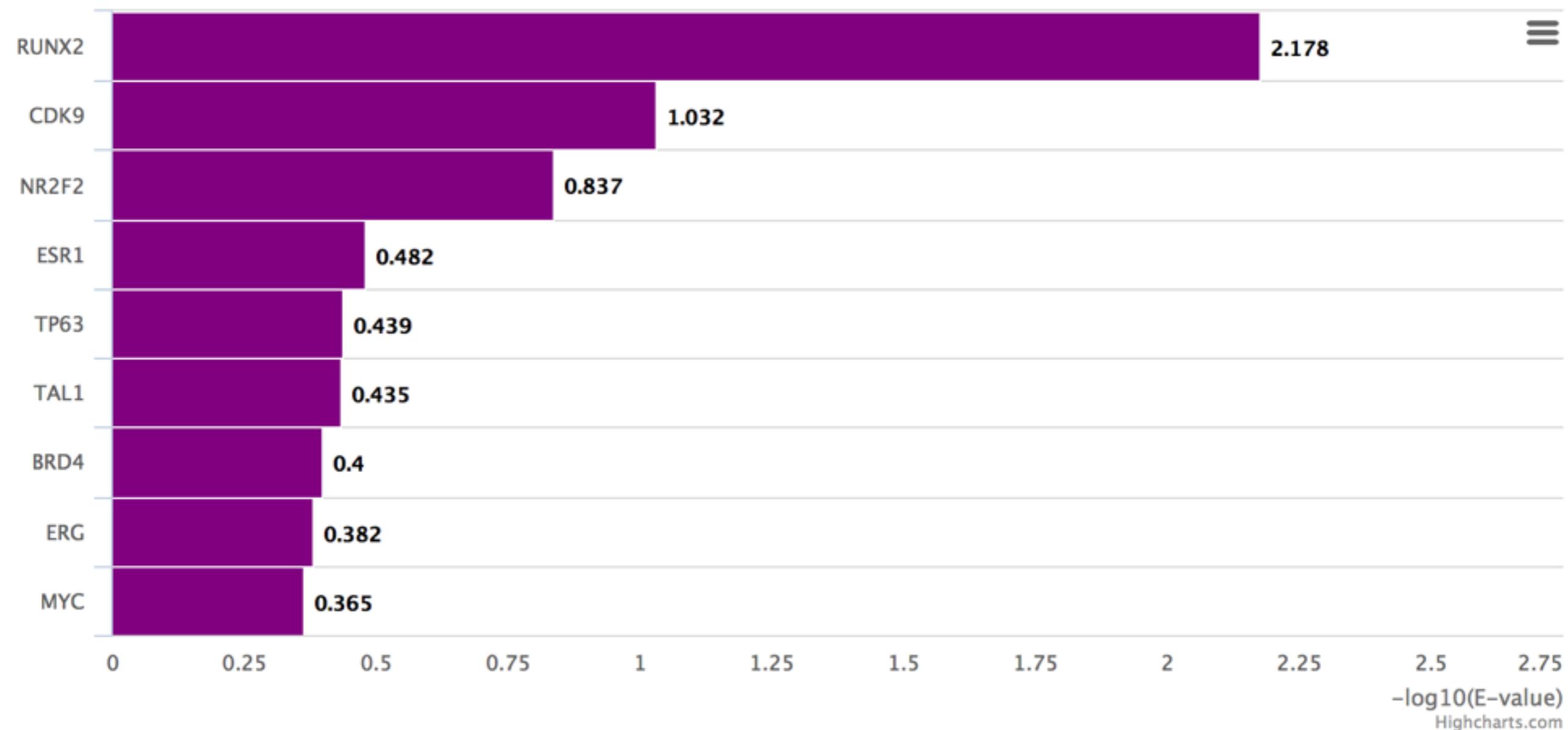
A donut chart showing the percentage of Encode sites that overlap with the Public dataset. The inner ring is grey and the outer ring is blue.

89% Encode sites overlap Public

<http://tagc.univ-mrs.fr/remap/>

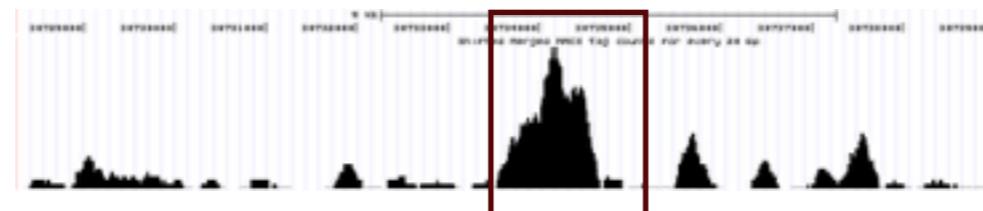
Highcharts.com

Enriched TFs in intersection



Motifs

Details in next session



ChIP-seq peaks

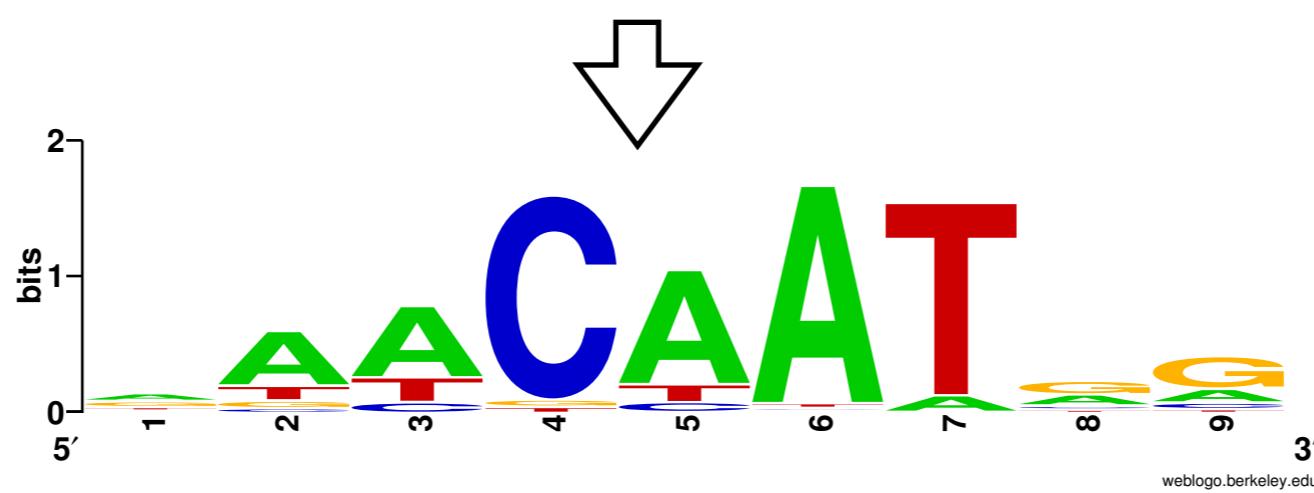
>mm9_chr1_39249116_39251316_+
gagagggaaagggggagaaagaggggaggggagGGTGATAGGTAGCCAGGAG
CCAATGGGGCGTTTCCTTGTCCAGGCCACTTGCTGGAATGTGAGATGT
AGAATGACCAAAGAGAGCTGCCAAGACAGAGCTTGCCCCAGGAATTGA
ACTCAAAGGGTGTCAAGAACAGCAGTGGCCTTGTGCACCTGGCGGGGA
CGTGGCTCCCCTTCCGGCTGGTAGCCAGGtgccctgcctgcctgcct
gccGTGATCTCTGGACGCCAGTAGAGGGTTGTTGGGTTGGGTGAAAC
ACGCCACCCCTGAGCTTCCGGGGCTAGCAATCCCCATCACCCA
TTCGCGCTCAGAACCCCTCAGCGAGTCTAACAGCAGGCCTGGTCCCCG

DNA sequence

↓

A	[24	54	59	0	65	71	4	24	9]
C	[7	6	4	72	4	2	0	6	9]
G	[31	7	0	2	0	1	1	38	55]
T	[14	9	13	2	7	2	71	8	3]

Discovered motif



Motif logo