Downstream Analysis of ChIP-seq Data

Shamith Samarajiwa
MRC Cancer Unit
University of Cambridge

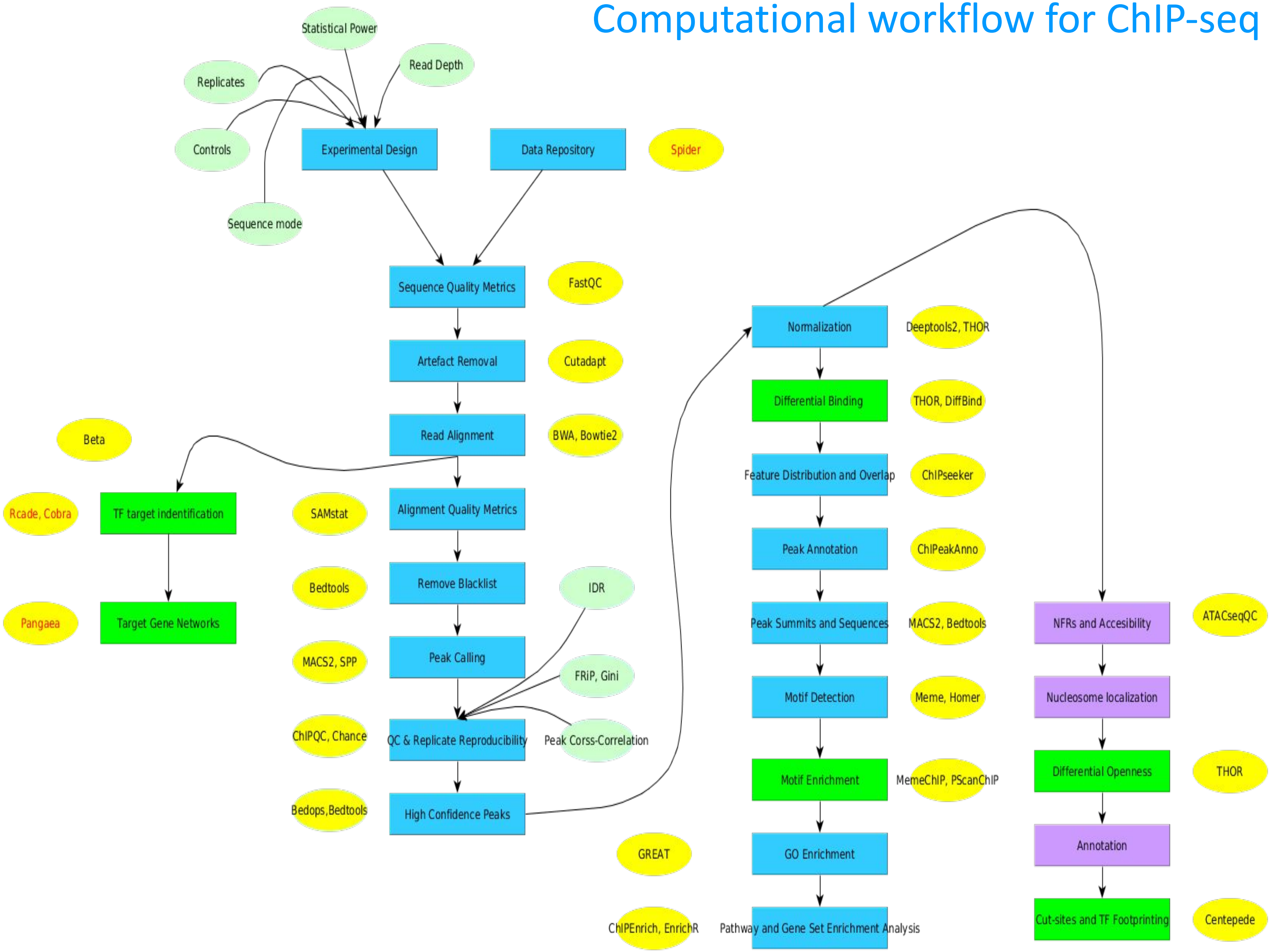
CRUK Bioinformatics Summer School 2021
27th July 2021

Summary

Downstream analysis for extracting meaningful biology :

- Normalization, Visualization and Interval Operations
- 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
- Differential binding analysis
- Integration with transcriptomic data to identify TF direct targets
- Network Biology applications

Computational workflow for ChIP-seq and ATAC-seq



Statistical Power

Replicates

Read Depth

Controls

Sequence mode

Experimental Design

Data Repository

Spider

Sequence Quality Metrics

FastQC

Artefact Removal

Cutadapt

Read Alignment

BWA, Bowtie2

Beta

Rcade, Cobra

TF target indentification

Target Gene Networks

Pangaea

SAMstat

Alignment Quality Metrics

Bedtools

Remove Blacklist

IDR

MACS2, SPP

Peak Calling

FRiP, Gini

ChIPQC, Chance

QC & Replicate Reproducibility

Peak Corss-Correlation

Bedops,Bedtools

High Confidence Peaks

GREAT

ChIPEnrich, EnrichR

Normalization

Deeptools2, THOR

Differential Binding

THOR, DiffBind

Feature Distribution and Overlap

ChIPseeker

Peak Annotation

ChIPeakAnno

Peak Summits and Sequences

MACS2, Bedtools

Motif Detection

Meme, Homer

Motif Enrichment

MemeChIP, PScanChIP

GO Enrichment

Pathway and Gene Set Enrichment Analysis

NFRs and Accesibility

ATACseqQC

Nucleosome localization

Differential Openness

THOR

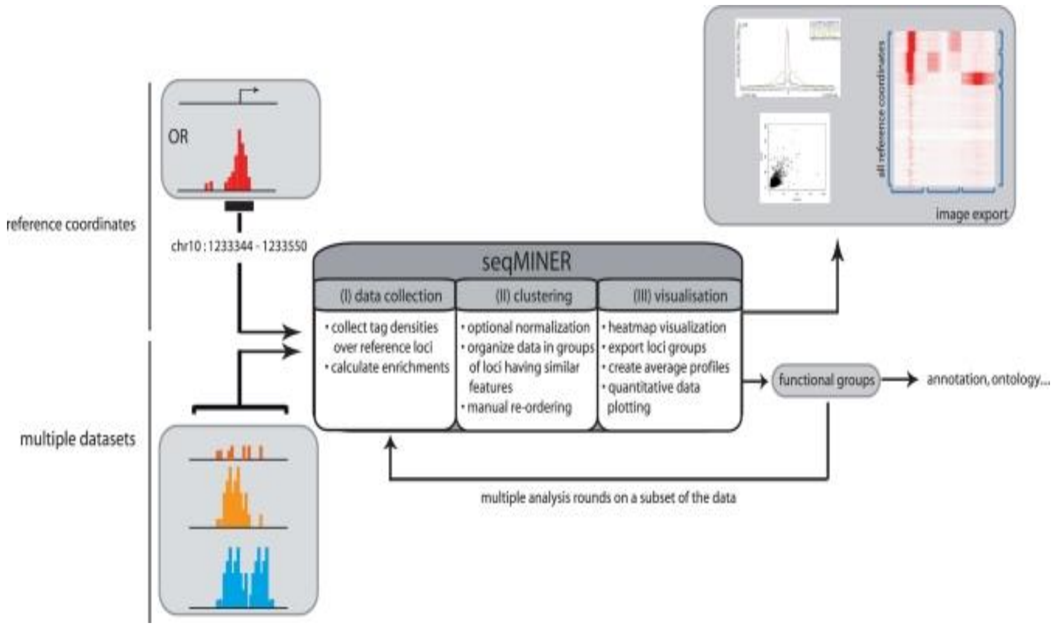
Annotation

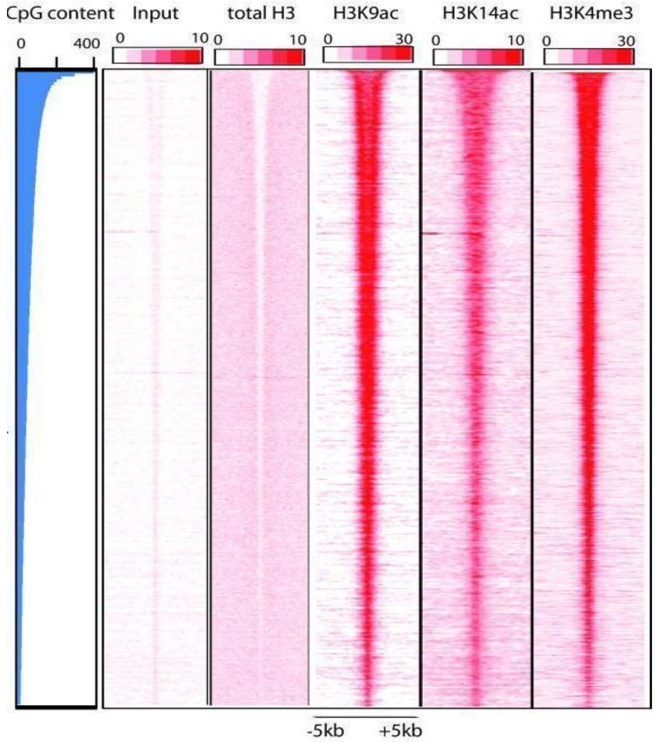
Cut-sites and TF Footprinting

Centepede

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.

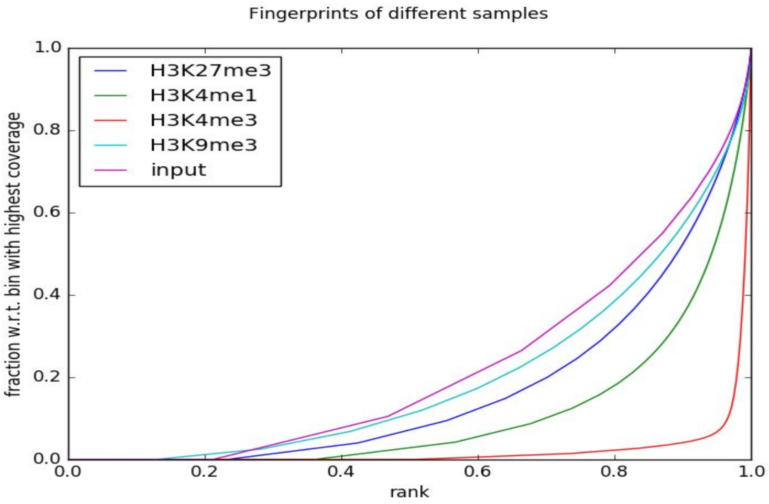




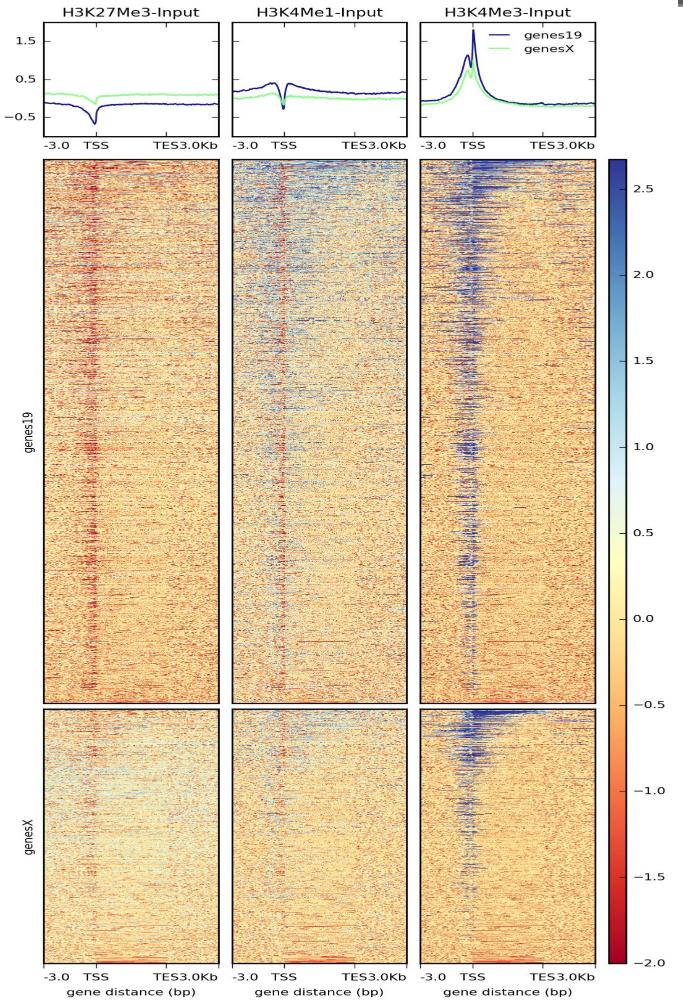
Ye *et al*., 2011, Nucleic Acids Res.

Compare, Normalize & Visualize 2

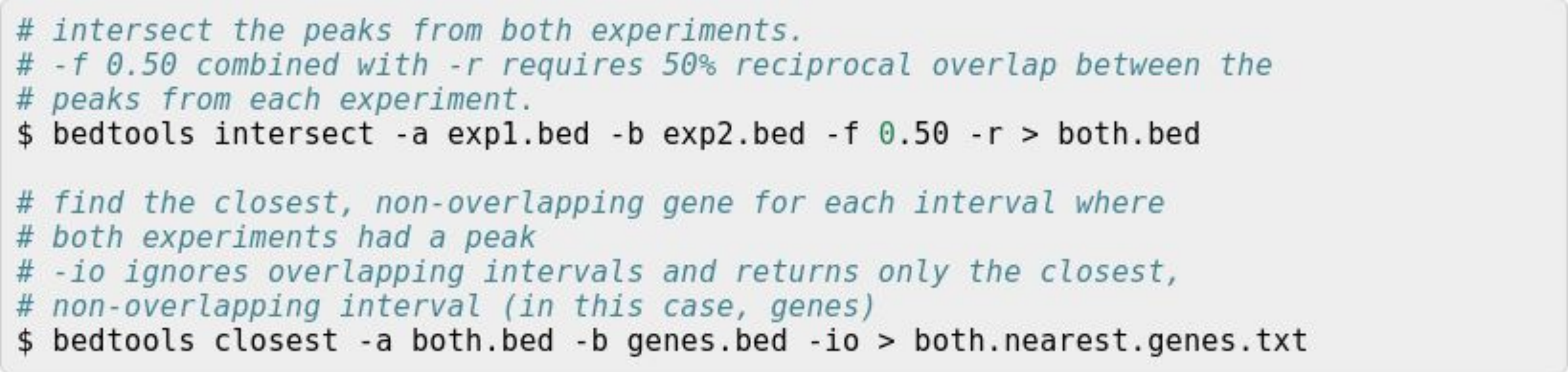
- **deepTools2** sequence depth or input normalization, GC bias correction
- Plot signal profiles
- Customized heat-maps
- PCA, correlation and fingerprint plots (chip enrichment)



Ramírez et al., 2016, Nucleic Acids Res.



BEDtools for genomic interval operations



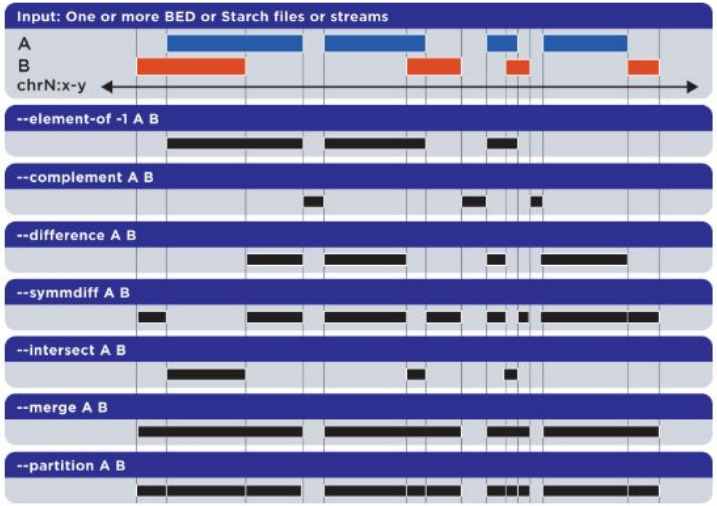
```
# intersect the peaks from both experiments.
# -f 0.50 combined with -r requires 50% reciprocal overlap between the
# peaks from each experiment.
$ bedtools intersect -a exp1.bed -b exp2.bed -f 0.50 -r > both.bed

# find the closest, non-overlapping gene for each interval where
# both experiments had a peak
# -io ignores overlapping intervals and returns only the closest,
# non-overlapping interval (in this case, genes)
$ bedtools closest -a both.bed -b genes.bed -io > both.nearest.genes.txt
```

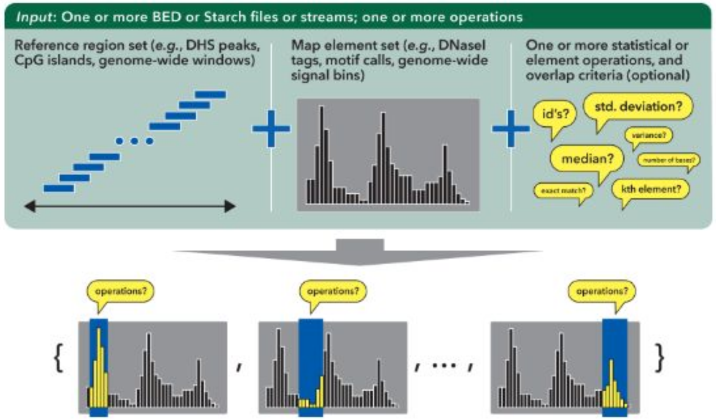
- bedtools are a swiss-army knife of tools for a wide-range of genomics analysis tasks. The most widely-used tools enable *genome arithmetic*: that is, set theory on the genome. **bedtools** allows one to
    - *Intersect*
    - *Merge*
    - *Count*
    - *Complement*
    - *shuffle* genomic intervals from multiple files in widely-used genomic file formats such as BAM, BED, GFF/GTF, VCF.

Bedops genome analysis toolkit

BEDOPS is an open-source command-line toolkit that performs highly efficient and scalable Boolean and other set operations, statistical calculations, archiving, conversion and other management of genomic data of arbitrary scale.



The bedmap core tool applies a wide variety of statistical and mapping operations to genomic inputs:


Input: One or more BED or Starch files or streams; one or more operations
Reference region set (e.g., DHS peaks, CpG islands, genome-wide windows)
Map element set (e.g., DNaseI tags, motif calls, genome-wide signal bins)
One or more statistical or element operations, and overlap criteria (optional)
id's?
std. deviation?
variance?
median?
number of bases?
exact match?
kth element?
operations?
operations?
operations?
{
,
, ... ,
}

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

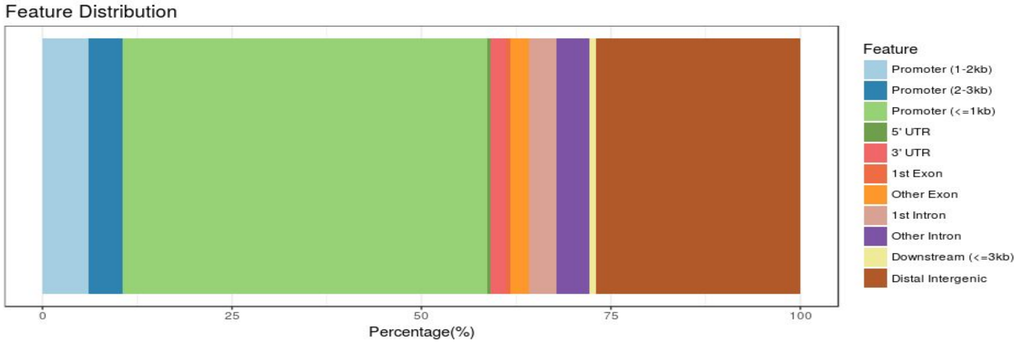
PeakAnalyzer

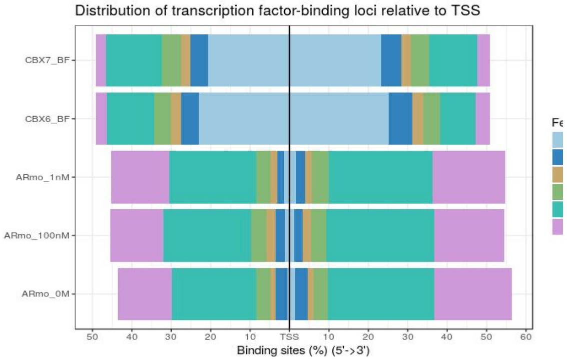
- A set of high-performance utilities for the automated processing of experimentally-derived peak regions and annotation of genomic loci.
- Consists of PeakSplitter and PeakAnnotator.
- Biologist' friendly tool.
- Get latest genome annotation files from Ensembl (gtf format) or UCSC (BED format).
- Map 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.

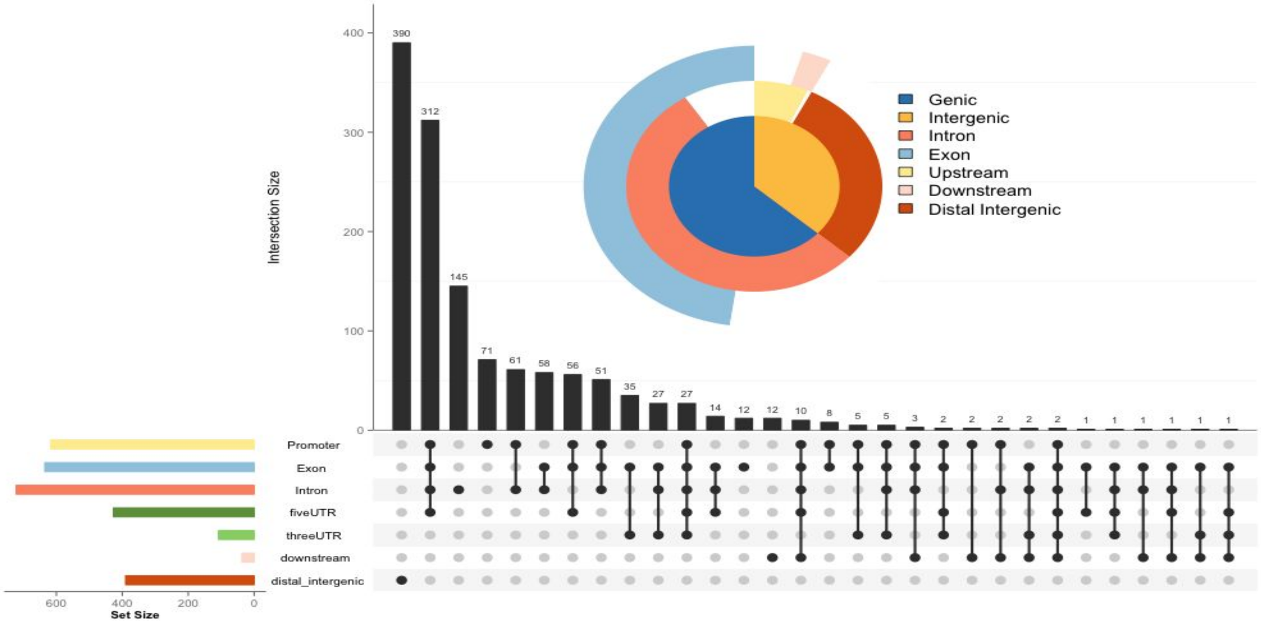
Salmon-Divon et al., 2010, BMC Bioinformatics.

Peaks distribution across features

ChIPseeker (BioC)



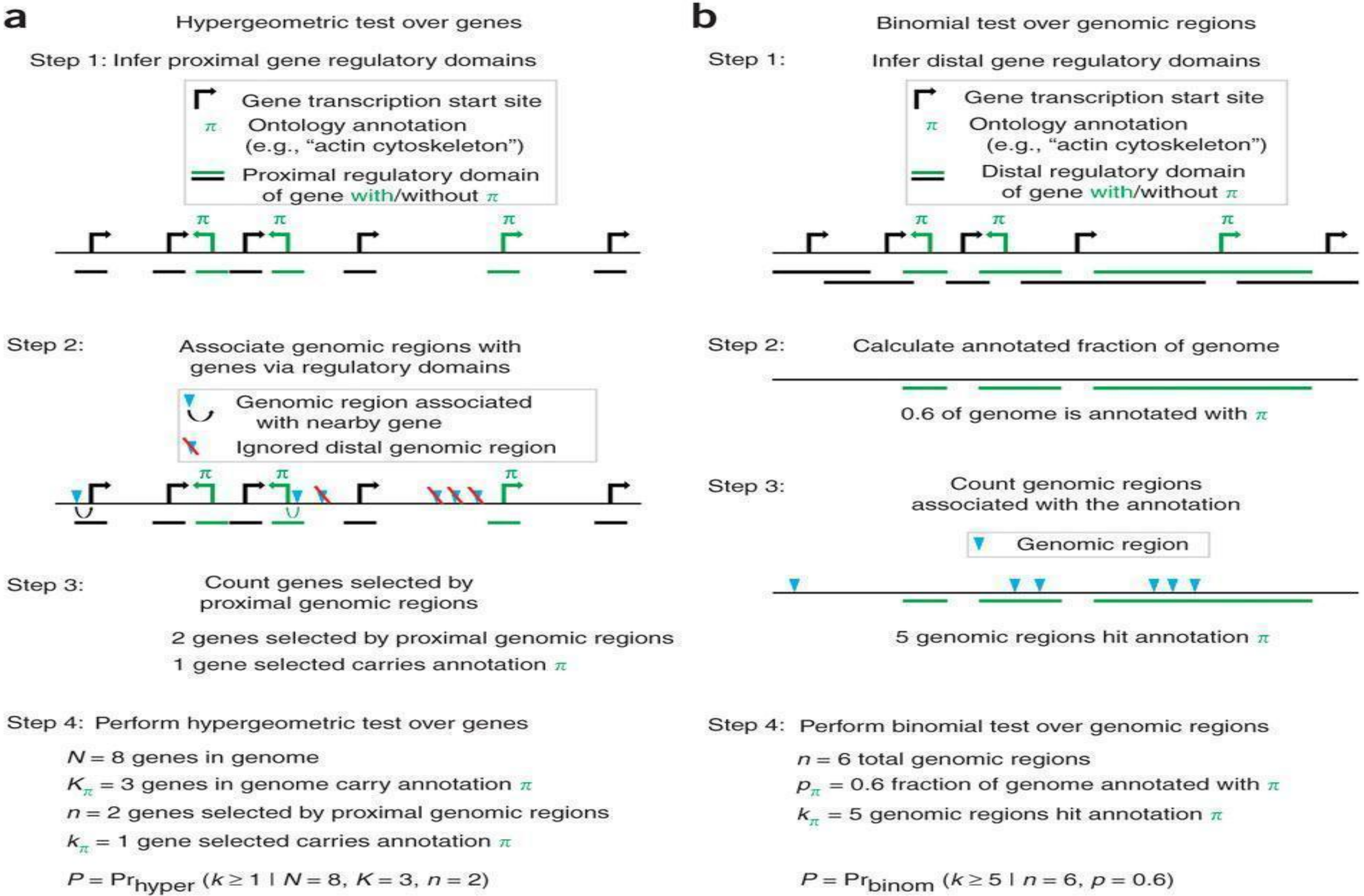




Yu et al., 2015, Bioinformatics

Functional Enrichment Analysis 1

GREAT & rGREAT: Genomic Regions Enrichment of Annotations Tool



Hypergeometric test over genes

Step 1: Infer proximal gene regulatory domains

Gene transcription start site

π Ontology annotation (e.g., "actin cytoskeleton")

Proximal regulatory domain of gene with/without π

Step 2: Associate genomic regions with genes via regulatory domains

Genomic region associated with nearby gene

Ignored distal genomic region

Step 3: Count genes selected by proximal genomic regions

2 genes selected by proximal genomic regions

1 gene selected carries annotation π

Step 4: Perform hypergeometric test over genes

$N = 8$ genes in genome

$K_{\pi} = 3$ genes in genome carry annotation π

$n = 2$ genes selected by proximal genomic regions

$k_{\pi} = 1$ gene selected carries annotation π

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

Binomial test over genomic regions

Step 1: Infer distal gene regulatory domains

Gene transcription start site

π Ontology annotation (e.g., "actin cytoskeleton")

Distal regulatory domain of gene with/without π

Step 2: Calculate annotated fraction of genome

0.6 of genome is annotated with π

Step 3: Count genomic regions associated with the annotation

Genomic region

5 genomic regions hit annotation π

Step 4: Perform binomial test over genomic regions

$n = 6$ total genomic regions

$p_{\pi} = 0.6$ fraction of genome annotated with π

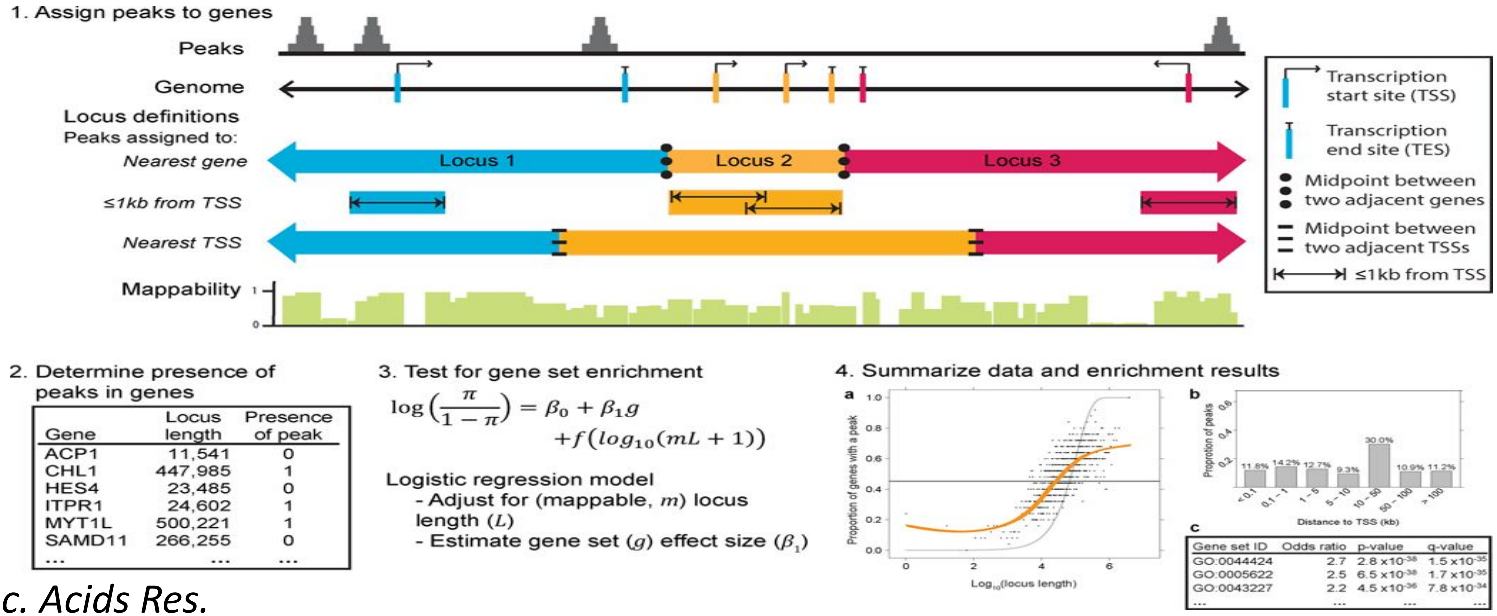
$k_{\pi} = 5$ genomic regions hit annotation π

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

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



Gene	Locus length	Presence of peak
ACP1	11,541	0
CHL1	447,985	1
HES4	23,485	0
ITPR1	24,602	1
MYT1L	500,221	1
SAMD11	266,255	0
...	...	...

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 g$$
$$+ f(log_{10}(mL + 1))$$

Logistic regression model
- Adjust for (mappable, m) locus length (L)
- Estimate gene set (g) effect size (β_1)

Gene set ID	Odds ratio	p-value	q-value
GO:0044424	2.7	2.8×10^{-38}	1.5×10^{-35}
GO:0005622	2.5	6.5×10^{-38}	1.7×10^{-35}
GO:0043227	2.2	4.5×10^{-36}	7.8×10^{-34}
...	...	...	...

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

Functional Enrichment Analysis 3

GAT tests if two sets of genomic intervals are associated more than expected by chance.

- Association typically means nucleotide overlap, but other measures such as the distance between elements or the number of overlapping segments can be used.
- gat-great implements a command line version of GREAT

https://gat.readthedocs.io/en/latest/

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.

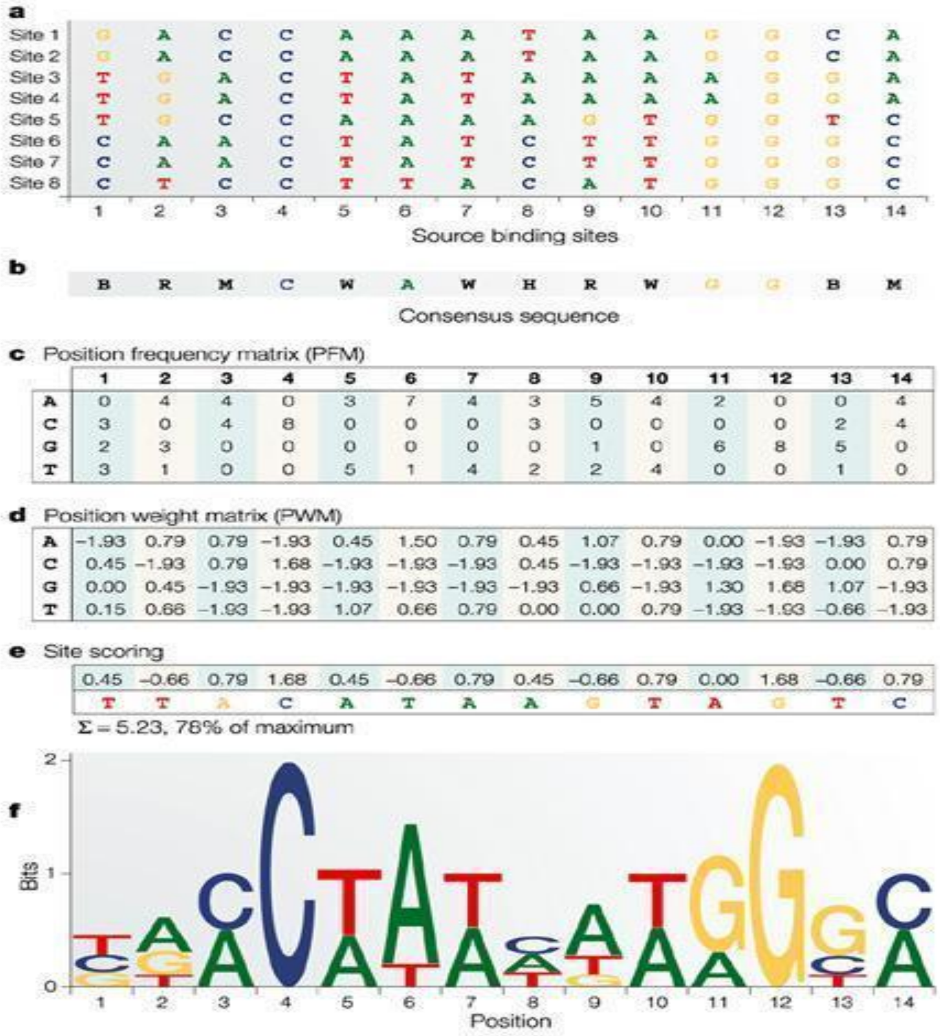
- Use *de novo* motif calling and alignment to build your own PWMs!

- 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



a

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

Source binding sites

b

B R M C W A W H R W G G B M

Consensus sequence

c Position frequency matrix (PFM)

	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

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
T	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

e Site scoring

0.45	–0.66	0.79	1.68	0.45	–0.66	0.79	0.45	–0.66	0.79	0.00	1.68	–0.66	0.79
T	T	A	C	A	T	A	A	G	T	A	G	T	C

Σ = 5.23, 78% of maximum

f

Nature Reviews | Genetics

PWM conversion:

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

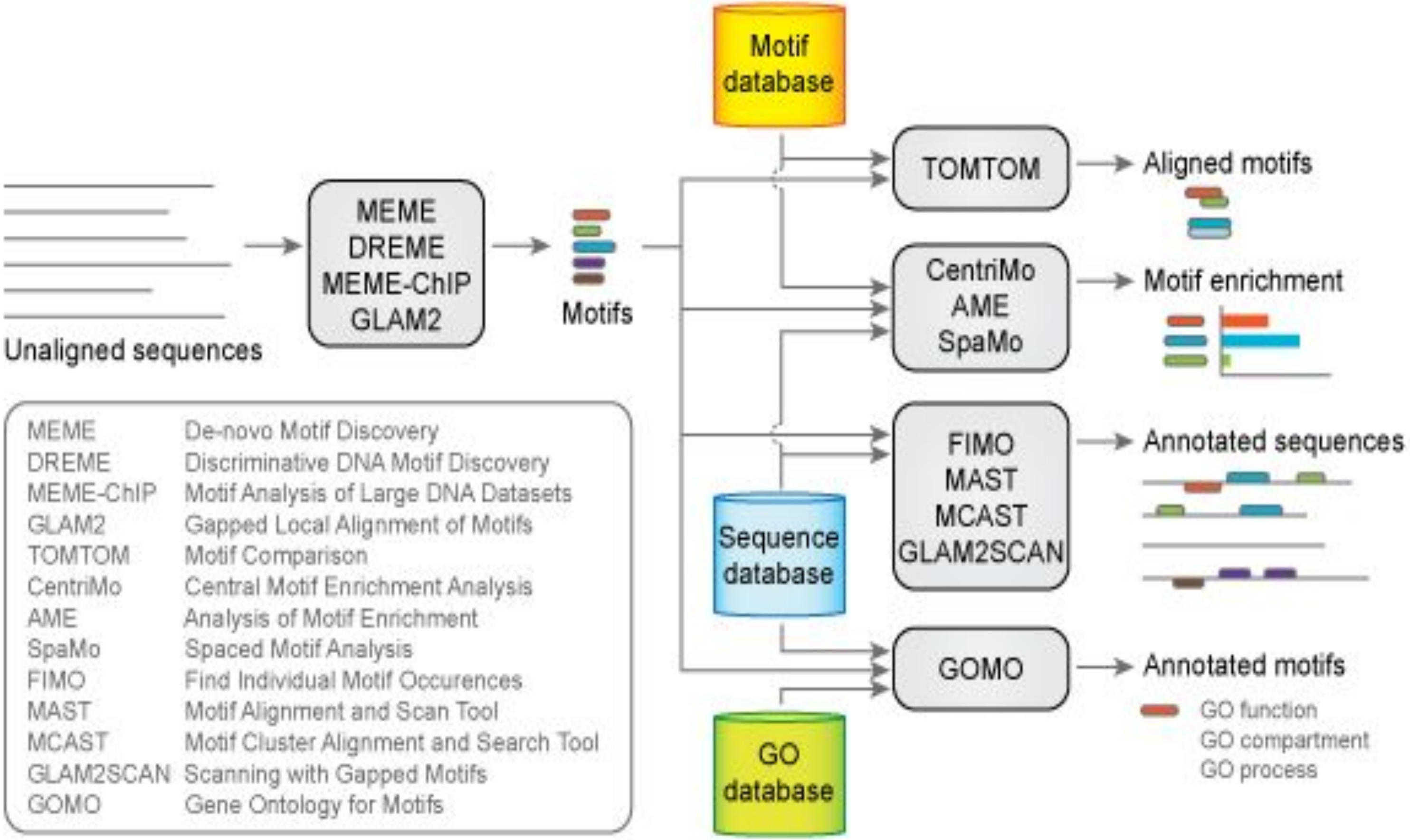
TFBS PWM/PFM sources

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 detection

- HOMER v4 http://homer.salk.edu/homer/index.html
- Large number of (Perl and C++) tools for ChIP-seq analysis.
- 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.

Meme Suite



MEME	De-novo Motif Discovery
DREME	Discriminative DNA Motif Discovery
MEME-ChIP	Motif Analysis of Large DNA Datasets
GLAM2	Gapped Local Alignment of Motifs
TOMTOM	Motif Comparison
CentriMo	Central Motif Enrichment Analysis
AME	Analysis of Motif Enrichment
SpaMo	Spaced Motif Analysis
FIMO	Find Individual Motif Occurences
MAST	Motif Alignment and Scan Tool
MCAST	Motif Cluster Alignment and Search Tool
GLAM2SCAN	Scanning with Gapped Motifs
GOMO	Gene Ontology for Motifs

Bailey TL. et al., "MEME SUITE: tools for motif discovery and searching." Nucleic Acids Res. 2009

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.

MEME-ChIP

url: http://meme.nbcr.net

- Given a set of genomic regions, it performs
  - 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)
- Uses two algorithms for motif discovery:
  - 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.
- Motif identification:
  - FIMO -identify individual motifs

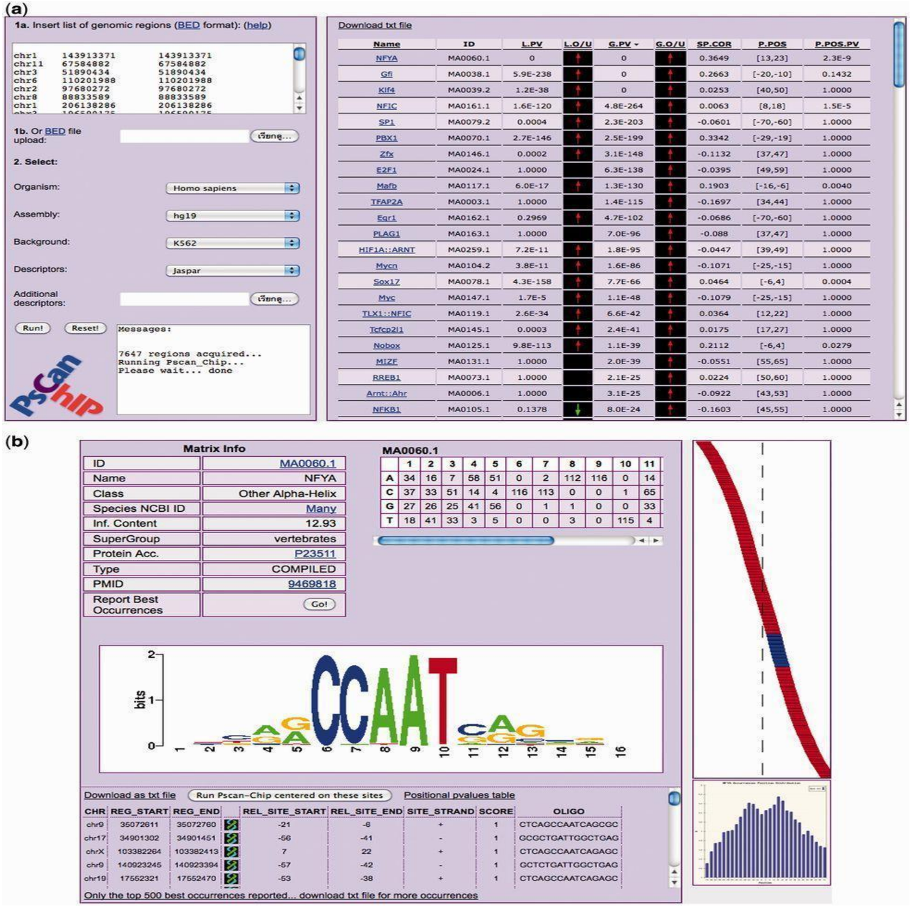
Machanick and Bailey, "MEME-ChIP: motif analysis of large DNA datasets." 2011 Bioinformatics

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.
- Produces sequence logo's and motif enrichment distribution plots.

Zambelli et al., 2013 Nucleic Acids Res.



(a)

1a. Insert list of genomic regions (BED format): (help)

```
chr1     143913371        143913371
chr11    67584882         67584882
chr3     51890434         51890434
chr6     110201988        110201988
chr2     97680272         97680272
chr8     88833589         88833589
chr1     206138286        206138286
```

1b. Or BED file upload: เรียกดู...

2. Select:

Organism: Homo sapiens

Assembly: hg19

Background: K562

Descriptors: Jaspar

Additional descriptors: เรียกดู...

Run! Reset!

```
Messages:

7647 regions acquired...
Running Pscan_Chip...
Please wait... done
```

Download txt file

Name	ID	L.PV	L.O/U	G.PV	G.O/U	SP.COR	P.POS	P.POS.PV
NFYA	MA0060.1	0	↑	0	↑	0.3649	[13,23]	2.3E-9
Gfi	MA0038.1	5.9E-238	↑	0	↑	0.2663	[-20,-10]	0.1432
Klf4	MA0039.2	1.2E-38	↑	0	↑	0.0253	[40,50]	1.0000
NFIC	MA0161.1	1.6E-120	↑	4.8E-264	↑	0.0063	[8,18]	1.5E-5
SP1	MA0079.2	0.0004	↑	2.3E-203	↑	-0.0601	[-70,-60]	1.0000
PBX1	MA0070.1	2.7E-146	↑	2.5E-199	↑	0.3342	[-29,-19]	1.0000
Zfx	MA0146.1	0.0002	↑	3.1E-148	↑	-0.1132	[37,47]	1.0000
E2F1	MA0024.1	1.0000		6.3E-138	↑	-0.0395	[49,59]	1.0000
Mafb	MA0117.1	6.0E-17	↑	1.3E-130	↑	0.1903	[-16,-6]	0.0040
TFAP2A	MA0003.1	1.0000		1.4E-115	↑	-0.1697	[34,44]	1.0000
Egr1	MA0162.1	0.2969	↑	4.7E-102	↑	-0.0686	[-70,-60]	1.0000
PLAG1	MA0163.1	1.0000		7.0E-96	↑	-0.088	[37,47]	1.0000
HIF1A::ARNT	MA0259.1	7.2E-11	↑	1.8E-95	↑	-0.0447	[39,49]	1.0000
Mycn	MA0104.2	3.8E-11	↑	1.6E-86	↑	-0.1071	[-25,-15]	1.0000
Sox17	MA0078.1	4.3E-158	↑	7.7E-66	↑	0.0464	[-6,4]	0.0004
Myc	MA0147.1	1.7E-5	↑	1.1E-48	↑	-0.1079	[-25,-15]	1.0000
TLX1::NFIC	MA0119.1	2.6E-34	↑	6.6E-42	↑	0.0364	[12,22]	1.0000
Tcfcp2l1	MA0145.1	0.0003	↑	2.4E-41	↑	0.0175	[17,27]	1.0000
Nobox	MA0125.1	9.8E-113	↑	1.1E-39	↑	0.2112	[-6,4]	0.0279
MIZF	MA0131.1	1.0000		2.0E-39	↑	-0.0551	[55,65]	1.0000
RREB1	MA0073.1	1.0000		2.1E-25	↑	0.0224	[50,60]	1.0000
Arnt::Ahr	MA0006.1	1.0000		3.1E-25	↑	-0.0922	[43,53]	1.0000
NFKB1	MA0105.1	0.1378	↓	8.0E-24	↑	-0.1603	[45,55]	1.0000

(b)

Matrix Info

ID	MA0060.1
Name	NFYA
Class	Other Alpha-Helix
Species NCBI ID	Many
Inf. Content	12.93
SuperGroup	vertebrates
Protein Acc.	P23511
Type	COMPILED
PMID	9469818
Report Best Occurrences	Go!

MA0060.1

	1	2	3	4	5	6	7	8	9	10	11
A	34	16	7	58	51	0	2	112	116	0	14
C	37	33	51	14	4	116	113	0	0	1	65
G	27	26	25	41	56	0	1	1	0	0	33
T	18	41	33	3	5	0	0	3	0	115	4

Download as txt file

Run Pscan-Chip centered on these sites

Positional pvalues table

CHR	REG_START	REG_END		REL_SITE_START	REL_SITE_END	SITE_STRAND	SCORE	OLIGO
chr9	35072611	35072760		-21	-6	+	1	CTCAGCCAATCAGCGC
chr17	34901302	34901451		-56	-41	-	1	GCGCTGATTGGCTGAG
chrX	103382264	103382413		7	22	+	1	CTCAGCCAATCAGAGC
chr9	140923245	140923394		-57	-42	-	1	GCTCTGATTGGCTGAG
chr19	17552321	17552470		-53	-38	+	1	CTCAGCCAATCAGAGC

Only the top 500 best occurrences reported... download txt file for more occurrences

Meta-Motif Analyzers

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. http://131.174.198.125/bioinfo/gimmemotifs/

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/

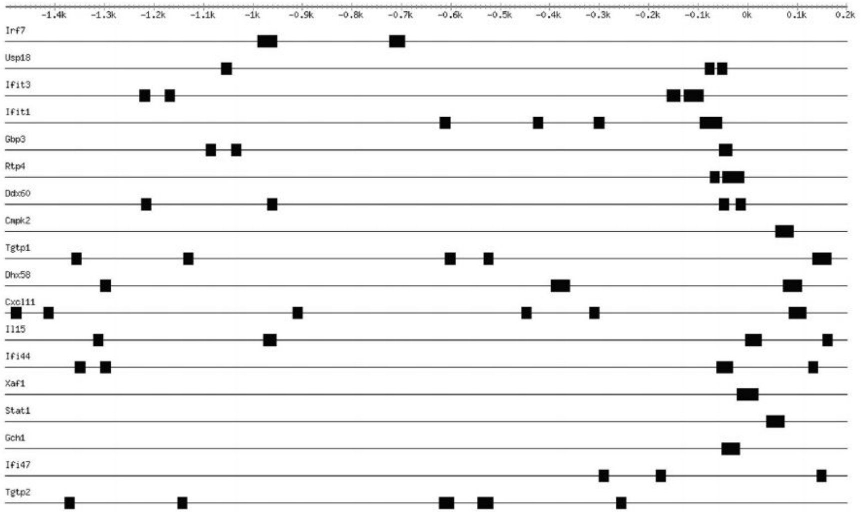
MotifSampler http://homes.esat.kuleuven.be/~sistawww/bioi/thijs/download.html

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

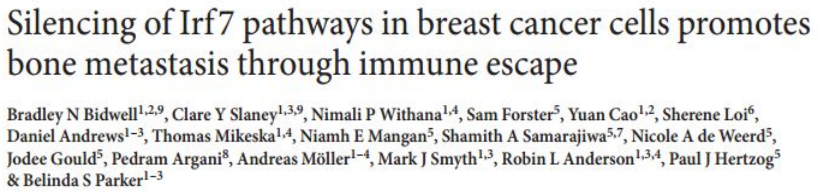
Weeder http://159.149.160.51/modtools/

Real world applications of Motif Enrichment Analysis

- Breast cancer metastasis is a key determinant of long-term patient survival.
- By comparing the transcriptomes of primary and metastatic tumor cells in a mouse model of spontaneous bone metastasis, we identified that a substantial number of genes suppressed in bone metastases are targets of the interferon regulatory factor IRF7.
- Restoration of *Irf7* activity in tumor cells or administration of interferon (which induces IRF7) led to reduced bone metastases and prolonged survival time.



nature medicine



Silencing of Irf7 pathways in breast cancer cells promotes bone metastasis through immune escape

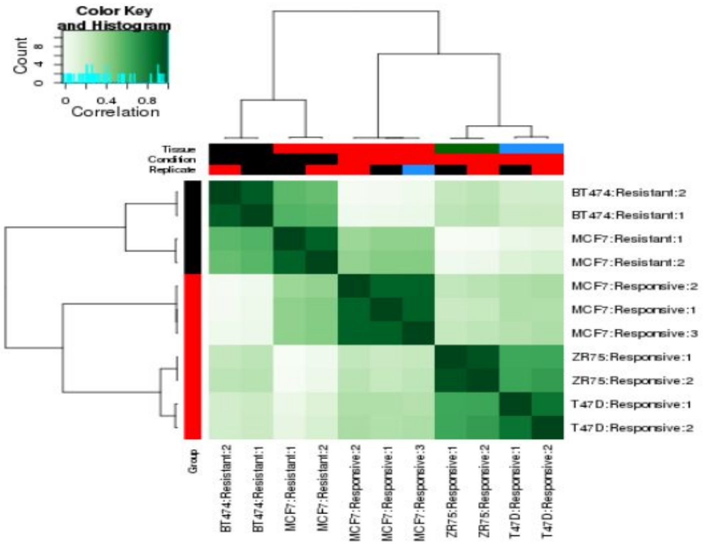
Bradley N Bidwell[1,2,9], Clare Y Slaney[1,3,9], Nimali P Withana[1,4], Sam Forster[5], Yuan Cao[1,2], Sherene Loi[6], Daniel Andrews[1–3], Thomas Mikeska[1,4], Niamh E Mangan[5], Shamith A Samarajiwa[5,7], Nicole A de Weerd[5], Jodee Gould[5], Pedram Argani[8], Andreas Möller[1–4], Mark J Smyth[1,3], Robin L Anderson[1,3,4], Paul J Hertzog[5] & Belinda S Parker[1–3]

Differential binding analysis 1

- **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 differences in read densities).

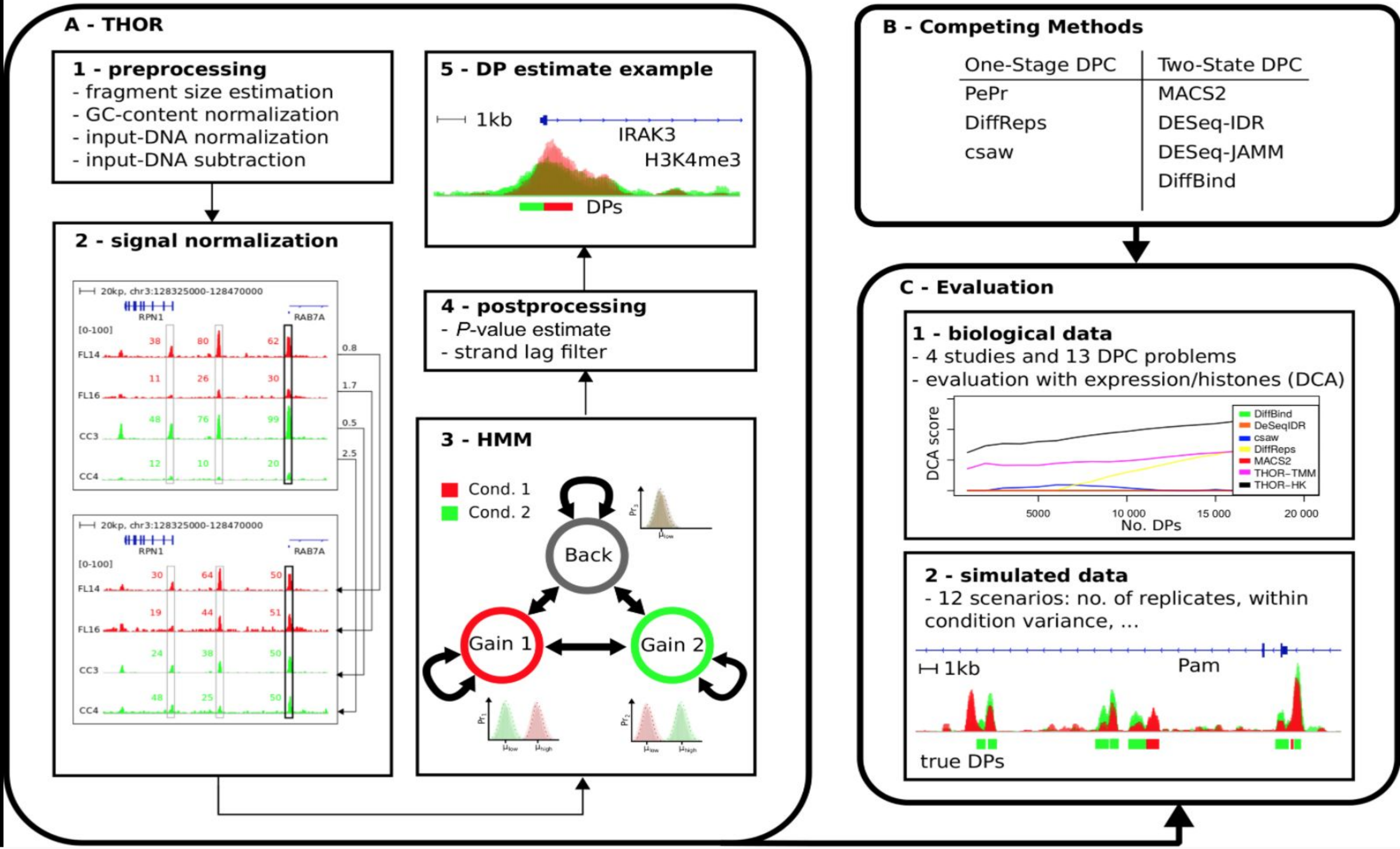
- More on DiffBind @ the practical!



Differential binding analysis 2

- **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.



A - THOR

1 - preprocessing
- fragment size estimation
- GC-content normalization
- input-DNA normalization
- input-DNA subtraction

2 - signal normalization

3 - HMM

4 - postprocessing
- *P*-value estimate
- strand lag filter

5 - DP estimate example

B - Competing Methods

One-Stage DPC	Two-State DPC
PePr	MACS2
DiffReps	DESeq-IDR
csaw	DESeq-JAMM
	DiffBind

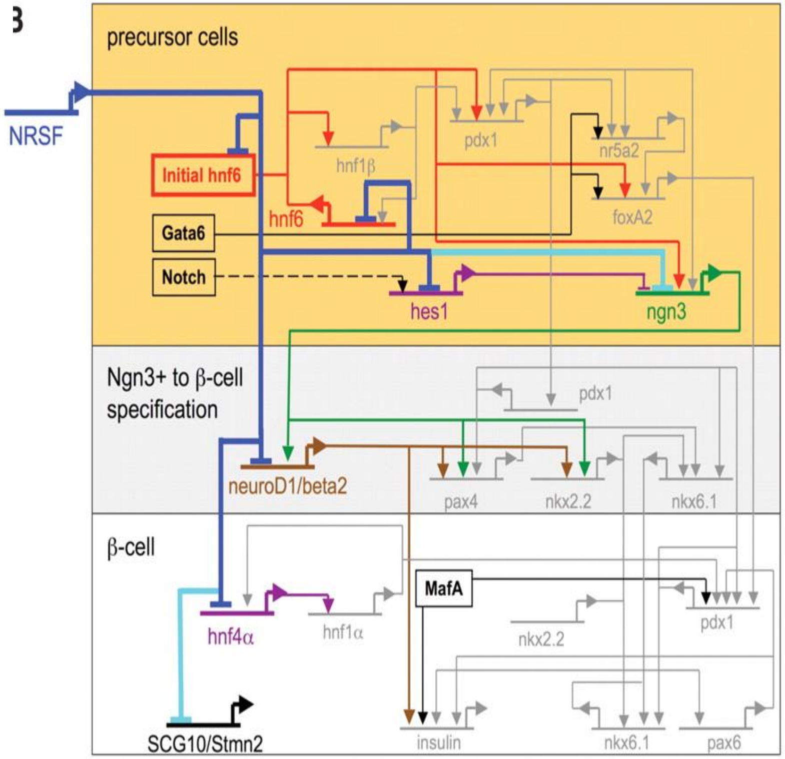
C - Evaluation

1 - biological data
- 4 studies and 13 DPC problems
- evaluation with expression/histones (DCA)

2 - simulated data
- 12 scenarios: no. of replicates, within condition variance, ...

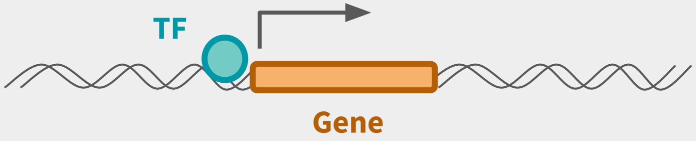
Regulomes: from target genes to networks

- Not all TF binding sites (cistrome) are transcriptionally active. The collection of transcriptionally active targets of a TF is it's regulome.
- Regulomes can be used to "explain" the phenotype under consideration and understand aspects of biological systems.
- Regulomes in combination with pathway and network modelling approaches can then be used deconvolute the networks underlying phenotypes.
- These networks provide information on connectivity, information flow, and regulatory, signaling and other interactions between cellular components.



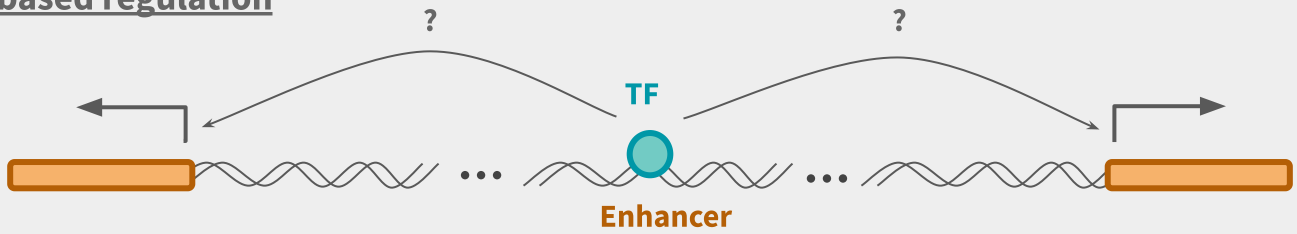
Detection of TF Direct Target

Promoter based regulation



- TF binding with **ChIP-seq**
- Gene expression with **RNA-seq/microarray**
- 3D architecture with **Hi-C**
- Regulatory element activity with **Histone ChIP-seq**

Enhancer based regulation



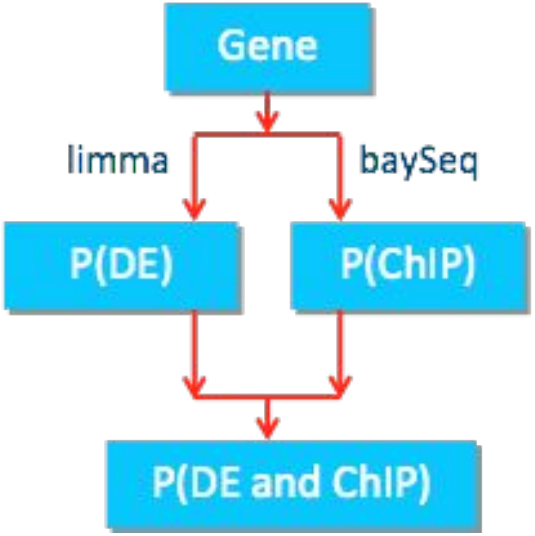
- Rcade (Bioconductor)
- COBRA
- Beta

Rcade: R-based analysis of ChIP-seq And Differential Expression

- Rcade is a Bioconductor package developed by Cairns *et al.*, that utilizes Bayesian methods to integrates ChIP-seq TF binding, with a transcriptomic Differential Expression (DE) analysis.
- The method is read-based and independent of peak-calling.
- Rcade can infer the direct targets of a transcription factor (TF).
- These targets should exhibit TF binding activity, and their expression levels should change in response to a perturbation of the TF.

Rcade

- **Rcade: R based analysis of ChIPseq And Differential Expression**
- Bayesian approach used to integrate ChIP-seq with differential expression to identify direct transcriptional targets of transcription factors.



Gene

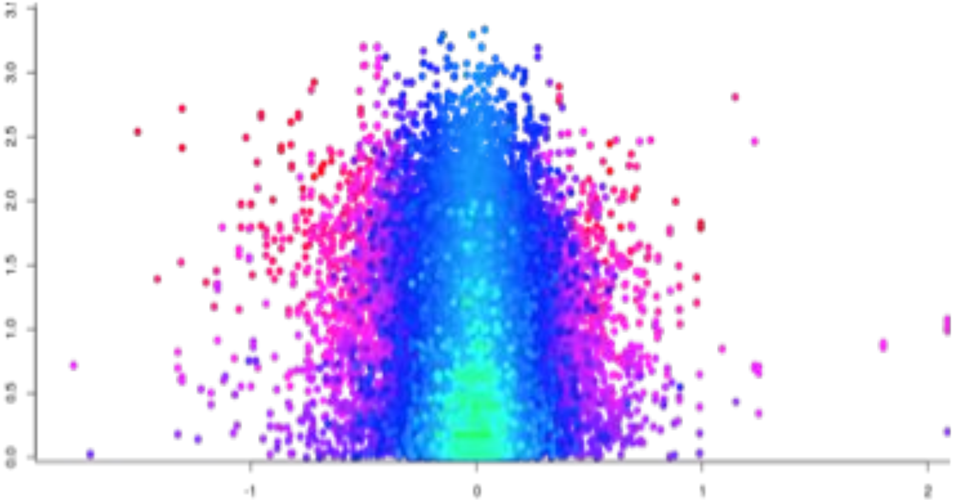
limma

baySeq

P(DE)

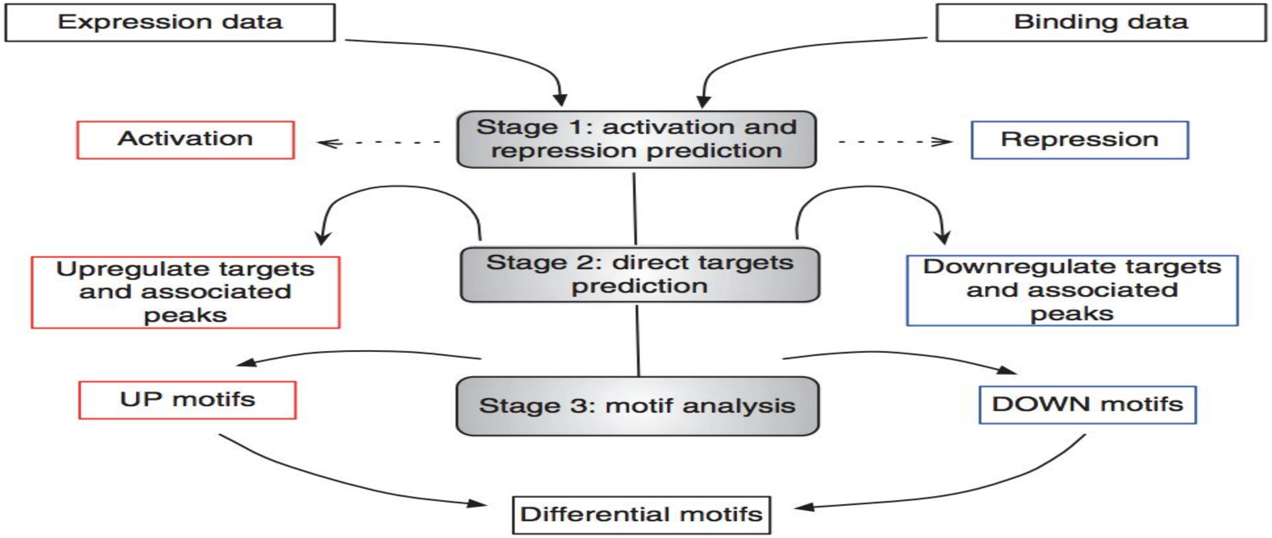
P(ChIP)

P(DE and ChIP)



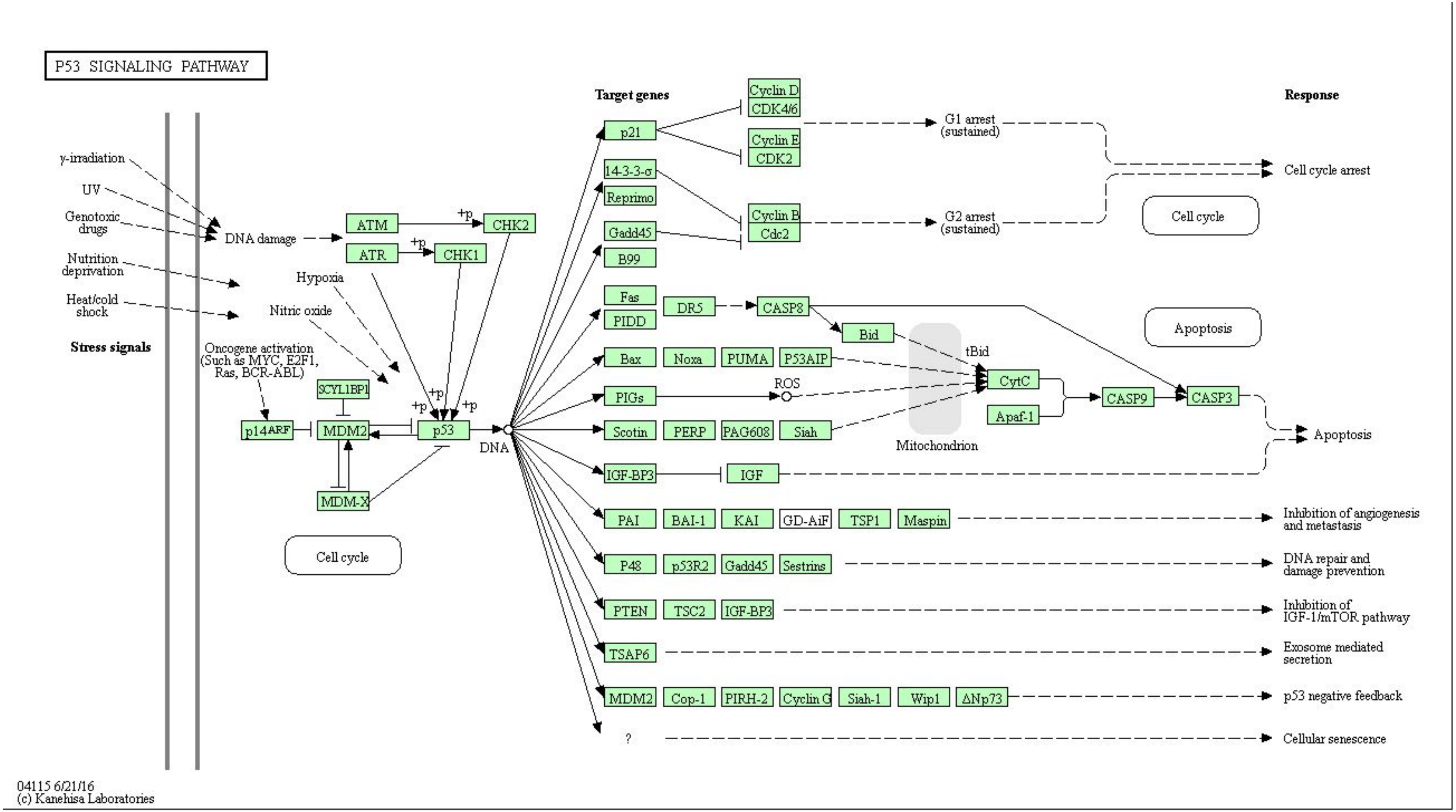
Beta

- Three main functionalities:
  - to predict whether a factor has activating or repressive function
  - to *infer* the factor's target genes
  - to identify the binding motif of the factor and its collaborators


Expression data
Binding data
Activation
Stage 1: activation and repression prediction
Repression
Upregulate targets and associated peaks
Stage 2: direct targets prediction
Downregulate targets and associated peaks
UP motifs
Stage 3: motif analysis
DOWN motifs
Differential motifs

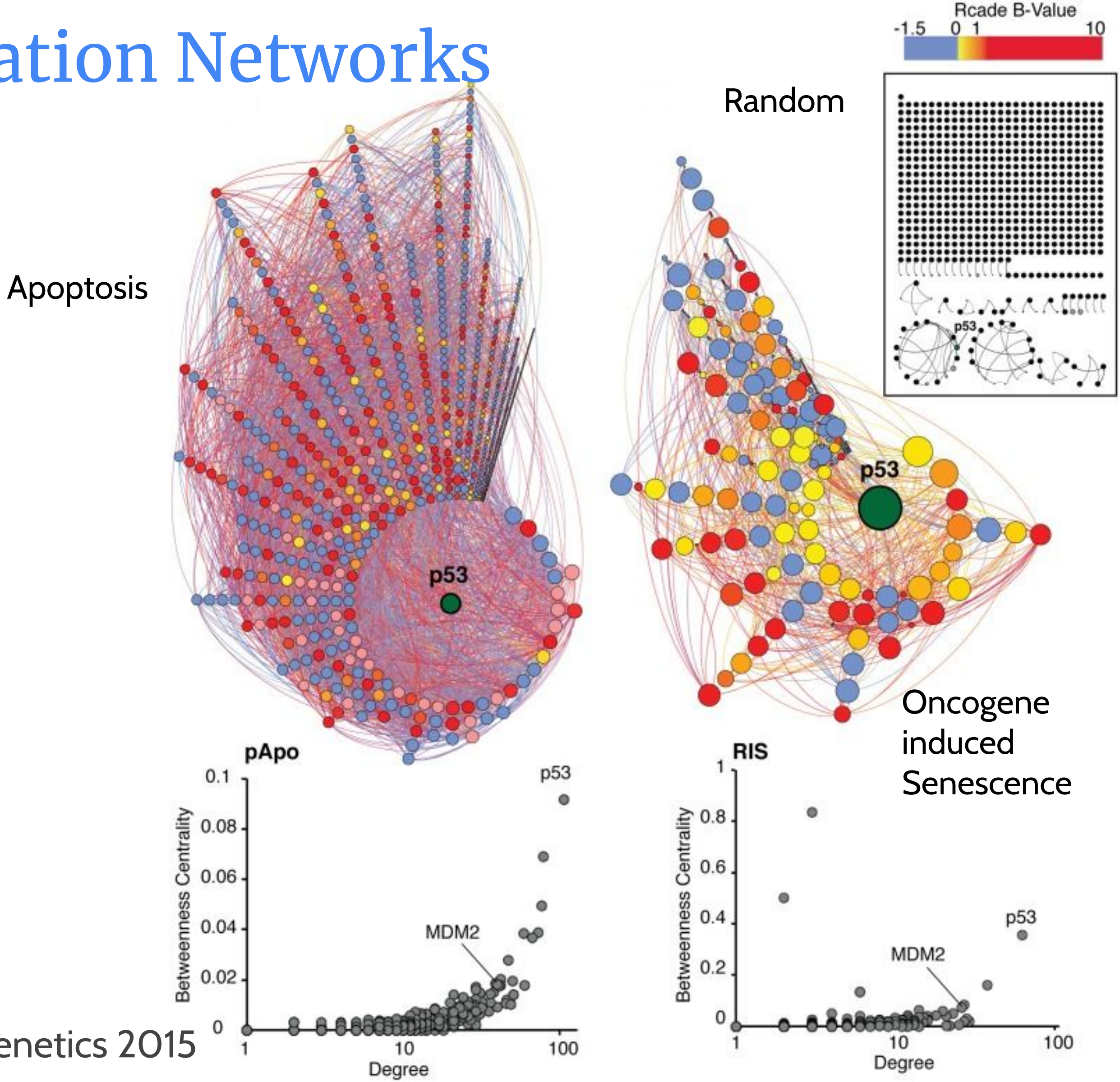
Wang, 2013 Nat Protoc. 2013

KEGG: p53 signalling pathway



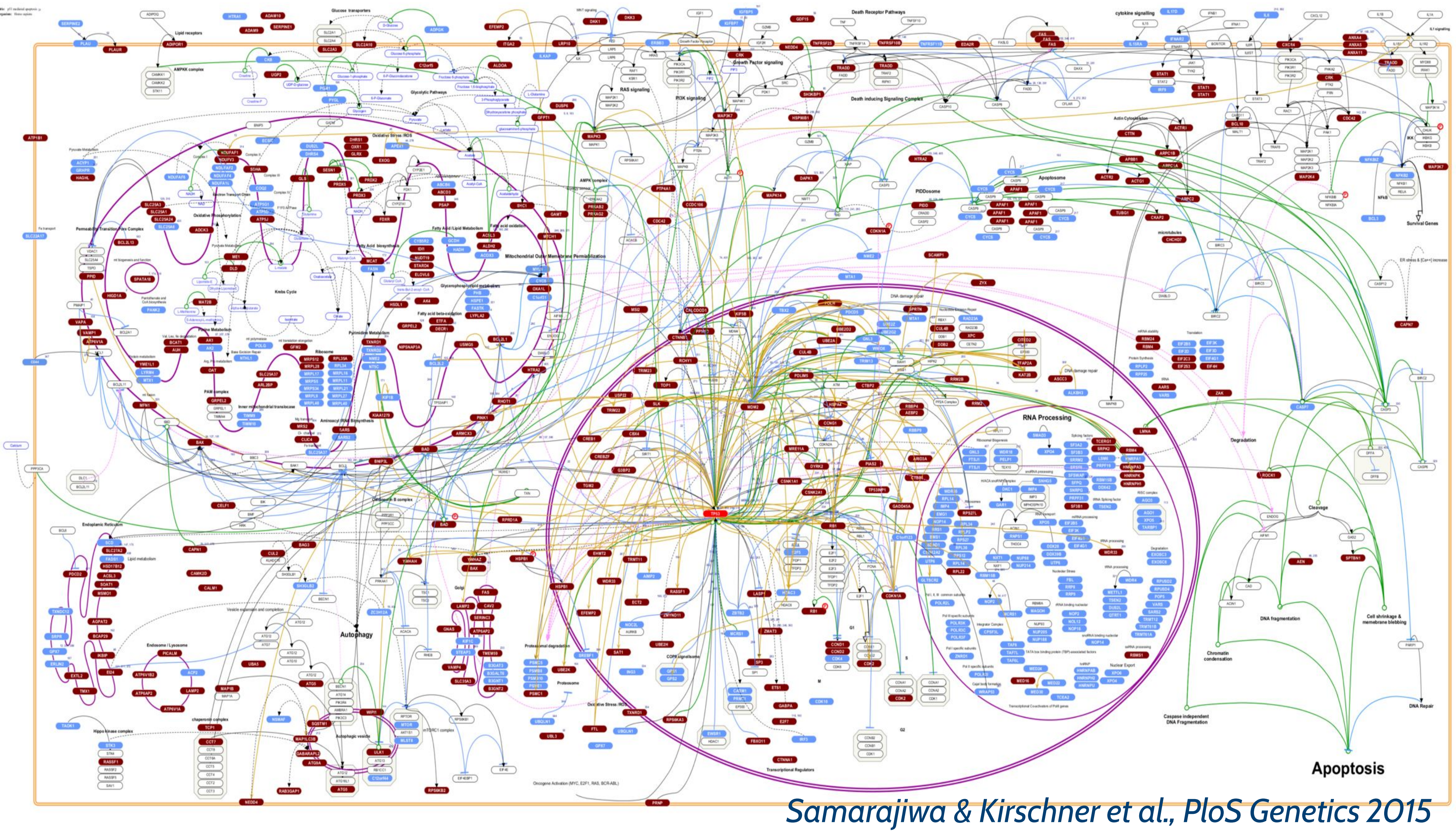
Functional Association Networks

Semi Supervised Network Generation and Topology analysis



Kirschner & Samarajiwa et al., PloS Genetics 2015

The TP53 Regulome



Samarajiwa & Kirschner et al., PloS Genetics 2015

References