

ChIP-seq Annotation and Visualization

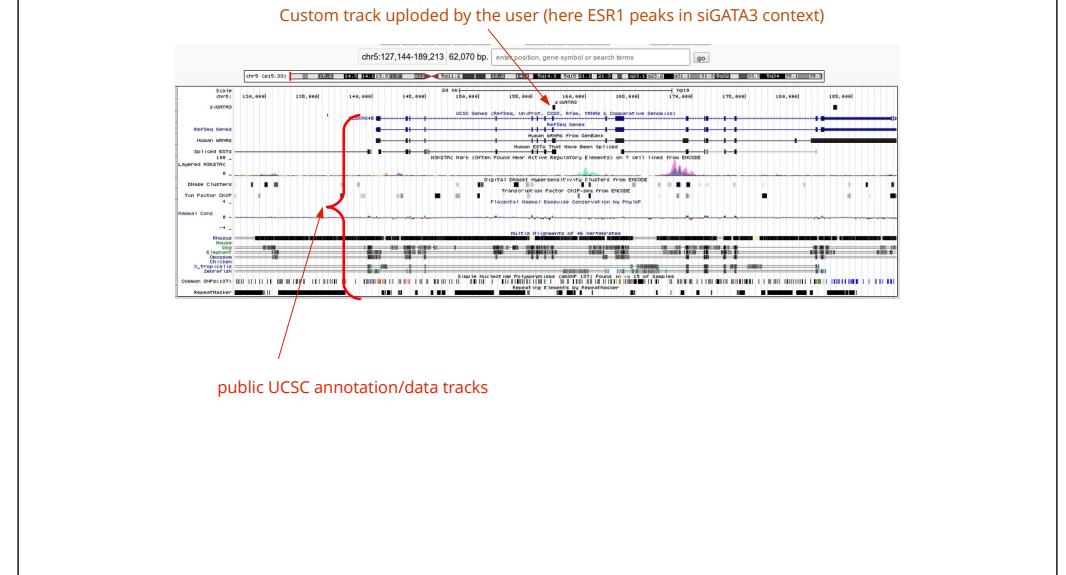
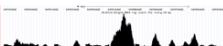
How to add biological meaning to peaks

M. Defrance, M. Thomas-Chollier, C. Herrmann, D. Puthier

Typical questions

- What are the genes associated to the peaks?
- Are some genomic categories over-represented? (promoter?)
- Are some functional categories over-represented?
- Are the peaks close to the TSS?

ChIP-seq peaks



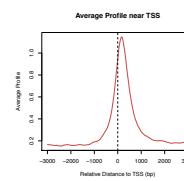
ChIP-seq peaks



Annotation
Visualisation



Enrichment profiles

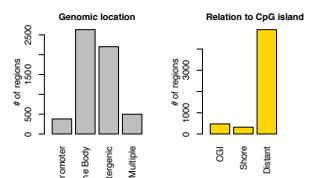


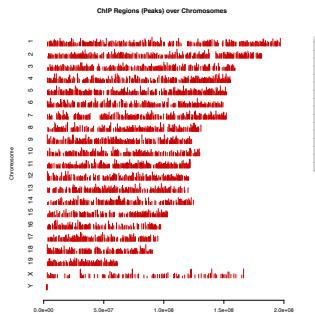
ChIP-seq peaks and their association with genes and genomic features.

Annotated peaks

chr	start	end	length	GCI
chr2	98506153	98507733	1581	Distant
chr9	35112555	35113849	1295	Distant
chr17	39979555	39986015	6461	CGI
chr11	108872720	108873618	899	Distant
chr2	98502147	98504490	2344	Distant
chr4	70038914	70039533	620	Distant
chr6	103598803	103599611	809	Distant
chr5	147072331	147073103	773	Distant
chr9	2999758	3005280	3523	Distant
chr12	3109627	3110366	740	Distant
chr12	75904653	75905731	1079	Distant

Genomic & functional
Annotation



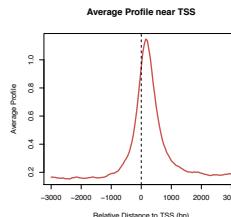


Chromosomal distribution of ChIP peaks

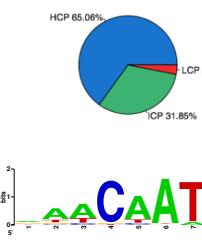
(bed, xls, txt file)

ChIP-seq peaks

(bed, xls, txt file)



Promoter CpG density



Mouse Phenotype

Term name	Brion Rank	Binom Raw	Binom P-Val	FDR	FDR P-Val	Element
abnormal limb digit/club morphology	2	2.059e-91	6.633e-69	2.1467		
abnormal craniofacial morphology	3	3.352e-91	2.033e-67	2.0082		
abnormal limb morphology	5	2.499e-91	3.249e-77	2.3077		
abnormal appendicular skeleton morphology	10	3.625e-70	1.947e-67	2.3450		
abnormal skeleton extremities morphology	12	3.267e-69	1.771e-66	2.3724		
abnormal paw/hand/foot morphology	13	4.030e-66	2.015e-66	2.6813		
abnormal head morphology	14	4.467e-67	3.029e-64	2.0134		
abnormal digit morphology	18	3.654e-63	3.898e-69	2.6982		
abnormal cartilage morphology	23	1.372e-58	2.054e-55	2.1602		
abnormal skeleton development	24	3.370e-56	9.699e-54	2.0833		
abnormal long bone morphology	25	4.693e-56	1.211e-53	2.3374		

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

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

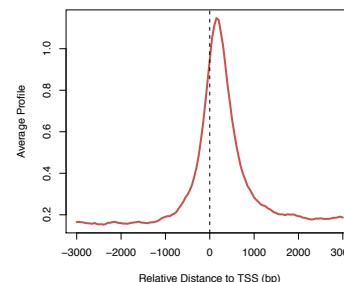
Statistical significance
-10 log(P-value)

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
chr7	48499365	48499865	370	48499615	24	798.58	20.62	0.0
chr18	72429446	72429946	531	72429694	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

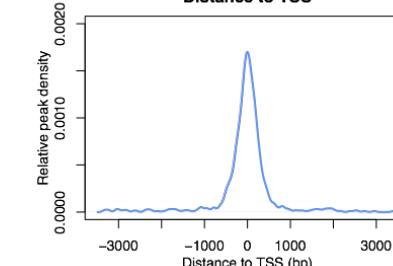
Profile arround the TSS
using profile in wig

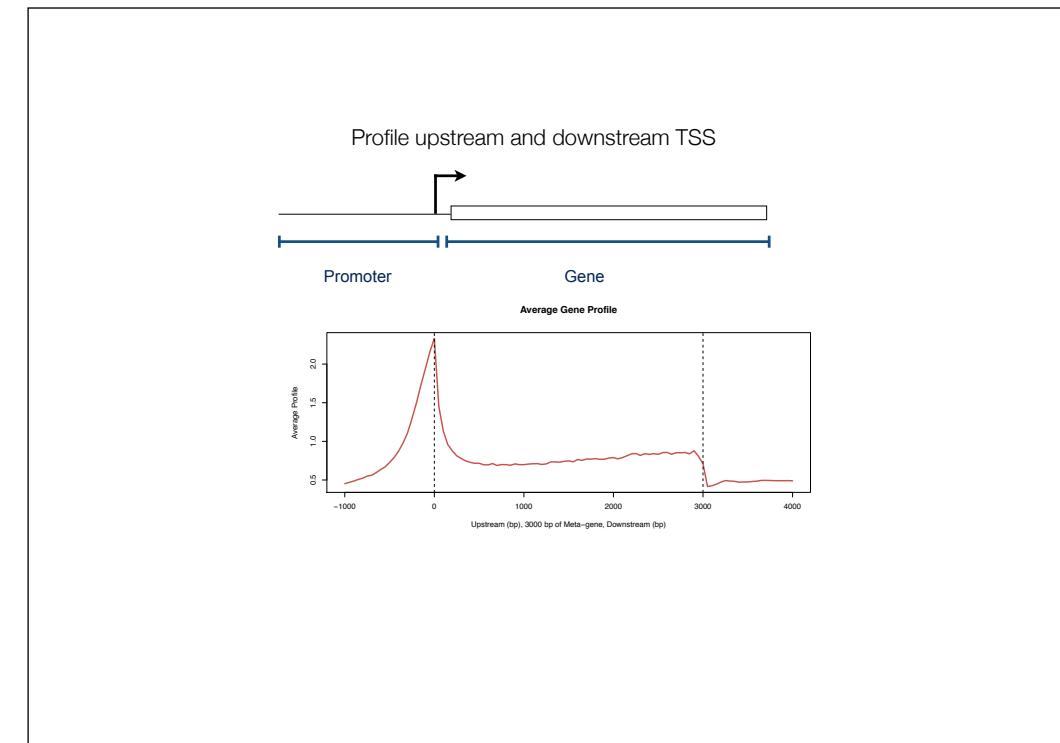
Average Profile near TSS



Peak distance to TSS distribution
using peaks in bed

Distance to TSS

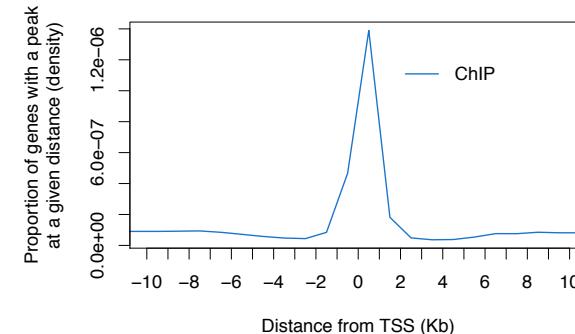




Galaxy: MakeTSSdist

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

PAVIS

PAVIS: a tool for Peak Annotation and Visualization
Wei-chun Huang^{1,*}, Rasiah Loganathan^{2,†,‡}, Bryce Schroeder^{1,‡,§}, David Fargo² and Leping Li^{1,*}

Annotation and visualization

Species/Genome Assembly/Gene Set: Human Ensembl_GRCm38/hg38

Upstream Length: 5000

Downstream Length: 1000

The query peak file to be annotated: no file selected

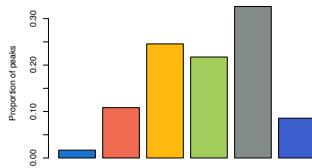
File format: UCSC BED GFF3 EpiCenter Report Other text file
If other, please specify the delimiter and column numbers:
field delimiter: tab whitespace comma semicolon pipe
column number: chromosome:1 , start position:2 , end position:3

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:1 , start position:2 , end position:3

Search distance to query peaks: 2000

SUBMIT RESET



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

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

PAVIS

Example output

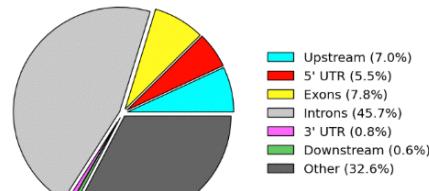
Peak Location Annotation (Annotation File 119.01 kB)

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

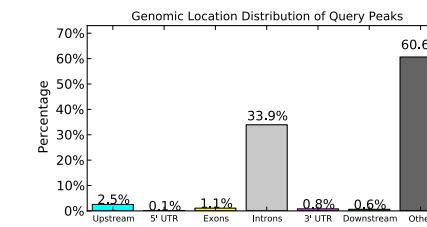


[Click here to download the Pie-Chart in the PDF format](#)

PAVIS

Detailed view

Chromosome	Loci	StartLoci	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	StartLoci	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	

GREAT

Functional annotation of cis-regulatory regions

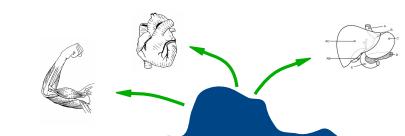
GREAT improves functional interpretation of cis-regulatory regions

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

ChIP-seq peaks

Ontology terms

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



GREAT

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:

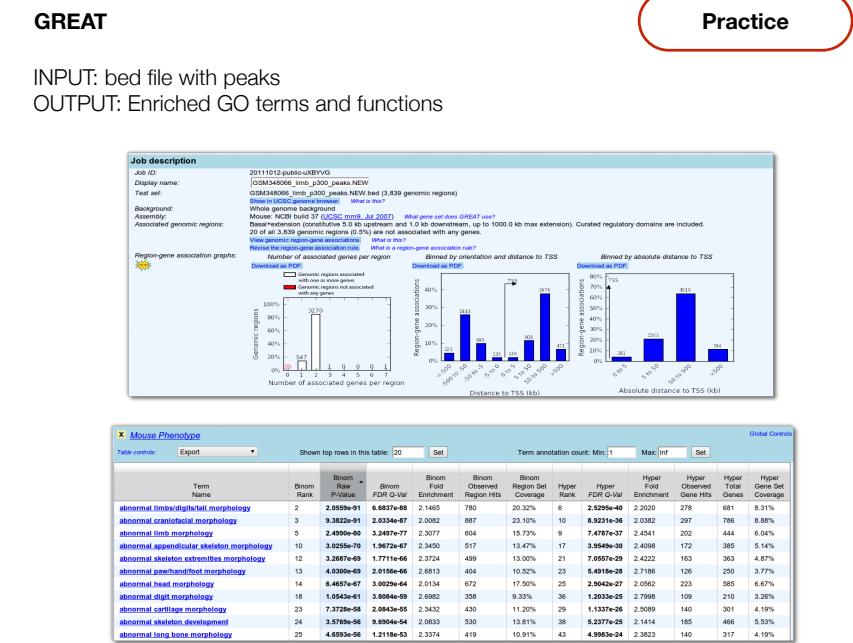
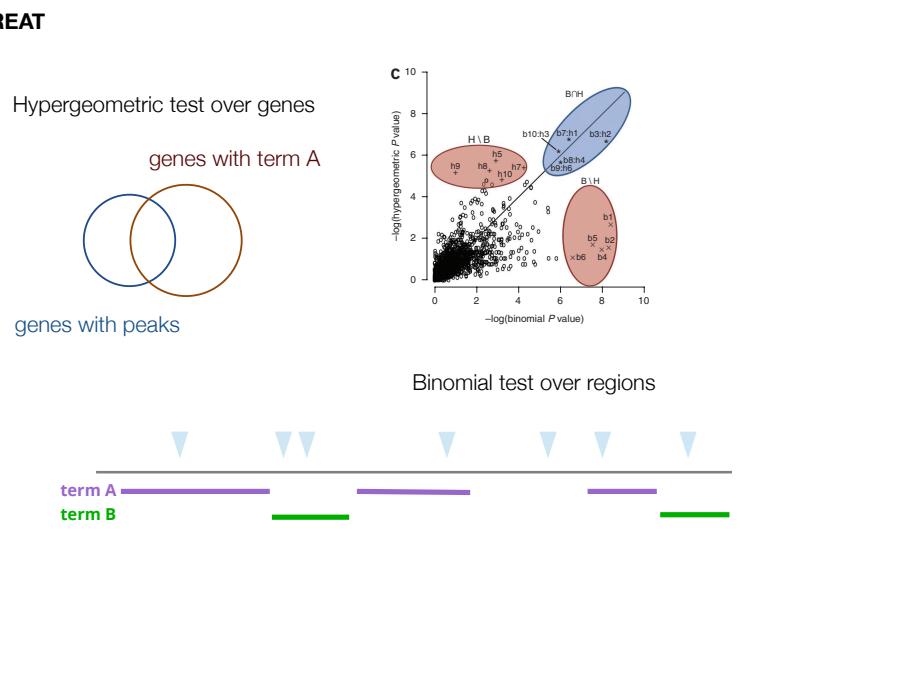
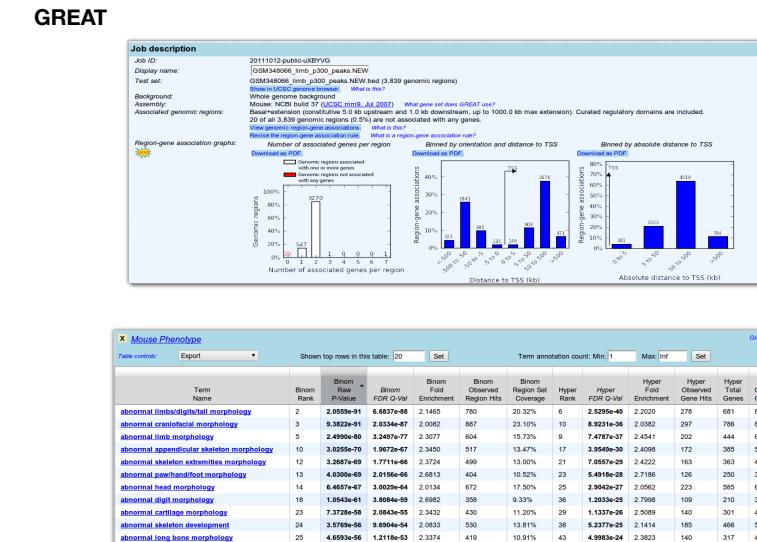
[When should I use a background set?](#)
[What should my background regions file contain?](#)

Association rule settings

[Show settings »](#)

Submit **Reset**

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

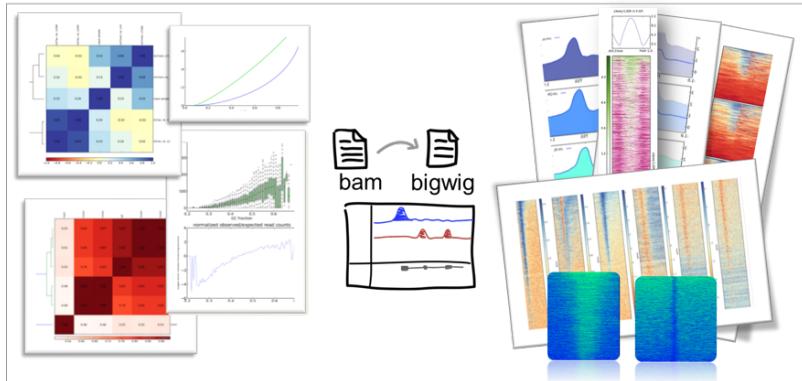


deepTools

deepTools: a flexible platform for exploring deep-sequencing data

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

QUALITY CHECKS – FORMAT CONVERSION & NORMALIZATION – PLOTTING

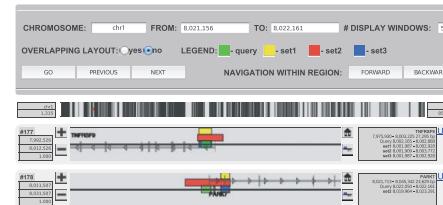
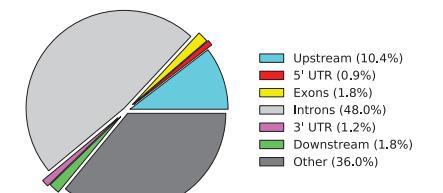


ChIPseequer

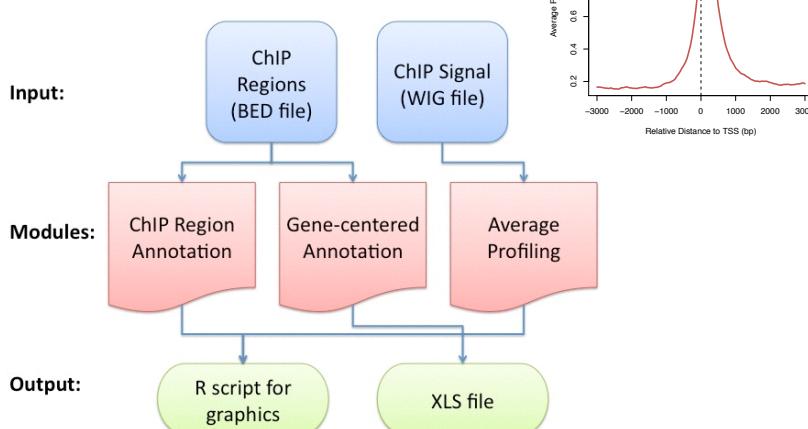
An integrated ChIP-seq analysis platform with customizable workflows

Eugenia G Giannopoulou^{1,2} and Olivier Elemento^{1,2*}

A comprehensive framework for the analysis of ChIP-seq data



CEAS (Cis-regulatory Element Annotation System)

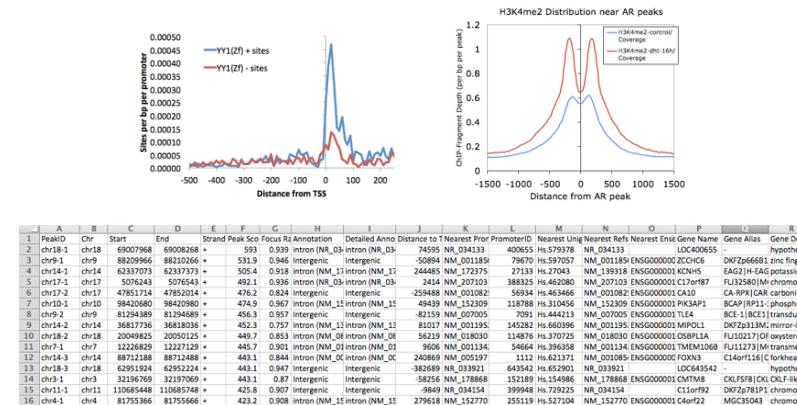


HOMER

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

Sven Heinz,^{1,2} Christopher Benner,^{1,2} Nathaniel Spann,^{1,2} Eric Bertolini,⁴ Yin C. Lin,² Peter Lasko,² Jason X. Cheng,⁴ Cornelis Mure,² Harinder Singh,^{4,5} and Christopher K. Glass^{1,2*}

Software for motif discovery and next-gen sequencing analysis



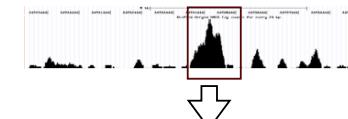
<http://homer.salk.edu/homer/>

HOMER: annotate peaks

- 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

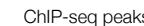
Motifs

Details in next session



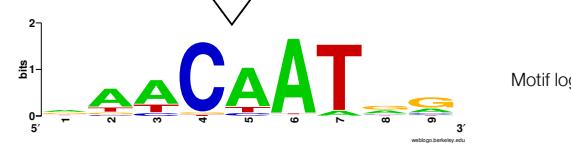
```
>mm9_chrl_39249116_39251316_+
gagaaaggaaaggaaaggaggggggggGGTGTAGTGACCGAGG
CCAACTGGGGCGCTTCTTCGGCTCAGGCCACTTGTGGAATCTGAGAT
ACTAACGACCCAAAAGACAGGCAACGAGCTCTGCCGGCAAAAGT
ACTCAAAGGCTGTGAAACAGCTGGCTTGTGCAAGCTGCCGGGG
CTGGCTCCCTCTGGCTCAGTGGGGGGGGGGGGGGGGGGGGGGGG
ggcGGTATCTGGACGCGTACAGGGGGGGGGGGGGGGGGGGGGGG
AGCCGGACCCCCGGACTCTGGCGGGGGGGGGGGGGGGGGGGGGGG
TTGGCGCTTCTGGACCCCCGGACTCTGGCGGGGGGGGGGGGGGGGG
```

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

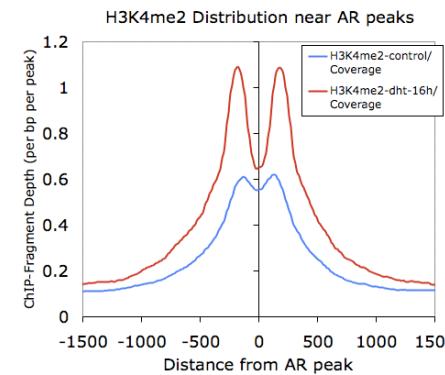


DNA sequences

Discovered motifs



HOMER: compare peaks



Peak Co-Occurrence Statistics

Co-Bound Peaks

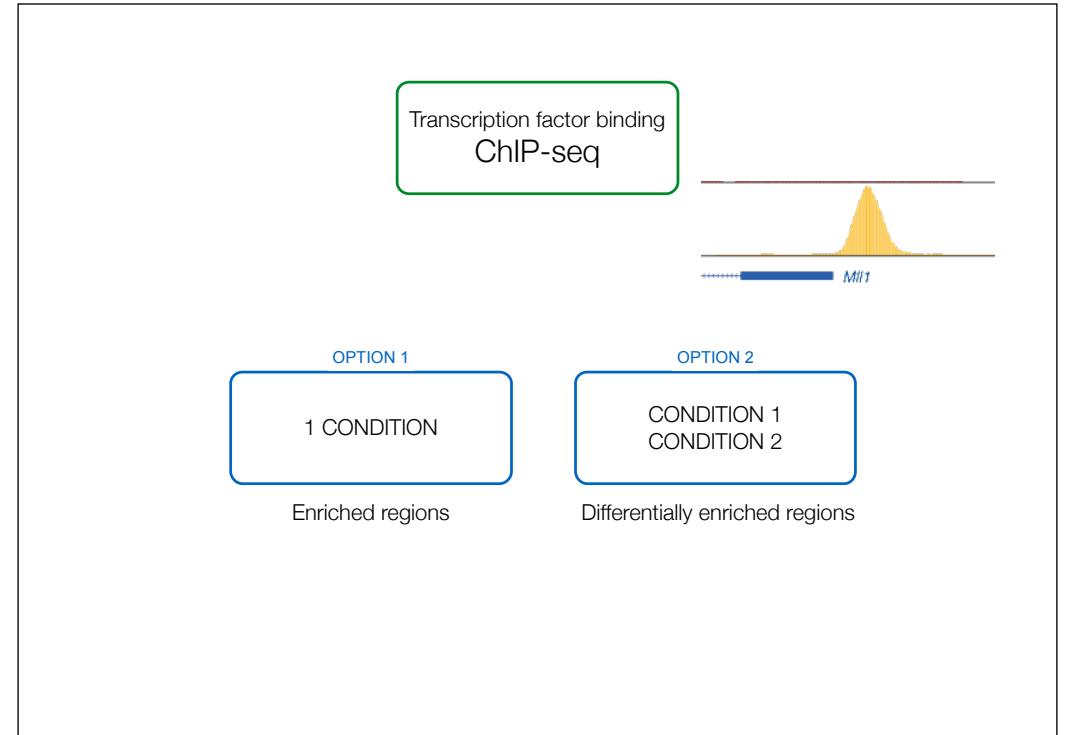
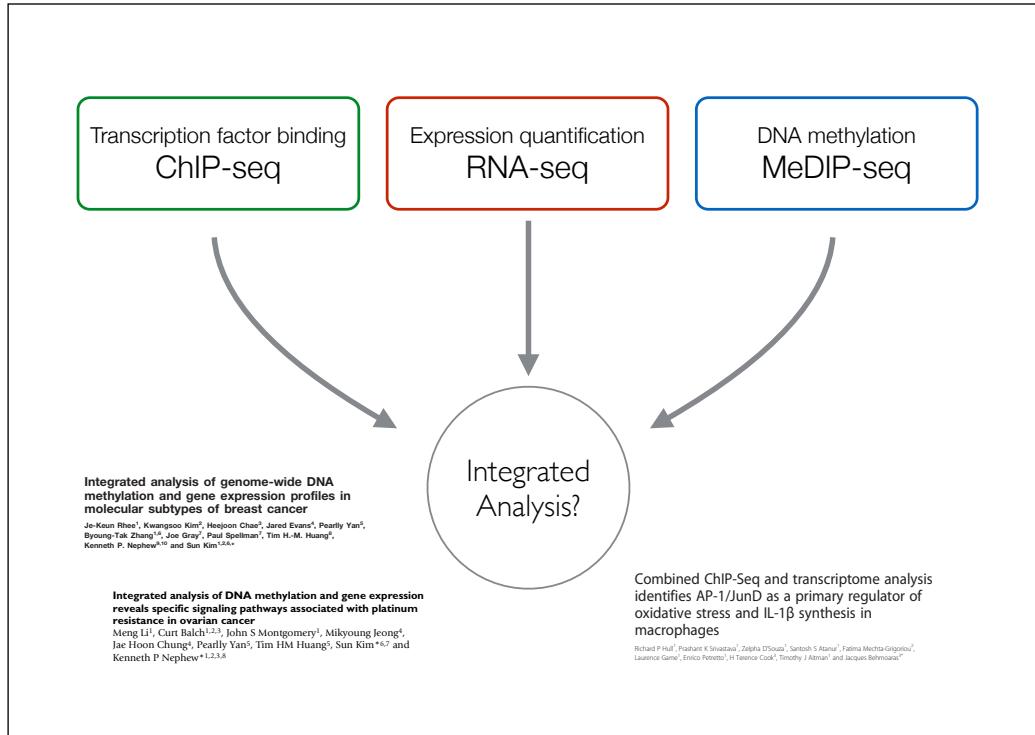
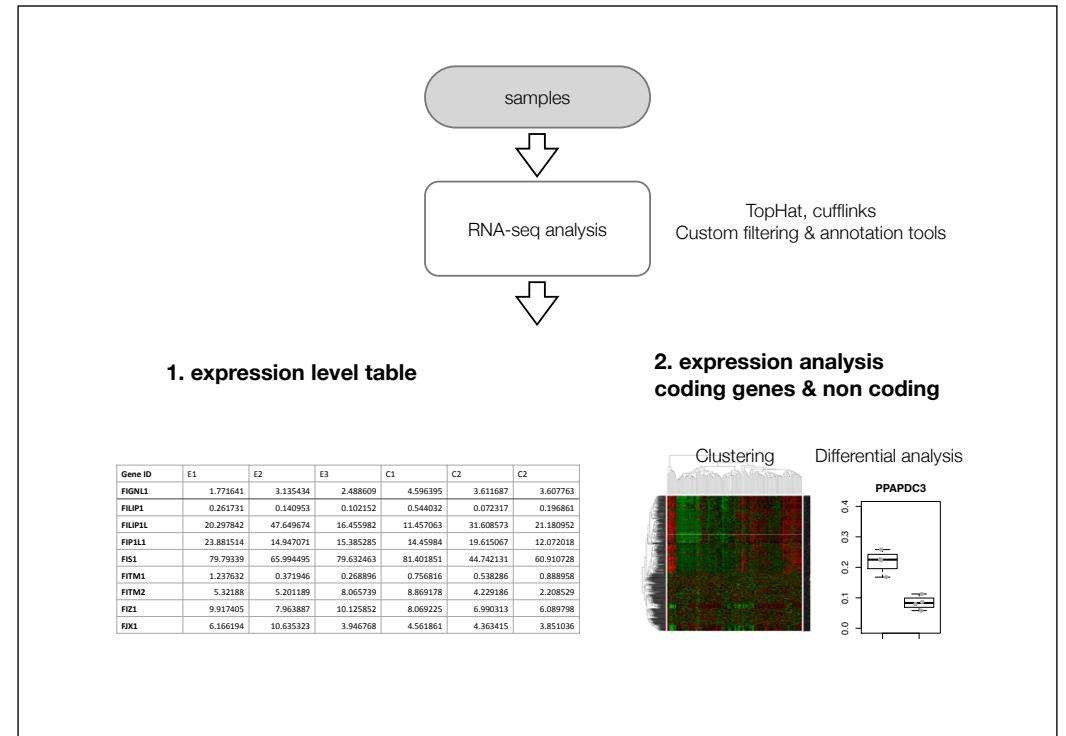
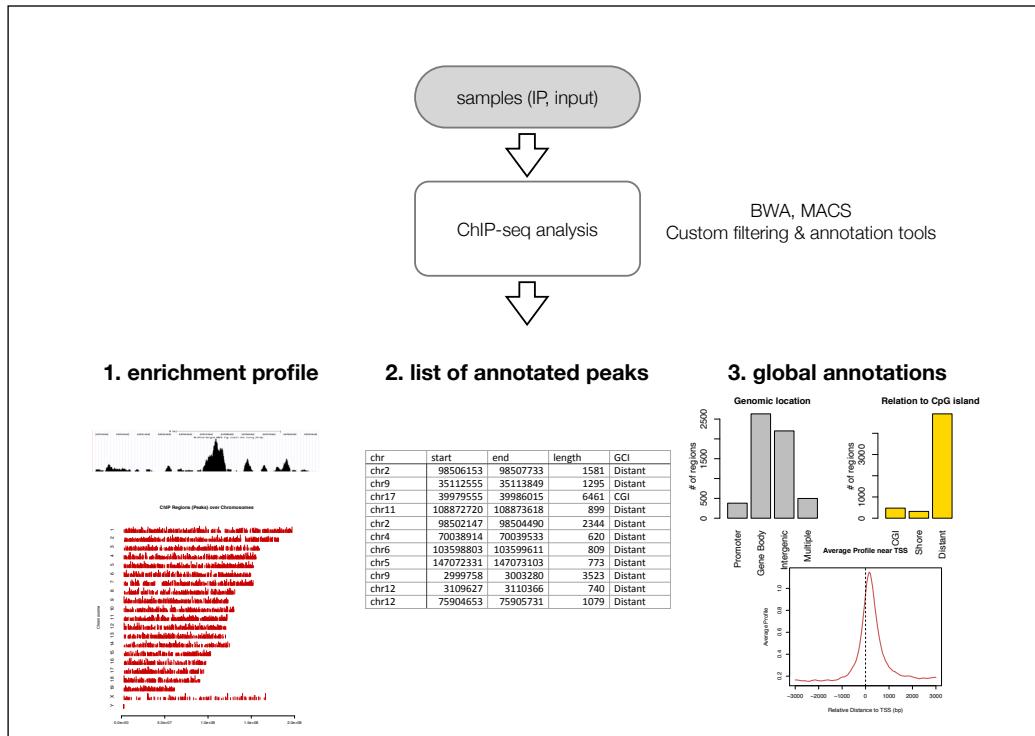
Differentially Bound Peaks

Integrated analysis of sequencing data

How to combine *-seq data

M. Defrance, M. Thomas-Chollier, C. Herrmann, D. Puthier

*ChIP-seq, RNA-seq, MeDIP-seq, ...



μ-array
Commercial / custom
Coding, non coding

Expression quantification
RNA-seq

RNA-seq
Coding, non coding
Alternative transcripts

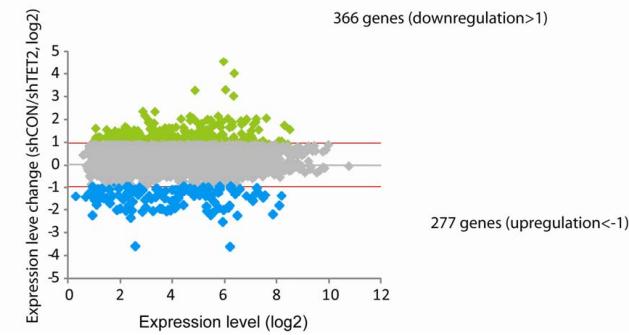
CONDITION 1
CONDITION 2

Differentially expressed genes

Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks

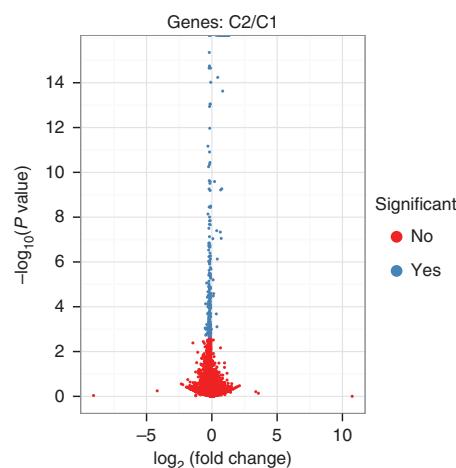
Cole Trapnell^{1,2}, Adam Robert¹, Loyal Goff^{1,3,4}, Geo Pertea^{1,4}, Dachwan Kim^{1,2}, David R Kelley^{1,2}, Harold Pimentel¹, Steven I Salzberg^{1,4}, John L Rinn^{1,2} & Lior Pachter^{3,4,5}

Differentially expressed genes?



2 condition x 1 sample

Differentially expressed genes?



Filter on both Pvalue and Fold change!

DNA methylation-seq

Bisulphite sequencing
DNA treatment with bisulphite specifically introduces mutations at unmethylated Cs. These mutations are mapped by next-generation sequencing

OPTION 1

1 CONDITION

Methylation level
(CpG, region)

OPTION 2

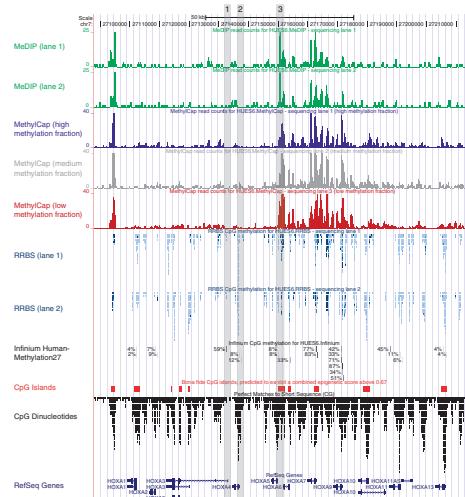
CONDITION 1
CONDITION 2

Differentially methylated regions

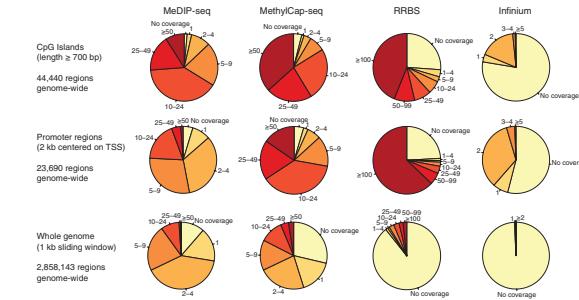
Enrichment-based methods
Methylated (alternatively, unmethylated) DNA fragments are enriched in a DNA library. The library composition is quantified by next-generation sequencing

DNA methylation mapping technologies

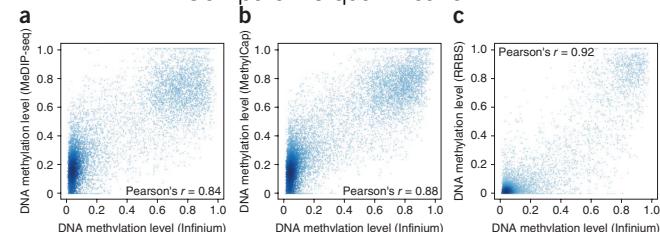
MethylCap-seq
MeDIP-seq
RRBS
WG bisulfite-seq
 μ -array (infinium)



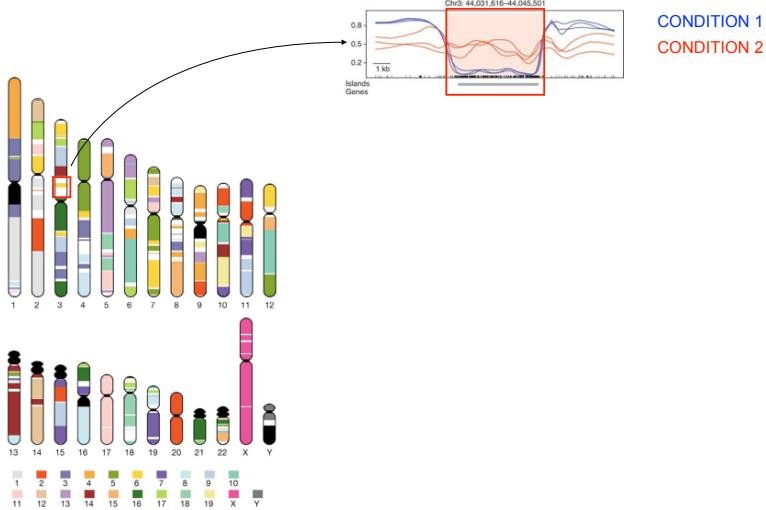
DNA methylation mapping technologies: Genomic coverage



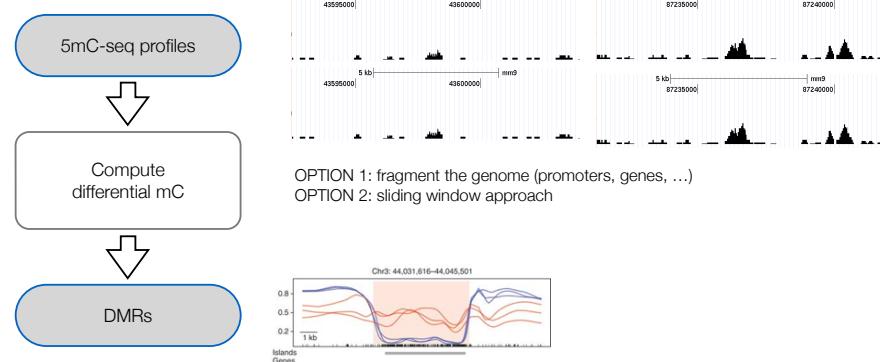
Comparative quantification



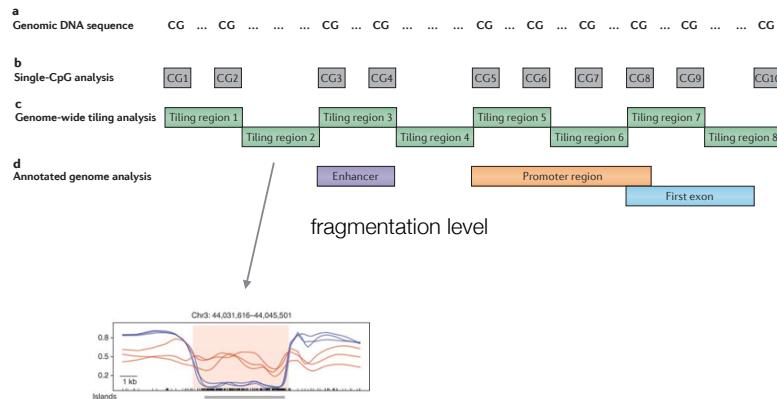
Differentially Methylated Regions (DMRs)?



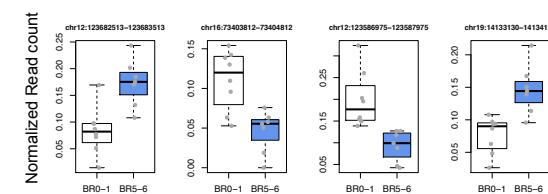
Identifying differentially methylated regions (DMRs)



Differentially methylated region, CpG or promoter?



Differentially methylated regions

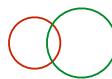


Region	Log ratio	Fold change	Pvalue	RefSeq.ID
chr12:123682513-123683513	-1.09	2.14	1.78E-03	
chr16:73403812-73404812	1.11	2.16	1.98E-03	
chr12:123586975-123587975	0.84	1.79	2.14E-03	NM_006312
chr19:14133130-14134130	-0.68	1.60	2.31E-03	NM_014921
chr5:134526390-134527390	0.88	1.84	3.21E-03	
chr14:66964536-66965536	0.83	1.78	3.47E-03	
chr4:17081181-17082181	0.70	1.63	3.49E-03	
chr6:21353381-21354381	1.16	2.23	3.53E-03	NM_013401
chr3:14347979-14348979	0.95	1.93	3.65E-03	NM_00114382
chr11:61421968-61422968	-0.93	1.91	3.85E-03	NM_032427
chr2:12995930-12996930	0.98	1.98	4.05E-03	NM_032251
chr11:95557089-95558069	0.95	1.93	4.83E-03	NM_007368
chr11:63865332-63866332	-1.12	2.17	4.90E-03	
chr13:113901339-113902339	-0.82	1.77	4.92E-03	NM_001077183
chr17:77586207-77587207	-1.22	2.34	5.43E-03	NM_004252
chr16:2059877-2060877	0.95	1.93	5.58E-03	NM_139057

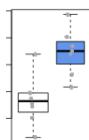
Combining RNA-seq and ChIP-seq



OPTION 1: focus on genes, on/off state



OPTION 2: quantitative analysis / correlation



OPTION 1: focus on genes, on/off state

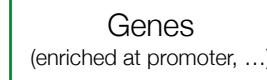


PAVIS: a tool for Peak Annotation and Visualization

Weichun Huang^{1,5}, Rasiah Loganathanraj^{2,3,6}, Bryce Schroeder^{1,4,5}, David Fargo² and Leping Li^{1,*}

ChiPeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data

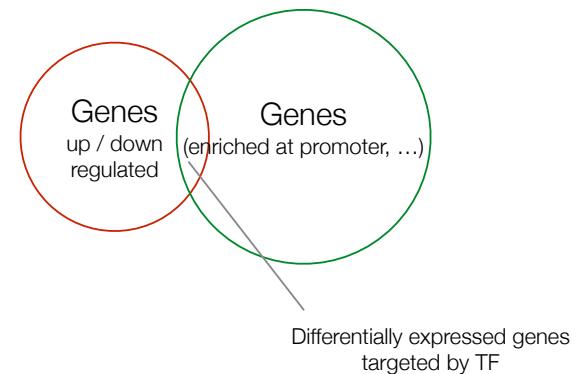
Lihua J Zhu^{1,2}, Claude Gazin³, Nathan D Lawson^{1,2}, Hervé Pagès⁴, Simon M Lin⁵, David S Lapointe⁶, Michael R Green^{1,2}



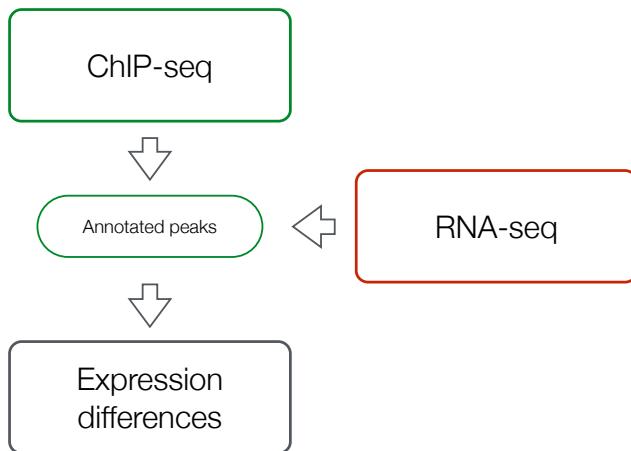
OPTION 1: focus on genes, on/off state



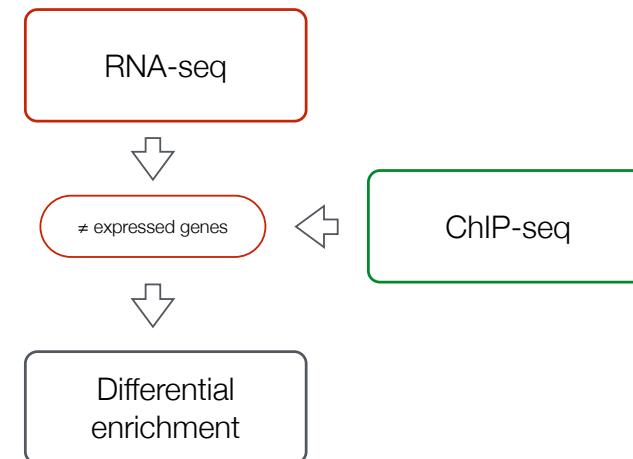
OPTION 1: focus on genes, on/off state



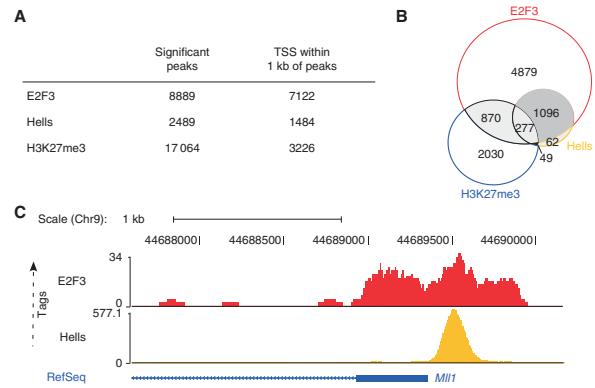
OPTION 2: quantitative analysis



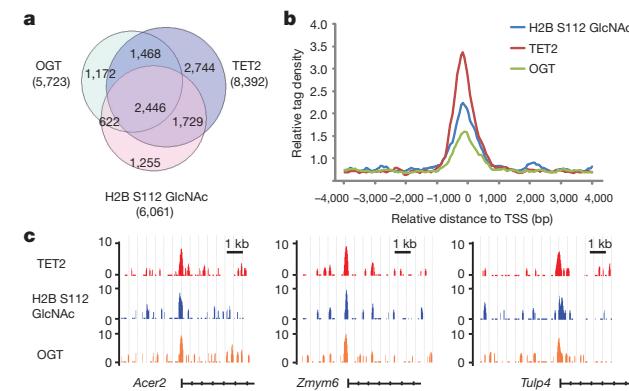
OPTION 2: quantitative analysis



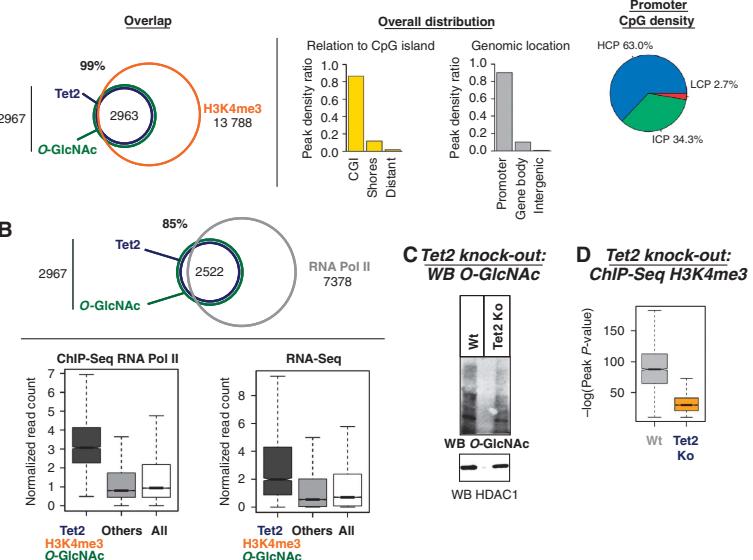
Histone marks and TF binding



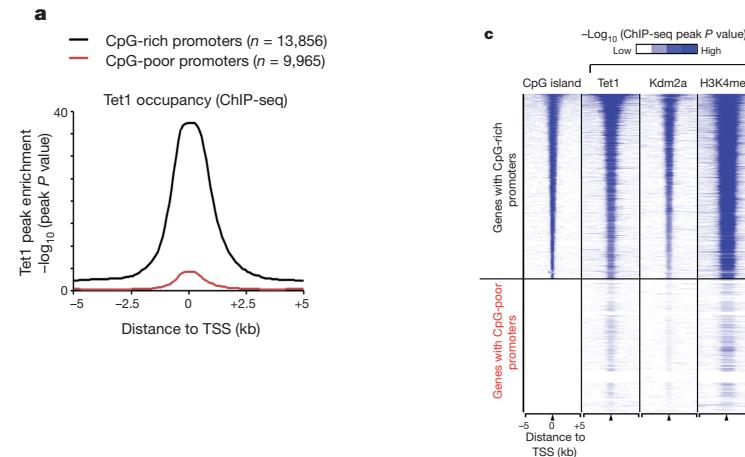
Co-localization



RNA-seq ChIP-seq



Histone marks and TF binding



Combining Histone marks and TF binding and expression

