University of Massachusetts Medical School

Integrated Analysis Of ChIP-seq/chip using ChIPpeakAnno, GeneNetworkBuilder and TrackViewer

Bioconductor Annual Meeting
Boston
July 28th 2017

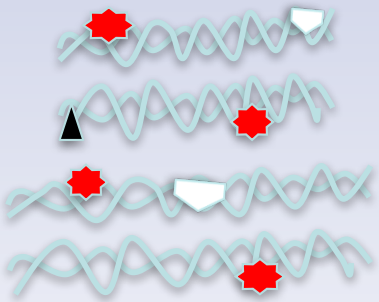
Jianhong Ou, Jun Yu, Haibo Liu and Lihua Julie Zhu

Outline

- Introduction of ChIP-seq and ChIP-chip analysis workflow
- ChIPpeakAnno
- GeneNetworkBuilder
- TrackViewer
- Demo

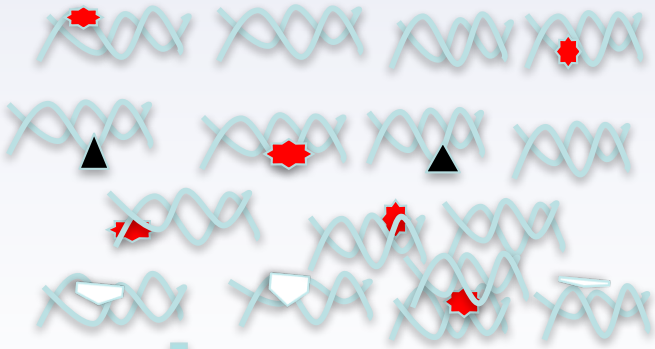
High-throughput Identification of DNA Binding Sites

- ChIP-seq
  - ChIP followed by high-throughput sequencing
- ChIP-chip
  - ChIP followed by genome tiling array analysis

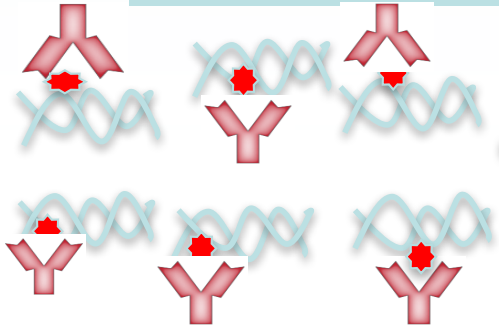


Cross-link with formaldehyde

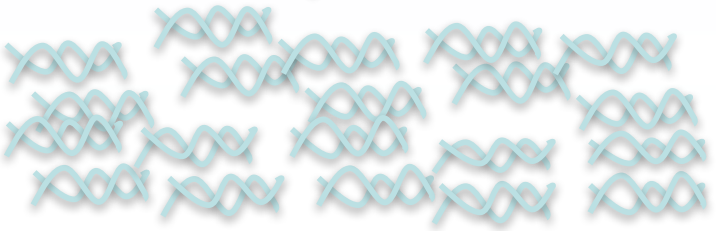
Fragment DNA



Add specific antibody to immunoprecipitate



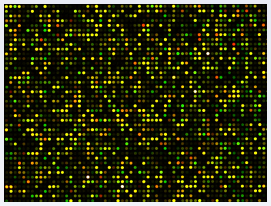
Reverse cross-link, purify and amplify DNA



High- throughput sequencing

Fastq sequence file

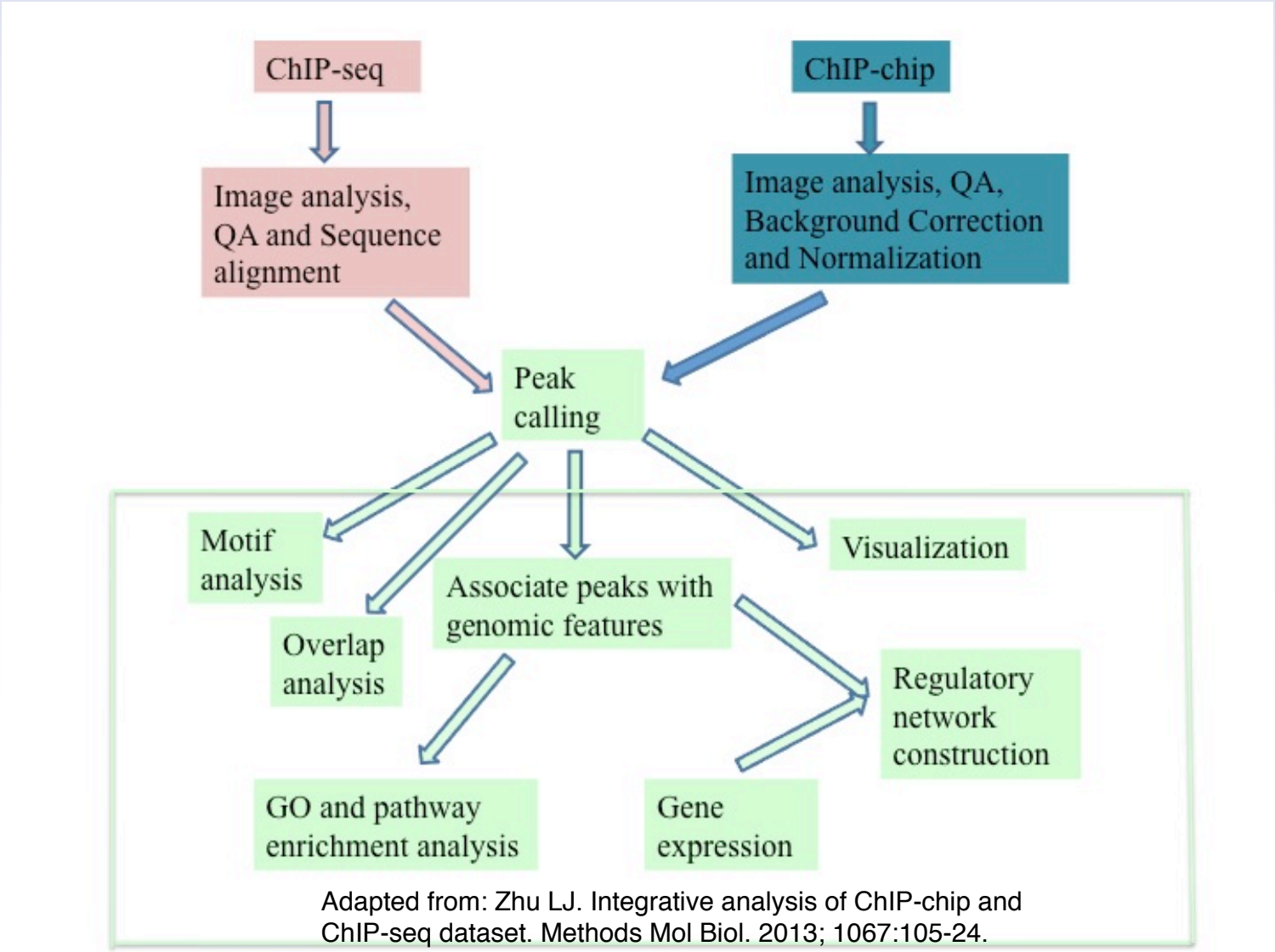
@HWI-ST570:42:D0PJJACXX:4:1101:1436:2323 1:N:0:
CATGGATCGGAAGAGGGAANTCATCTTTGGCCCGGTGTTTCGTCCTTTCC
+
CCCFFFFFHHHHHHHHIJGG#2AEGHIGJIJJJJJJ?FFHJGIGHIIIJIJ



Hybridize to DNA microarray post DNA labeling

Adapted from: Zhu LJ. Integrative analysis of ChIP-chip and ChIP-seq dataset. Methods Mol Biol. 2013; 1067:105-24.

ANALYSIS WORKFLOW



Adapted from: Zhu LJ. Integrative analysis of ChIP-chip and ChIP-seq dataset. Methods Mol Biol. 2013; 1067:105-24.

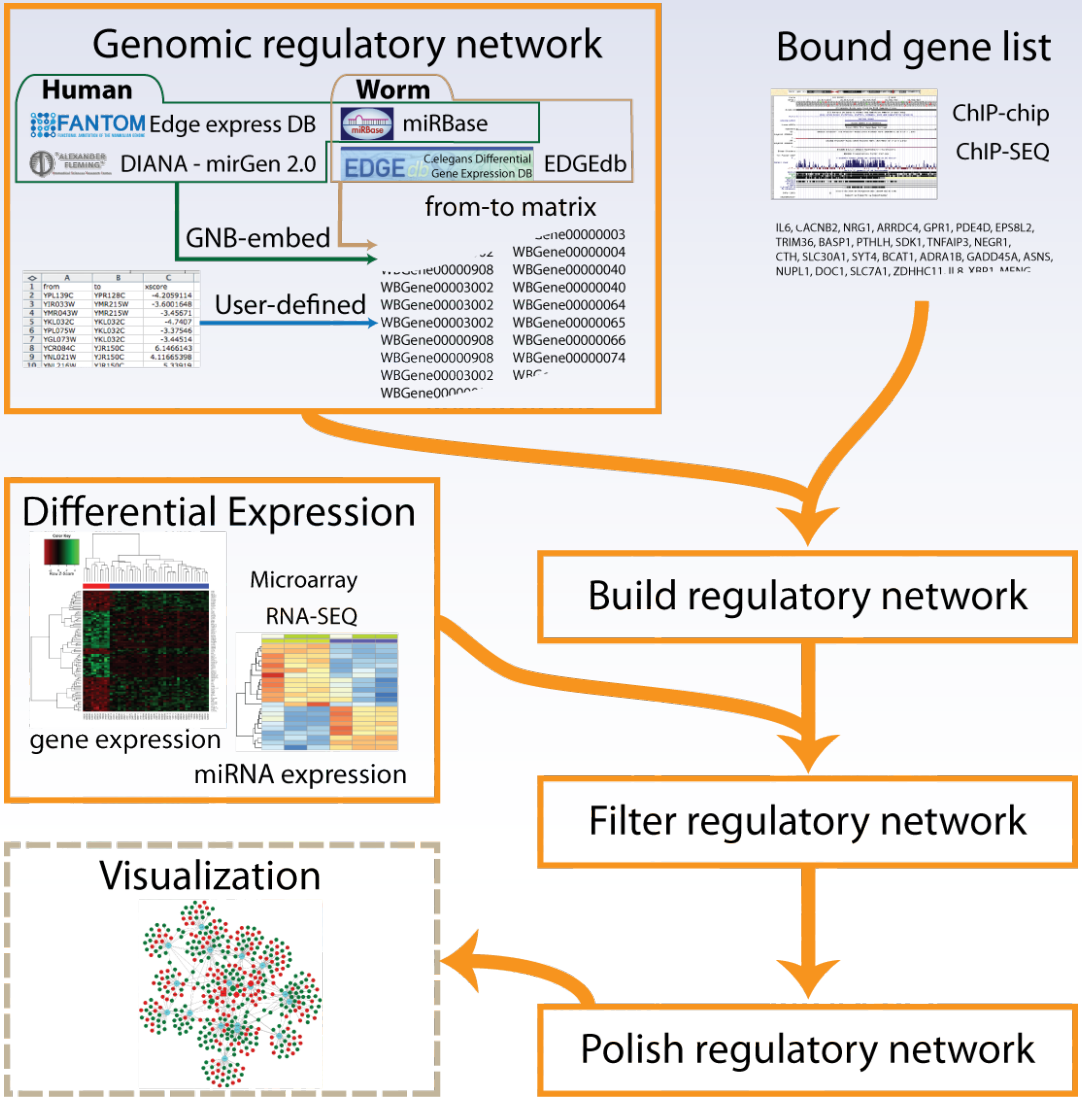
ChIPpeakAnno

- Batch annotate enriched peaks
  - ChIP-seq
  - ChIP-chip
  - PAS-seq (Poly(A) Site Sequencing)
  - Cap Analysis of Gene Expression (CAGE)
  - Any experiments resulting in a large number of enriched genomic regions

FUNCTIONALITY

- Find the nearest genes for each set of peaks and graph the distribution around features.
- Find all genes within a certain distance from the peaks
- Identify enriched Gene Ontology (GO) terms and pathways associated with adjacent genes of the peaks.
- Label peaks with any annotation of interest
  - a dataset from the literature
  - CpG island
  - conserved element
  - histone modification marks
- Determine the significance of overlap and drawing Venn diagrams to visualize the extent of the overlap
  - binding sites among replicates
  - binding sites among transcription factors within a complex
  - binding sites among different experiments such as yours and the ones in literature
- Retrieve genomic sequences flanking putative binding sites for motif discovery, cloning or PCR amplification
- Find the peaks with bi-directional promoters with summary statistics
- Summarize motif occurrence in peaks
- Irreproducibility Discovery **Rate** (IDR)

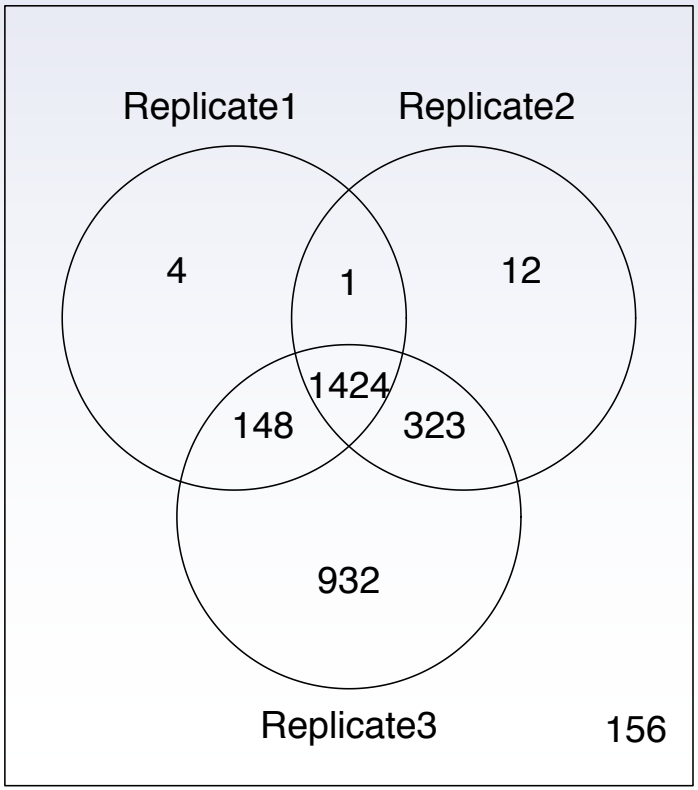
GeneNetWorkBuilder

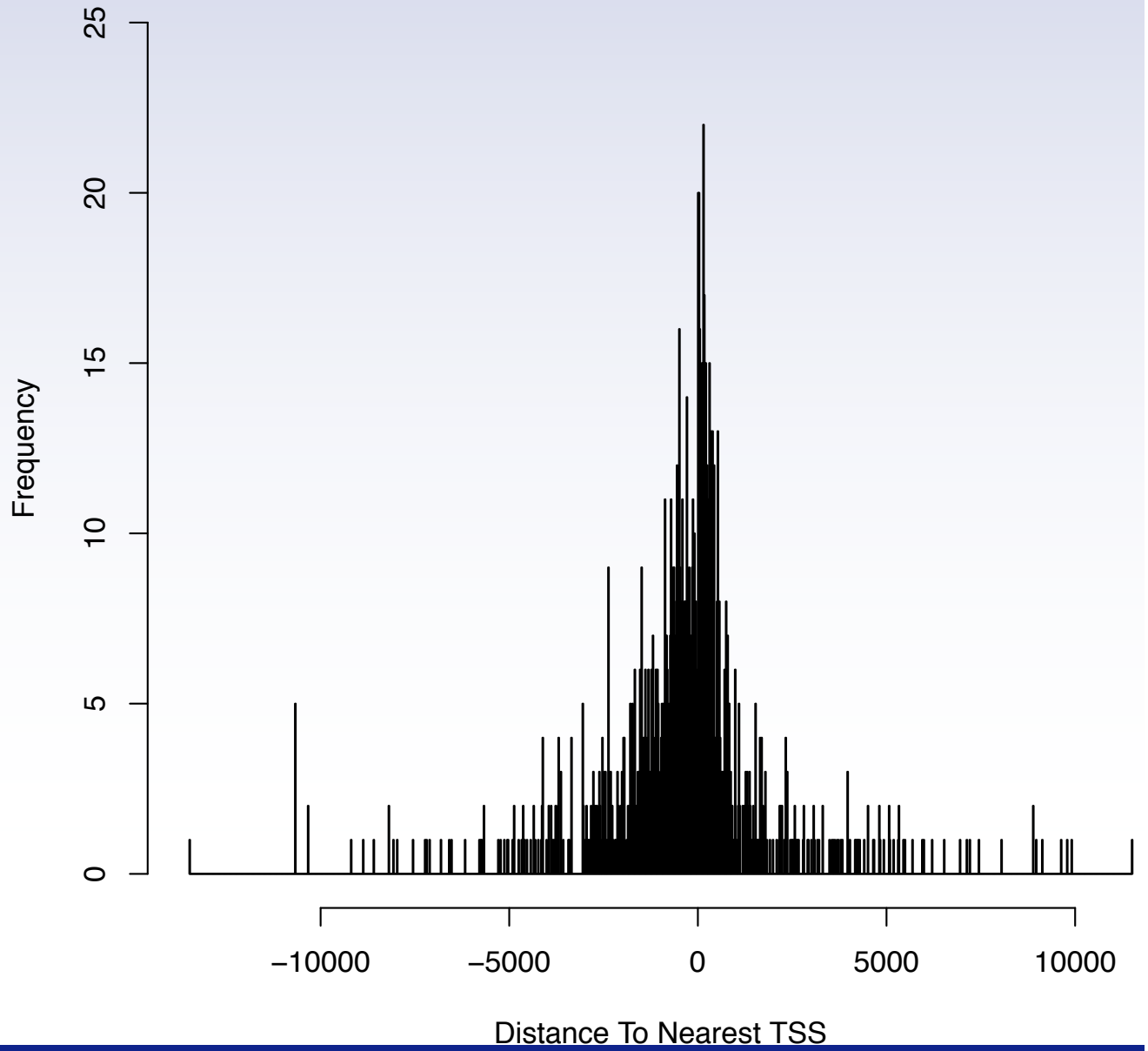


DAF-12 Example Dataset

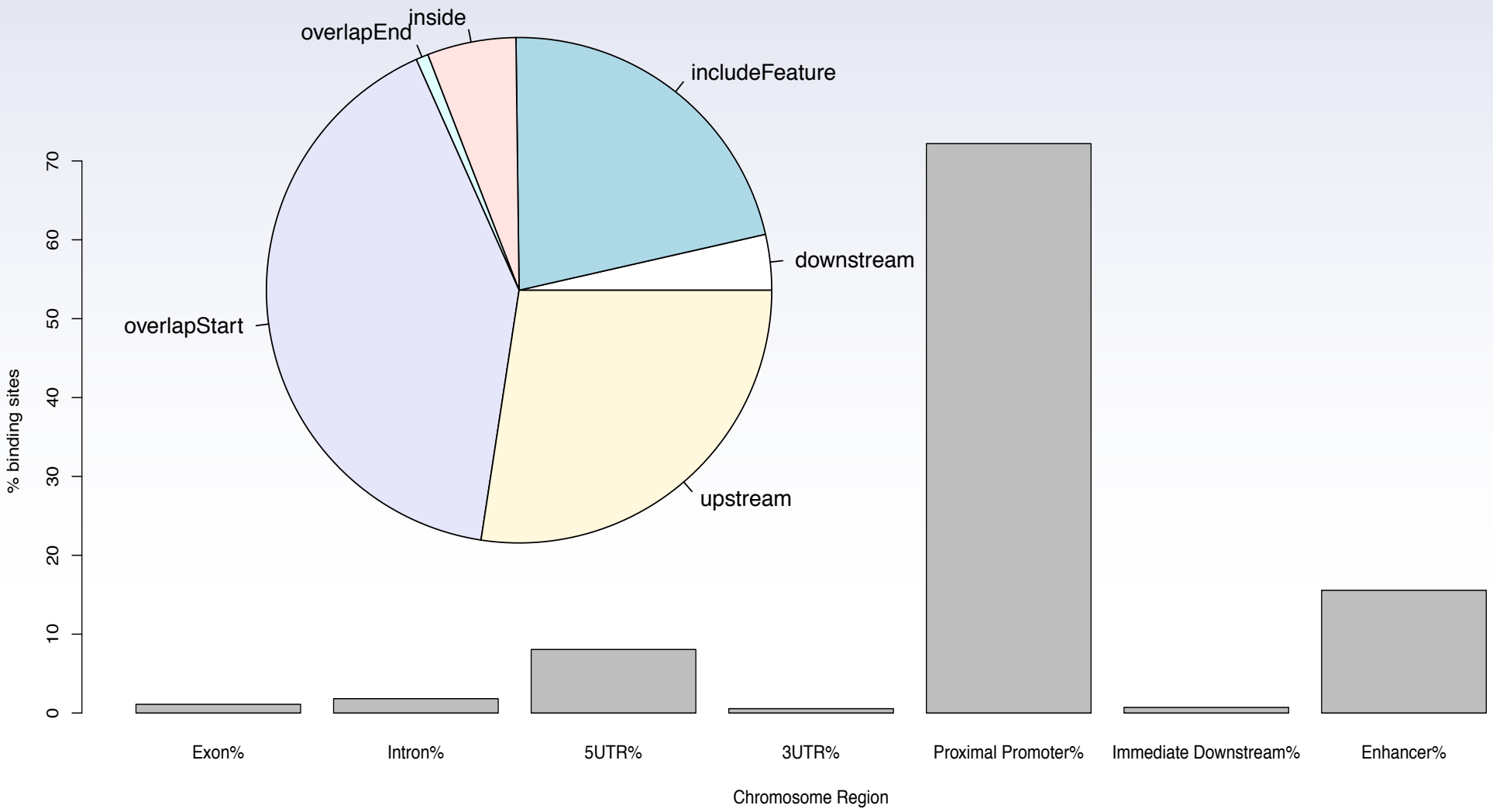
- ChIP-chip peaks were downloaded from GEO at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE28350 (Hochbaum, Zhang et al. 2011, PLoS Genet **7**(7): e1002179)
- Expression Microarray results were downloaded from (Fisher and Lithgow 2006, Aging Cell **5**(2): 127-138).

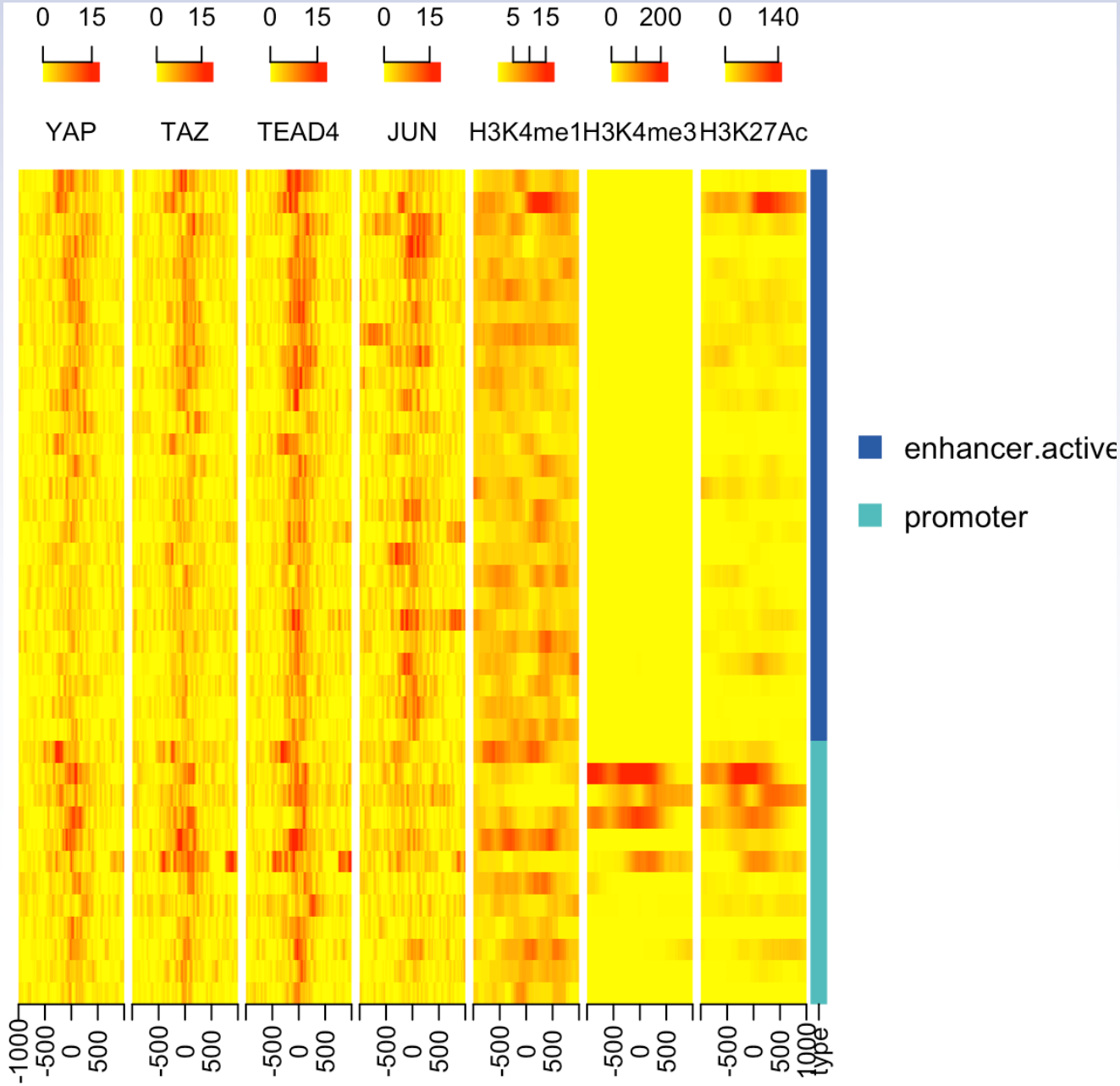
Overlap Analysis and distribution of peaks around TSS





DISTRIBUTION OF DAF-12-BINDING SITES




0 15
0 15
0 15
0 15
5 15
0 200
0 140
YAP
TAZ
TEAD4
JUN
H3K4me1
H3K4me3
H3K27Ac
-1000
-500
0
500
type
enhancer.active
promoter

Top Motifs

Logo

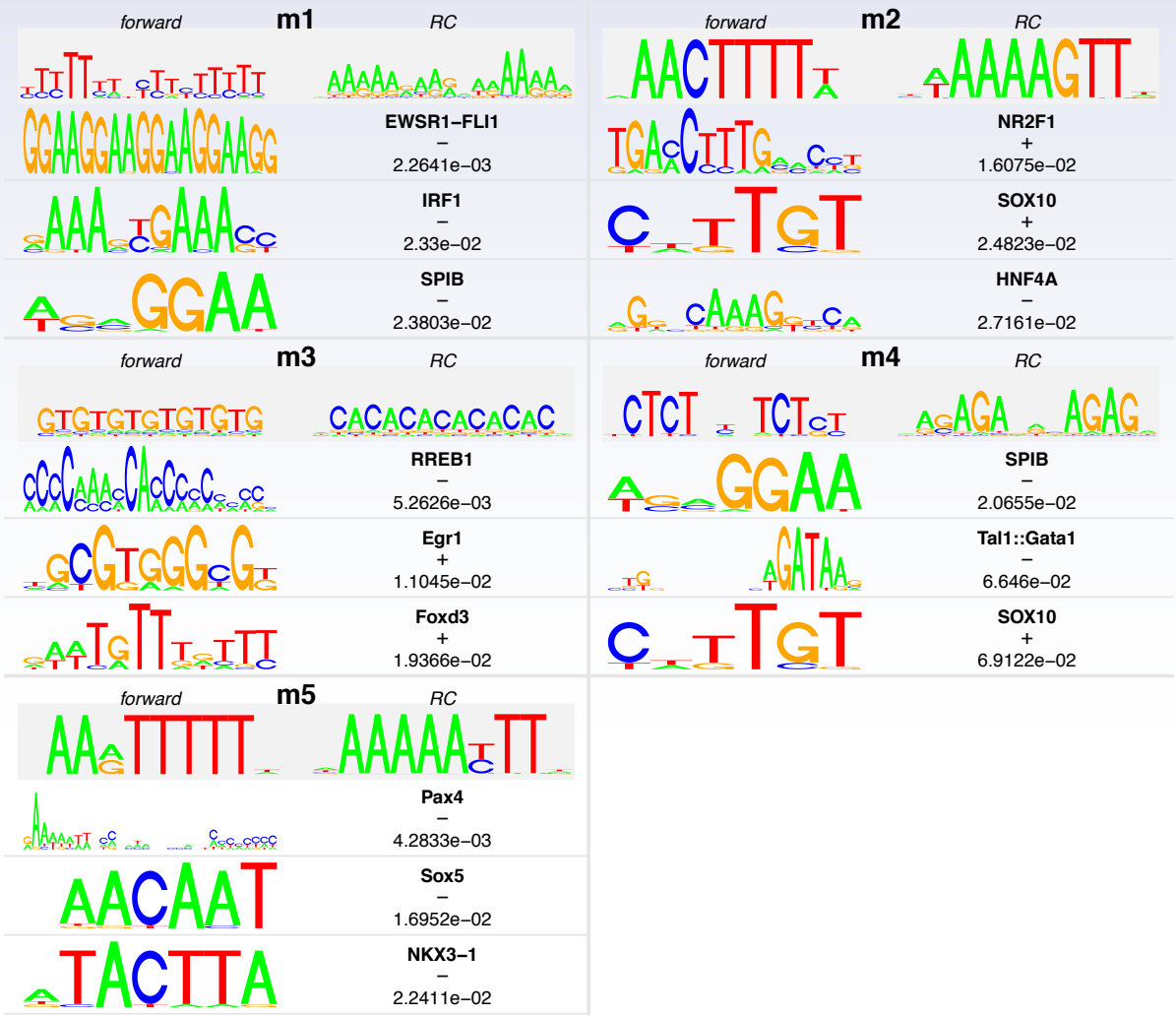


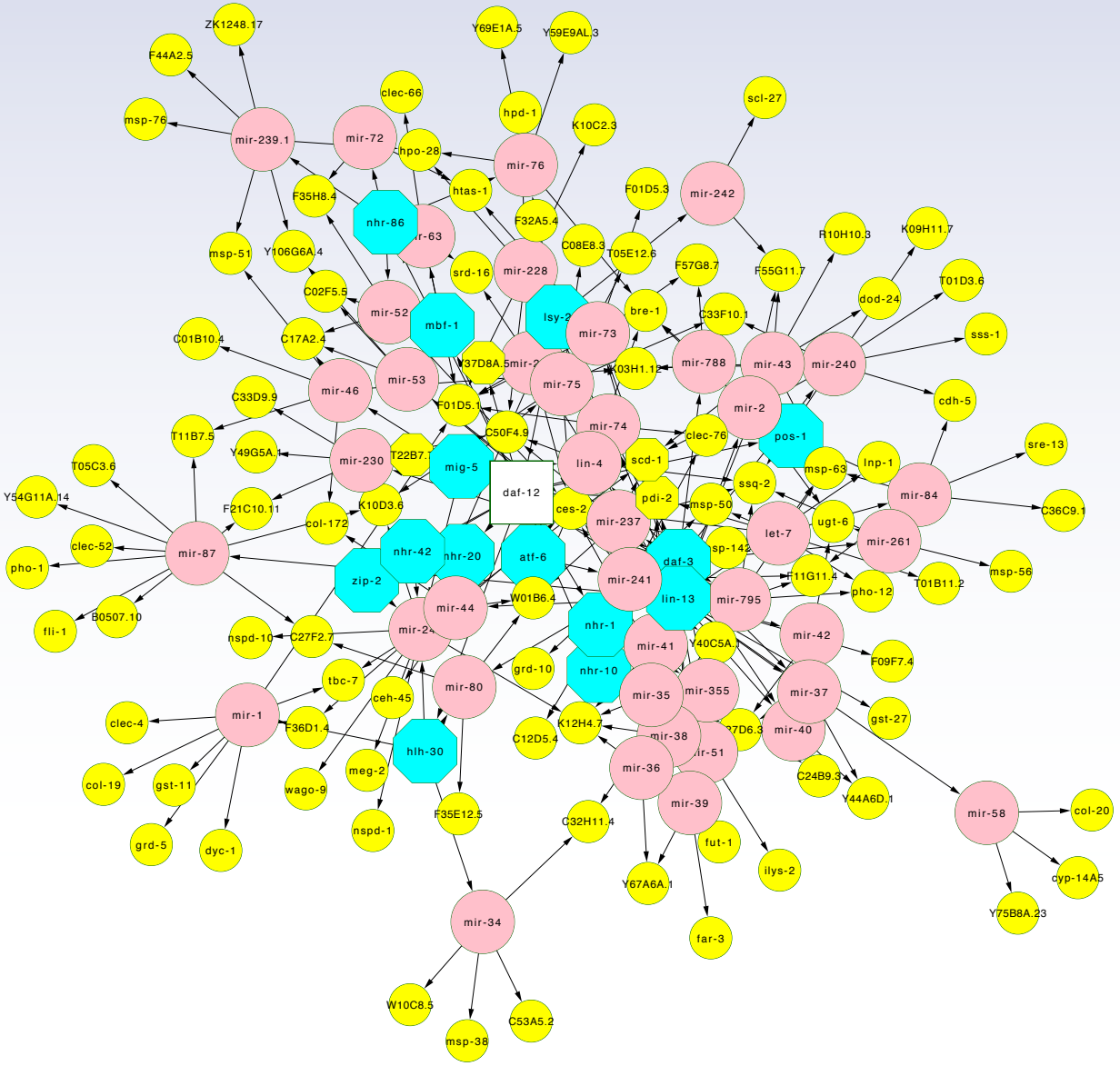
Table 2. Enriched GO molecular functions and biological processes of DAF-12-binding sites in worm ordered by false discovery rate (FDR).

GO ID	GO Term	GO Definition	Category	FDR
GO:0005515	protein binding	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	MF	4.85E-08
GO:0003735	structural constituent of ribosome	The action of a molecule that contributes to the structural integrity of the ribosome.	MF	0.0002
GO:0002119	nematode larval development	The process whose specific outcome is the progression of the nematode larva over time, from its formation to the mature structure. Nematode larval development begins with the newly hatched first-stage larva (L1) and ends with the end of the last larval stage (for example the fourth larval stage (L4) in C. elegans). Each stage of nematode larval development is characterized by proliferation of specific cell lineages and an increase in body size without alteration of the basic body plan. Nematode larval stages are separated by molts in which each stage-specific exoskeleton, or cuticle, is shed and replaced anew.	BP	0.0013
GO:0002164	larval development	The process whose specific outcome is the progression of the larva over time, from its formation to the mature structure. The larva is the early, immature form of an	BP	0.0013

Table 3. Enriched pathways in reactome database for DAF-12 binding sites in worm ordered by FDR.

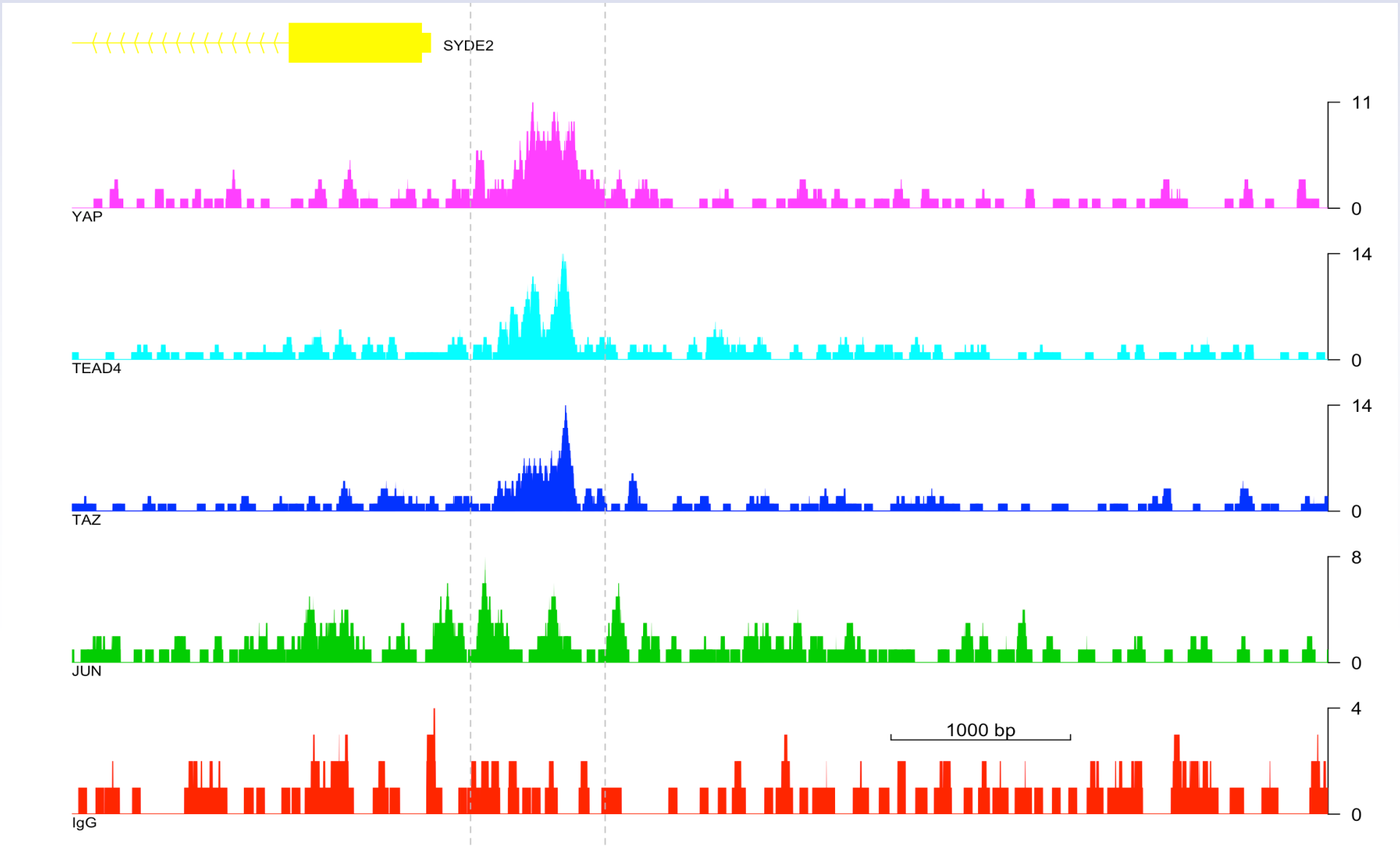
Pathway ID	Pathway Definition	FDR
1626134	Caenorhabditis elegans: Regulation of gene expression in beta cells	0.004
1626136	Caenorhabditis elegans: Diabetes pathways	0.004
1625991	Caenorhabditis elegans: Peptide chain elongation	0.004
1625992	Caenorhabditis elegans: Eukaryotic Translation Elongation	0.004
1625772	Caenorhabditis elegans: GTP hydrolysis and joining of the 60S ribosomal subunit	0.006
1626131	Caenorhabditis elegans: Regulation of beta-cell development	0.006
1625983	Caenorhabditis elegans: Eukaryotic Translation Termination	0.007
1626135	Caenorhabditis elegans: Insulin Synthesis and Processing	0.023
1625880	Caenorhabditis elegans: Developmental Biology	0.032
1625773	Caenorhabditis elegans: Formation of a pool of free 40S subunits	0.036

DAF-12 Regulatory Network



University of Massachusetts Medical School

TrackViewer



REFERENCES

- Zhu LJ, Gazin C, Lawson ND, Pagès H, Lin SM, Lapointe DS, Green MR. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics. 2010 May 11; 11:237. PMID: 20459804.
- Zhu LJ. Integrative analysis of ChIP-chip and ChIP-seq dataset. Methods Mol Biol. 2013; 1067:105-24. PMID: 23975789.