Downstream analysis of ChIP-seq data

Shamith Samarajiwa

CRUK Bioinformatics Summer School July 2019





Summary

Downstream analysis for extracting meaningful biology :

- Normalization and Visualization
- Annotation of genomic features to peaks
- Feature distribution of binding sites
- Feature overlap analysis
- Functional enrichment analysis: Ontologies, Gene Sets, Pathways
- Motif identification and Motif Enrichment Analysis
- Integration with transcriptomic data to Identify direct targets
- Network Biology applications
- Differential binding analysis



Compare, Normalize & Visualize 1

- seqMiner enables qualitative comparisons between a reference set of genomic positions and multiple ChIP-seq data-sets.
- Useful for comparing and visualizing replicates or conditions.





Ye et al., 2011, Nucleic Acids Res.



Compare, Normalize & Visualize 2

- deepTools2 sequence depth or input normalization, GC bias correction
- Plot signal profiles
- Customized heat-maps
- PCA, correlation and fingerprint plots



Ramírez et al., 2016, Nucleic Acids Res.



Peak annotation 1

- ChIPpeakAnno (BioC) map peaks to nearest feature (TSS, gene, exon, miRNA or custom features)
 - extract peak sequences
 - find peaks with bidirectional promoters
 - obtain enriched gene ontology
 - map different annotation and gene identifiers to peaks
- Use **biomaRt** package to get annotation from Ensembl.
- IRanges, GenomicFeatures, GO.db, BSgenomes, multtest (BioC)
- converts BED and GFF data formats to *RangedData* object before calling *peak annotate* function.

Zhu et al., 2010, BMC Bioinformatics

Peak annotation 2

PeakAnalyzer

- A set of high-performance utilities for the automated processing of experimentally-derived peak regions and annotation of genomic loci.
- Consists of PeakSplitter and PeakAnnotator.
- Biologist' friendly tool.
- Get latest genome annotation files from Ensembl (gtf format) or UCSC (BED format).
- Map to either nearest downstream gene, TSS or user defined annotation.
- Determine overlap between peak sets.
- Split peaks to sub-peaks. May be useful for *de novo* motif analysis.

Salmon-Divon et al., 2010, BMC Bioinformatics.

Peaks distribution across features



Yu et al., 2015, Bioinformatics

Functional Enrichment Analysis 1

GREAT & rGREAT: Genomic Regions Enrichment of Annotations Tool



$$P = \Pr_{\text{hyper}} (k \ge 1 \mid N = 8, K = 3, n = 2)$$

 $P = Pr_{binom}$ ($k \ge 5 \mid n = 6, p = 0.6$)

Functional Enrichment Analysis 2

chipenrich

- Includes 3 different enrichment methods:
 - Broadenrich broadpeaks or histone modifications
 - Chipenrich -TF narrow peaks 1000-10000's
 - Polyenrich -TF >100,000
- Includes annotation, and can use custom user provided annotation



Welch et al., 2014, Nuc. Acids Res.

Motif detection

- Don't scan a sequence with a motif and expect all sites identified to be biologically active. Random matches will swamp the biologically relevant matches! This is a well known problem in motif searching, amusingly called the "Futility Theorem" of motif finding. *Wasserman & Sandelin, 2004, Nat Rev Genet.*
- 1. PWM based sequence scanning or word search methods. These methods uses prior information about TF binding sites and therefore can only be used to detect known Transcription Factor Binding Sites (TFBS).
- 2. *De novo* motif identification Pattern discovery methods:
- Word based Occurrence of each 'word' of nucleotides of a certain length is counted and compared to a background distribution.
- **Probabilistic** seek the most overrepresented pattern using algorithmic approaches like Gibbs sampling and Expectation maximization. These iteratively evolve an initial random pattern until a more specific one is found.
- Use *de novo* motif calling and alignment to build your own PWMs!
- **Biostrings & Motiv** packages have PFM to PWM conversion methods.

BioConductor motif analysis packages

- rGADEM -motif discovery
- MotifRG -motif discovery
- MotIV -map motif to known TFBS, visualize logos
- motifStack -plot sequence logos
- MotifDb -motif database
- PWMenrich -motif enrichment analysis
- TFBSTools R interface to the JASPAR database

Position Weight Matrices



Consensus sequence

PWM conversion:

c Position frequency matrix (PFM)

- 1	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A	0	4	4	0	3	7	4	3	5	4	2	0	0	4
C	3	0	4	8	0	0	0	3	0	0	0	0	2	4
G	2	3	0	0	0	0	0	0	1	0	6	8	5	0
T	3	1	0	0	5	1	4	2	2	4	0	0	1	0

 $W_{b,i} = \log_2 \frac{p(b,i)}{p(b)}$

d Position weight matrix (PWM)

A	-1.93	0.79	0.79	-1.93	0.45	1.50	0.79	0.45	1.07	0.79	0.00	-1.93	-1.93	0.79
C	0.45	-1.93	0.79	1.68	-1.93	-1.93	-1.93	0.45	-1.93	-1.93	-1.93	-1.93	0.00	0.79
G	0.00	0.45	-1.93	-1.93	-1.93	-1.93	-1.93	-1.93	0.66	-1.93	1.30	1.68	1.07	-1.93
т	0.15	0.66	-1.93	-1.93	1.07	0.66	0.79	0.00	0.00	0.79	-1.93	-1.93	-0.66	-1.93
	-													

Site scoring



Nature Reviews | Genetics

TFBS PWM/PFM sources

TRANSFAC public	Matys et al., 2006	Multiple species	v7.0 2005, Not been updated for a while!
TRANSFAC professional	Matys et al., 2006	Multiple species	v2017
JASPAR 2014	Mathelier et al., 2014	Multiple species	(656)
ORegAnno		Multiple species	Curated collection from different sources.
hPDI	Xie et al., 2010	Human	(437)
SwissRegulon	Pachkov et al., 2010	mammalian	(190)
HOMER	Heinz et al., 2010	Human	(1865)
UniPROBE	Newburger & Bulyk, 2009	Multiple species	
Dimers	Jonawski et al., 2013	Human	(603) predicted dimers
FactorBook	Wang et al., 2012	Human	(79) ENCODE ChIP-seq motifs
SCPD, YetFasco		Yeast	
Elemento, Redfly FlyFactorSurvey,Tiffin		Drosophila	
Prodoric		Prokaryotic	

Motif Enrichment Analysis

- Identifies over and under-represented known motifs in a set of regions
- The TFs whose DNA binding motifs are enriched in a set of regulatory regions are candidate transcription regulators of that gene/promoter/enhancer set.
- Without ChIP-seq, identifying a co-regulated gene sets is difficult. Use Ontologies, pathways, GSEA etc.
- Picking the right background model will determine the success of the motif enrichment analysis:
 - All core-promoters from protein coding or non-coding genes etc.
 - Higher order Markov model based backgrounds
 - A sequence set similar in nucleotide composition, length and number to the test set
 - Open chromatin regions or a shuffled test sequence set.

Motif detection and enrichment analysis

- MEME Suite and MEME-Chip http://meme.nbcr.net
- Given a set of genomic regions, it performs
 - Motif detection (FIMO)
 - *ab initio* motif discovery -novel TF binding sites (MEME, DREME)
 - motif enrichment analysis -known TF enrichment (Centrimo, AME)
 - motif visualization (MAST and AMA)
 - binding affinity analysis
 - motif identification -compare to known motifs (TOMTOM)
- MEME -expectation maximization (EM) to discover probabilistic models of DNA-binding by single TFs or TF complexes.
- DREME -simpler, non-probabilistic model (regular expressions) to describe the short binding motifs.

Machanick and Bailey, 2011, Bioinformatics



Motif detection

- HOMER v4 http://homer.salk.edu/homer/index.html
- Large number of (Perl and C++) tools for ChIP-seq analy
- Provides both *de novo* and PWM scanning based motif identification and enrichment analysis.
- User can specify custom background. (Randomly selected, GC or CGI matched backgrounds.)
- Uses a collection of ChIP-seq derived PWMs or user can specify PWM.
- Can help with Peak annotation, GO enrichment analysis, Extract peak sequences, Visualization.



Motif Enrichment Analysis

Pscan-Chip

- Motif enrichment analysis using PWM databases and user defined background models.
- Optimized for ChIP-seq.
- Ranked lists of enriched motifs.
- Sequence logo's and motif enrichment distribution plots.

1a. Insert list	t of genomic regions (BED	format): (help)	Download txt file								
			Name	ID	L.PV	L.0/U	G.PV +	G.0/U	SP.COR	P.POS	P.POS.PV
hr1 143	3913371 1439133	71 🕕	NEYA	MA0060.1	0	+	0	A	0.3649	[13,23]	2.3E-9
hr3 516	584882 67584882 890434 51890434		Gfi	MA0038.1	5.9E-238	+	0	4	0.2663	[-20,-10]	0.1432
hr6 110	0201988 11020198	38		100000.1	5.50 250				0.0000	(10, 10)	
hr8 888	833589 8883358	X X	Kite	MA0039.2	1.20-38	1	0		0.0253	[40,50]	1.0000
hr1 200	6138286 20613828	36 74	NFIC	MA0161.1	1.6E-120	+	4.8E-264	t.	0.0063	[8,18]	1.SE-5
			<u>SP1</u>	MA0079.2	0.0004	- +	2.3E-203	t.	-0.0601	[-70,-60]	1.0000
1b. Or BED file	le	(เวียกตุ)	P8X1	MA0070.1	2.7E-146	+	2.5E-199	4-	0.3342	[-29,-19]	1.0000
			Zfx	MA0146.1	0.0002	+	3.1E-148	4	-0.1132	[37,47]	1.0000
2. Select:				MA0024 1	1.0000		6 36 139		-0.0205	[40 60]	1.0000
			<u>EXT1</u>	MA0024.1	1.0000		0.32-138		-0.0395	[44,24]	1.0000
Organism:	Homo	sapiens 🔹	Mafb	MA0117.1	6.0E-17	Ţ.	1.3E-130	1 T	0.1903	[-16,-6]	0.0040
			TFAP2A	MA0003.1	1.0000		1.4E-115	+	-0.1697	[34,44]	1.0000
Assembly:	hg19	•	Egr1	MA0162.1	0.2969	+	4.7E-102	1	-0.0686	[-70,-60]	1.0000
			PLAG1	MA0163.1	1.0000		7.0E-96	4	-0.088	[37,47]	1.0000
Background:	K562	•	HIELAULAPAT	MA0359.1	7 35.11		1 95.05	4	-0.0447	100.001	1 0000
				100239.1	7.20-11		1.00-95		-0.0447	[39,49]	1.0000
Descriptors:	Jaspar	•	Mycn	MA0104.2	3.88-11	<u> </u>	1.62-86		-0.1071	[+25,-15]	1.0000
A statistic stati			<u>Sox17</u>	MA0078.1	4.3E-158	1	7.7E-66	1	0.0464	[-6,4]	0.0004
Additional descriptors		(เวียกตู)	Myc	MA0147.1	1.7E-5	+	1.1E-48	1	+0.1079	[-25,-15]	1.0000
			TLX1::NFIC	MA0119.1	2.6E-34	+	6.6E-42	+	0.0364	[12,22]	1.0000
Run!) (R	Reset! Messages:		Tcfcp211	MA0145 1	0.0003		2.4E-41	4	0.0175	[17,27]	1.0000
			Mahar	MA0125 :	0.95-112		1.15-20	A V	0.2112	[16 A]	0.0270
-	7647 regions a	cquired		PIA0125.1	9.00-113		1.16-39		0.2112	[-0,4]	0.0279
5	Running Pscan Please wait	done	MIZE	MA0131.1	1.0000		2.0E-39		-0.0551	[55,65]	1.0000
(9)			RREB1	MA0073.1	1.0000		2.1E-25	t	0.0224	[50,60]	1.0000
	A		Arnt::Ahr	MA0006.1	1.0000		3.1E-25	.	-0.0922	[43,53]	1.0000
22.1			MEKBI	MA0105 1	0 1278		9.05-24		-0.1607	TAE EET	1.0000
	Name	NFY	A 34	16 7 58 51	0 2	112	16 0 1				
				10 7 50 51				-			
	Class	Other Alpha-Hel	lix C 37	33 51 14 4	116 113	0	0 1 6	5		1	
	Class Species NCBI ID	Other Alpha-Hel Mar	lix C 37 ny G 27	33 51 14 4 26 25 41 56	116 113 0 1	0	0 1 6	* 5 3			
	Class Species NCBI ID Inf. Content	Other Alpha-Hel Mar 12.9	Iix C 37 ny G 27 7 18	33 51 14 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3	0 1 6 0 0 3 0 115 4	* 5 3		1	
	Class Species NCBI ID Inf. Content SuperGroup	Other Alpha-Hel Mar 12.9 vertebrate	lix C 37 ny G 27 7 18 es	33 51 14 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3	0 1 6 0 0 3 0 115 4	* 5 3			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc.	Other Alpha-Hel Mar 12.5 vertebrate P2351	lix C 37 hy G 27 7 18 es 11	10 7 50 51 33 51 14 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3	0 1 6 0 3 0 115 4	*			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type	Other Alpha-Hel Mar 12.5 vertebrate P2351 COMPILE	Iix C 37 ny G 27 7 T 18 es 11 10	10 1 00 01 33 51 14 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3	0 1 6 0 0 3 0 115 4	* 5 3 *			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID	Other Alpha-Hel Mar 12.5 vertebrate P235 COMPILE 946981	IIx C 37 Iny G 27 33 T 18 es 11 1D 18	10 1 1 4 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3	0 1 6 0 0 3 0 115 4	*			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best	Other Alpha-Hel Mar 12:5 vertebrate P235 COMPILE 946981	Iix C 37 ny G 27 7 18 es 11 1D 18	10 7 50 51 33 51 14 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3	0 1 6 0 0 3 0 115 4	* 5 3 *			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hel Mar 12.5 vertebrate P235 COMPILE 946981 Com	Iix C 37 ny G 27 33 T 18 11 ID 18 18 I I	10 7 30 31 33 51 14 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3	0 1 6 0 0 3 0 115 4	* 5 3 *			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hel Mar 12.5 vertebrate P235 COMPILE 946981 Con	lix C 37 ny G 27 33 11 12 12 13 14	10 7 30 31 33 51 14 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3	0 1 6 0 0 3 0 115 4	* 5 3 *			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hel Mar 12:6: vertebrate 2235 COMPILE 946981 (Cot	lik C 37 ny G 27 33 11 11 12 12	10 7 30 14 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3	0 1 6 0 0 3 0 115 4	*			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hel Mar 12:5: vertebrate P235 COMPILE 946981 (Cor	lix C 37 ny G 27 33 T 18 ess 11 11 12 12 12	10 7 30 11 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3		× 5 3 3			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hel Mar 12.6 vertebrat 2235 COMPILE 946981 (Cor	Ikk C 37 ny G 27 7 18 33 98 11 10 12 0 11	10 10 10 10 20	116 113 0 1 0 0	0 1 3	0 1 6 0 0 3 0 115 4	*			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hel Mai 12.5 verfebrat P235 COMPILE 946981 Com	IX C 37 NY G 27 33 T 18 55 II 10 18 D D	33 51 14 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3		* 5 3 *			
	Class Species NCB1 ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hela Mar 12.5 verebrat 235 COMPILE 945981 (Cor	IX C 37 IX G 27 33 S T 16 18 11 10 18 11 10 10 11	33 51 14 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3		× 5 3 *			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Held Mai 12.6 verfebrat P235 COMPILE 946981 (cor		33 51 14 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3	10 1 6 0 0 3 0 115 4	•			
	Class Species NCB1 ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hela Mar 12.5 verebrat 2235 COMPILE 945981 (Cor		33 51 14 4 26 25 41 56 41 33 3 5	116 113 0 1 0 0	0 1 3		•			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Held Mar 12.6 verfebrat P235 COMPILE 946981		103 51 14 4 26 25 41 56 41 33 3 5		0 1 3	10 1 6 0 0 3 0 115 4	*			
	Class Species NCB1 ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Held Mar 12.6 vertebrat P235 COMPILE 945981 Cor		1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		0 1 3		*			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Held		133 51 14 4 26 25 41 56 41 33 3 5		0 1 3		*			
	Class Species NCB1 ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hela Mar 12.5 verebrate P235 COMPILE 945981 Cor ComPiLE		133 61 14 4 26 25 41 66 41 33 3 5 2		0 1 3					
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hele Mar 12.2 vertebrat P2352 COMPILE S45881 (Cor		23 6 14 4 26 25 41 60 41 33 3 6		0 1 3		5			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Held Mar 12.6 vertebrate 2355 COMPILE 945981 Cor 0		33 51 14 5 28 25 41 6 41 33 3 5 41 33 3 5 28 25 41 6 41 33 3 5 41 5 415		0 1 3		5			
	Class Species NCBID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hele Mar 12.2 vertebrat P235 COMPILE S45881 ComPILE S4581 ComPILE S45811 ComPILE S45811		1 1 4 4 2 2 1 5 4 2 2 4 5 5 4 3 3 5 5 1 2 2 4 5 2 2 4 5 5 2 2 4 5 5 2 2 4 5 5 2 2 4 5 5 2 2 4 5 5 3 5 5 5 5 3 5 5 5 5 3 5 5 5 5 3 5 5 5 5 3 5 5 5 5 3 5 5 5 5 3 5 5 5 5 3 5 5 5 5		0 1 3		•			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Held Mar 12.6 vertebrate P235: COMPILE 945981 Cor ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 945981 Cor P350 ComPILE 9459810 Com	IX G G G IX G G G G IX G T T T T IX IX T <td>10 5 5 14 4 5 5 4 1 5 5 1 5 1 5 1 5 1 5 1</td> <td></td> <td></td> <td></td> <td>5 3 •</td> <td></td> <td></td> <td></td>	10 5 5 14 4 5 5 4 1 5 5 1 5 1 5 1 5 1 5 1				5 3 •			
	Class Species NCBID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hel Mar 12.2 vertebrat P235 COMPILE S45581 COMPILE S45581 ComPile S45		1 1 4 4 20 20 41 50 44 50 41 33 5 1 44 50 50 9 2 2 1 50 1 50 1 10 2 2 1 50 1 1 1 10 2 2 1 50 1		0 1 3 3	0 1 6 0 0 3 0 115 4 2 20 2	•			
	Class Species NCB1D Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hele Mar 12.5 vertebrate P235 COMPILE 945981 (Col Col P235 ComPILE 945981 (Col P235 ComPILE 945981 (Col P235 Col P2 Col P235 Col P235 Col P2 Col P2 Col P2 Col P2	IX C IX IX G IX	1 1 1 1 2 2 1 4 6 4 3 3 5 1 4 4 3 3 5 1 4 6 4 3 3 5 1 5 1 1 9 2 2 2 1 1 1 1 1 1	1116 113 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		0 1 6 0 0 3 0 115 4 0 13 0 115 4	5 3 			
	Class Species NCBI ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Hele Mar Mar 12.2 Vertebrat P235 COMPILE S45881 ComPILE S4581 ComPILE ComPILE S4581 ComPILE S4				0 1 3 1 3 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 1 1					
	Class Species NCB ID Inf. Content SuperGroup Protein Acc. Type PMID Report Best Occurrences	Other Alpha-Help Mar Mar 12.6 verlebrat P235 COMPIL 945981 Co Co Mar Co Verlebrat 945981 Co Co Verlebrat 945981 Co Co Verlebrat 945981 Verlebrat Co Verlebrat Second	IX C IX IX C IX IX G IX	1 1 1 1 2 2 1 1 4 6 2 2 1 1 6 6 4 3 3 5 1 5 6 4 3 3 5 1 5 6 1 5 1		0 1 3 1 3 1 1 3 1 1 1 1 1 1 1 1 1 1 1 1 1	0 1 6 1 0 1 8 1 0 115 4 0 115 4 1 15 1 1 1 15 1 1 15 1 15 1 1 15 1 15 1 1 15 1 15 1 1 15 1 1				

Meta-Motif Analyzers

http://131.174.198.125/bioinfo/gimmemotifs/

GimmeMotifs: a *de novo* motif prediction pipeline, especially suited for ChIP-seq datasets. It incorporates several existing motif prediction algorithms in an ensemble method to predict motifs and clusters these motifs using the weighted information content (WIC) similarity scoring metric.

BioProspector http://motif.stanford.edu/distributions/bioprospector/

GADEM http://www.niehs.nih.gov/research/resources/software/gadem/index.cfm

Improbizer http://users.soe.ucsc.edu/~kent/

MDmodule (included in the MotifRegressor Package) http://www.math.umass.edu/~conlon/mr.html

MEME http://meme.sdsc.edu/

MoAn http://moan.binf.ku.dk/

MotifSampler http://homes.esat.kuleuven.be/~sistawww/bioi/thijs/download.html

Trawler http://ani.embl.de/trawler/

Weeder http://159.149.160.51/modtools/

Differential binding analysis 1

• **THOR** is an HMM-based approach to detect and analyze differential peaks in two sets of ChIP-seq data from distinct biological conditions with replicates.

• Performs genomic signal processing and normalization, peak calling and p-value calculation in an integrated framework.



IRAK3

Gain



Differential binding analysis 2

- **Diffbind** is a Bioconductor package by **Stark** *et al.*, for identifying sites that are differentially bound between two sample groups.
- It includes functions to support the processing of peak sets, overlapping and merging peak sets, counting sequencing reads overlapping intervals in peak sets, and identifying statistically significantly differentially bound sites based on evidence of binding affinity (measured by differencessing read densities).
- More on THOR and DiffBind @ the practical!

