University of
Massachusetts
UMASS. Medical School

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

Lihua Julie Zhu

Bioconductor Developer Meeting
Zurich, Switzerland
December 13th -14th 2012

Outline

- Introduction of ChIP-seq and ChIP-chip analysis workflow
- ChIPpeakAnno
- GeneNetworkBuilder
- Analysis of DAF-12 ChIP-chip and Expression Dataset

HIGH-THROUGHPUT IDENTIFICATION OF DNA BINDING SITES

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

ANALYSIS WORKFLOW

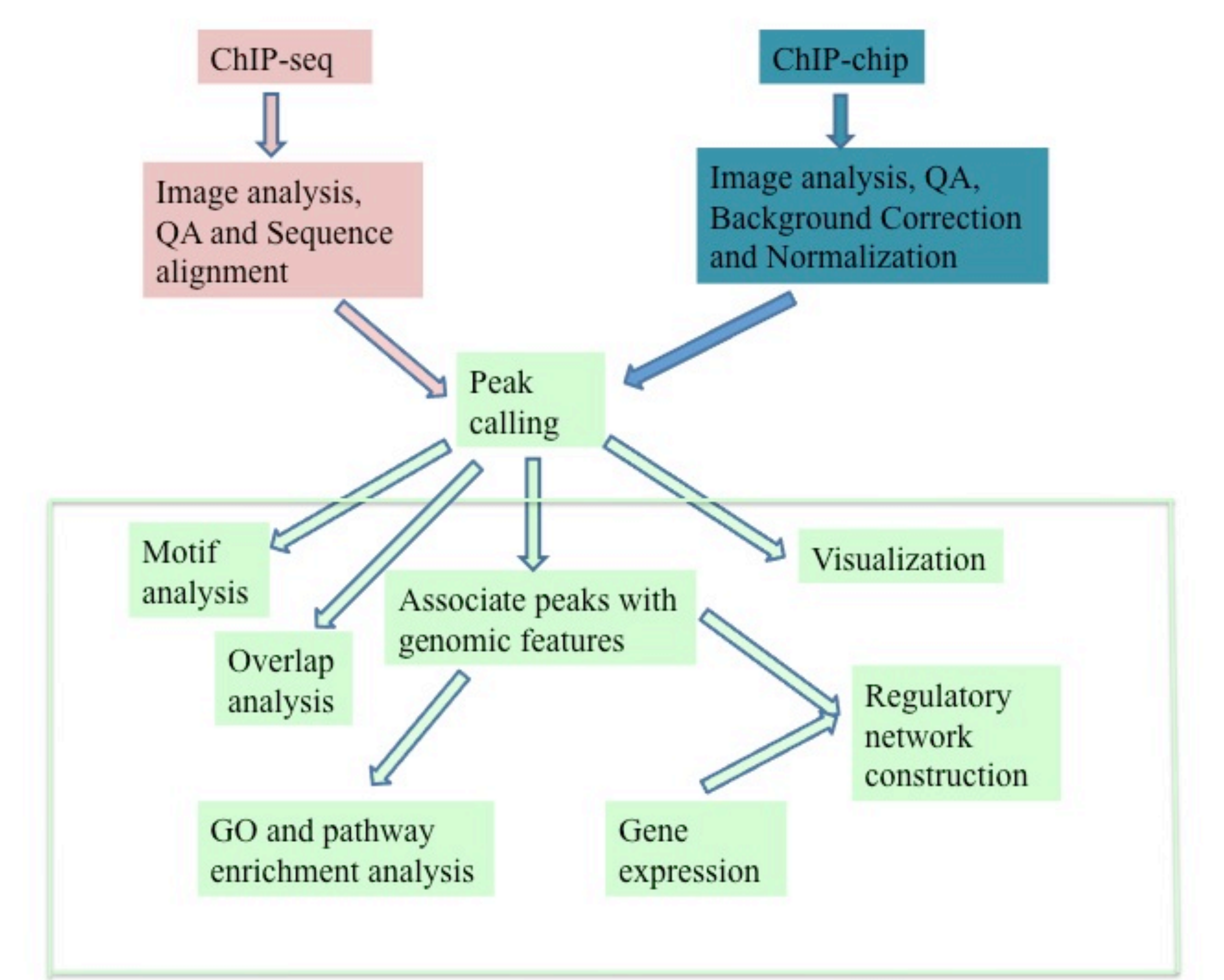

ChIP-seq

Image analysis, QA and Sequence alignment

ChIP-chip

Image analysis, QA, Background Correction and Normalization

Peak calling

Motif analysis

Overlap analysis

Associate peaks with genomic features

Visualization

Regulatory network construction

GO and pathway enrichment analysis

Gene expression

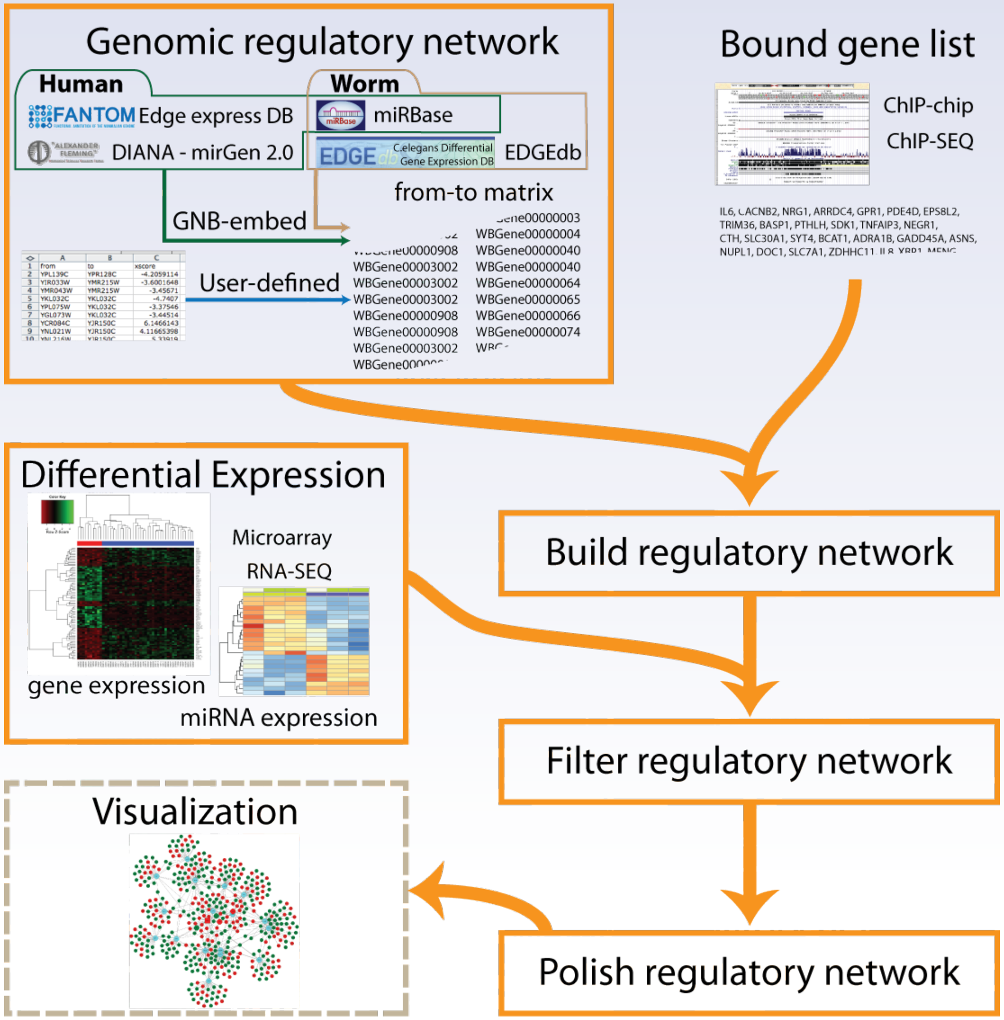
CHIPPEAKANNO

- 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

CHIPPEAKANNO

- 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

GENENETWORKBUILDER



Genomic regulatory network

Human

FANTOM Edge express DB

DIANA - mirGen 2.0

Worm

miRBase

EDGEdb C.elegans Differential Gene Expression DB EDGEdb

GNB-embed

User-defined

	A	B	C
1	from	to	xscore
2	YPL139C	YPR128C	-4.2059114
3	YIR033W	YMR215W	-3.6001648
4	YMR043W	YMR215W	-3.45671
5	YKL032C	YKL032C	-4.7407
6	YPL075W	YKL032C	-3.37546
7	YGL073W	YKL032C	-3.44514
8	YCR084C	YJR150C	6.1466143
9	YNL021W	YJR150C	4.11665398
10	YNL216W	YJR150C	5.33919

from-to matrix

ene00000003
02 WBGene00000004
ene00000908 WBGene00000040
WBGene00003002 WBGene00000040
WBGene00003002 WBGene00000064
WBGene00003002 WBGene00000065
WBGene00000908 WBGene00000066
WBGene00000908 WBGene00000074
WBGene00003002 WB
WBGene000

Bound gene list

ChIP-chip

ChIP-SEQ

IL6, CACNB2, NRG1, ARRDC4, GPR1, PDE4D, EPS8L2, TRIM36, BASP1, PTHLH, SDK1, TNFAIP3, NEGR1, CTH, SLC30A1, SYT4, BCAT1, ADRA1B, GADD45A, ASNS, NUPL1, DOC1, SLC7A1, ZDHHC11, IL8, YBP1, MENG

Differential Expression

Microarray

RNA-SEQ

gene expression

miRNA expression

Build regulatory network

Filter regulatory network

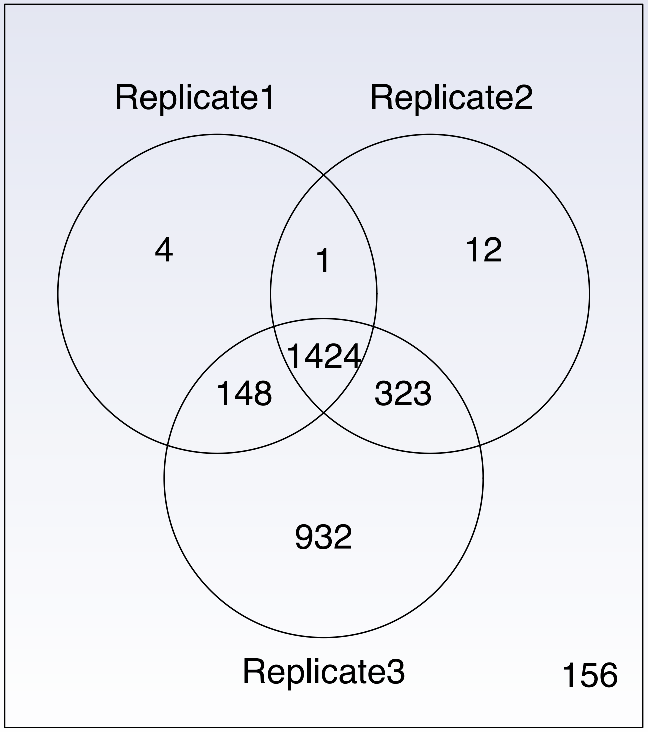
Polish regulatory network

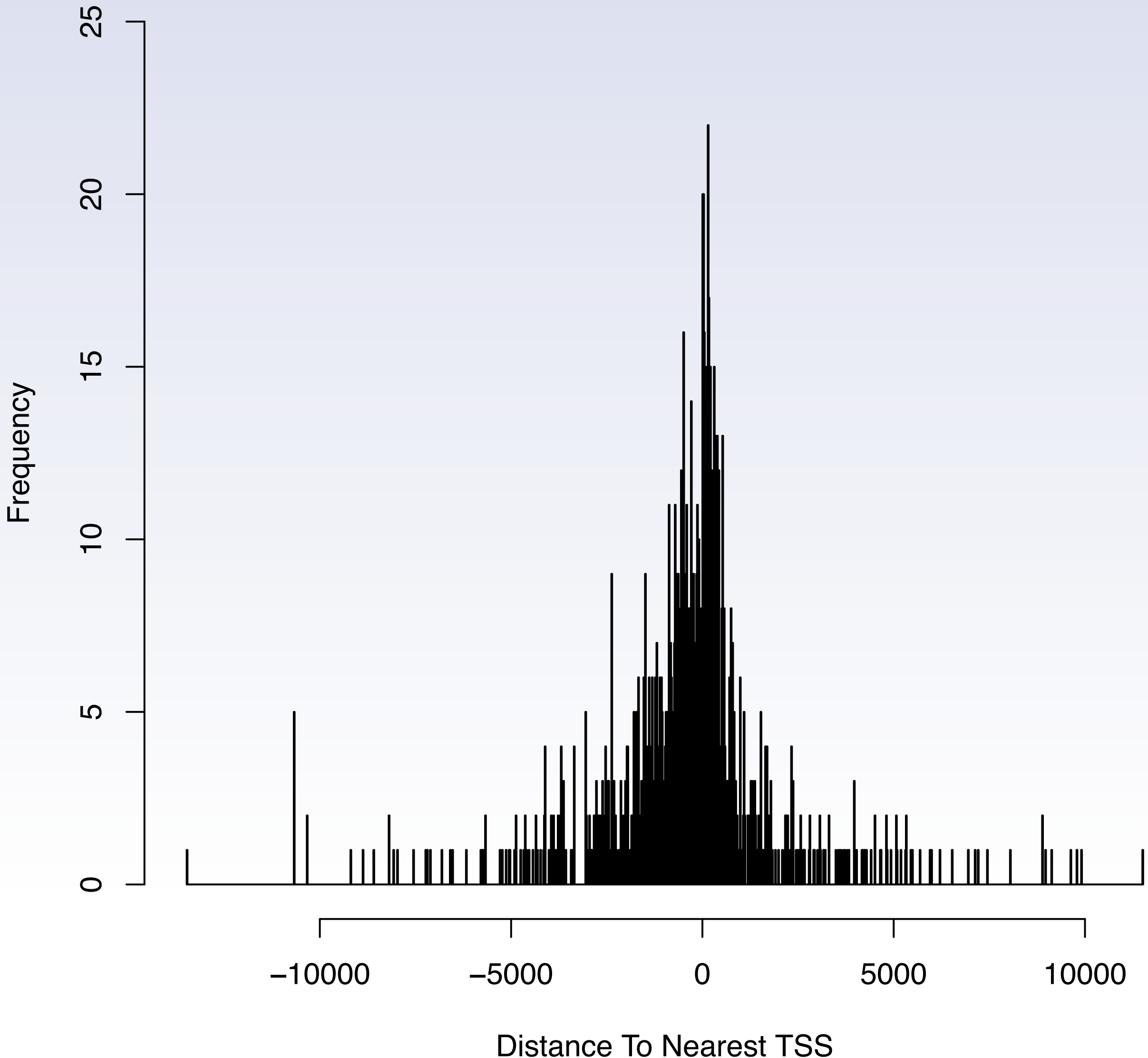
Visualization

DAF-12 EXAMPLE DATASET

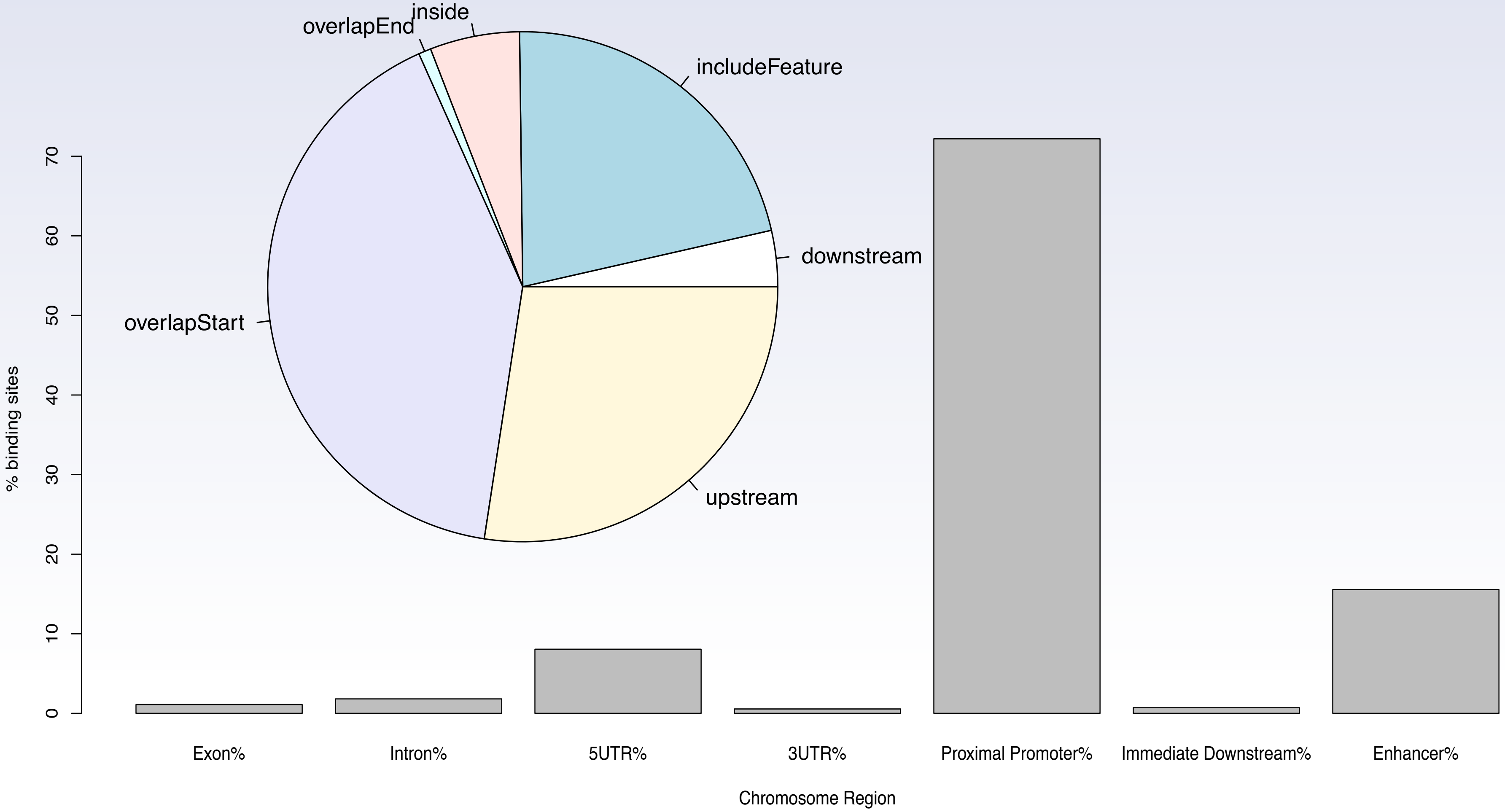
- ChIP-chip peaks were downloaded from GEO at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE28350 (Hochbaum, Zhang et al. 2011, PLoS Genet **7**(7): e1002179)
- Expression Microarray results were downloaded from (Fisher and Lithgow 2006, Aging Cell **5**(2): 127-138).

OVERLAP ANALYSIS AND DISTRIBUTION OF PEAKS AROUND TSS





DISTRIBUTION OF DAF-12-BINDING SITES



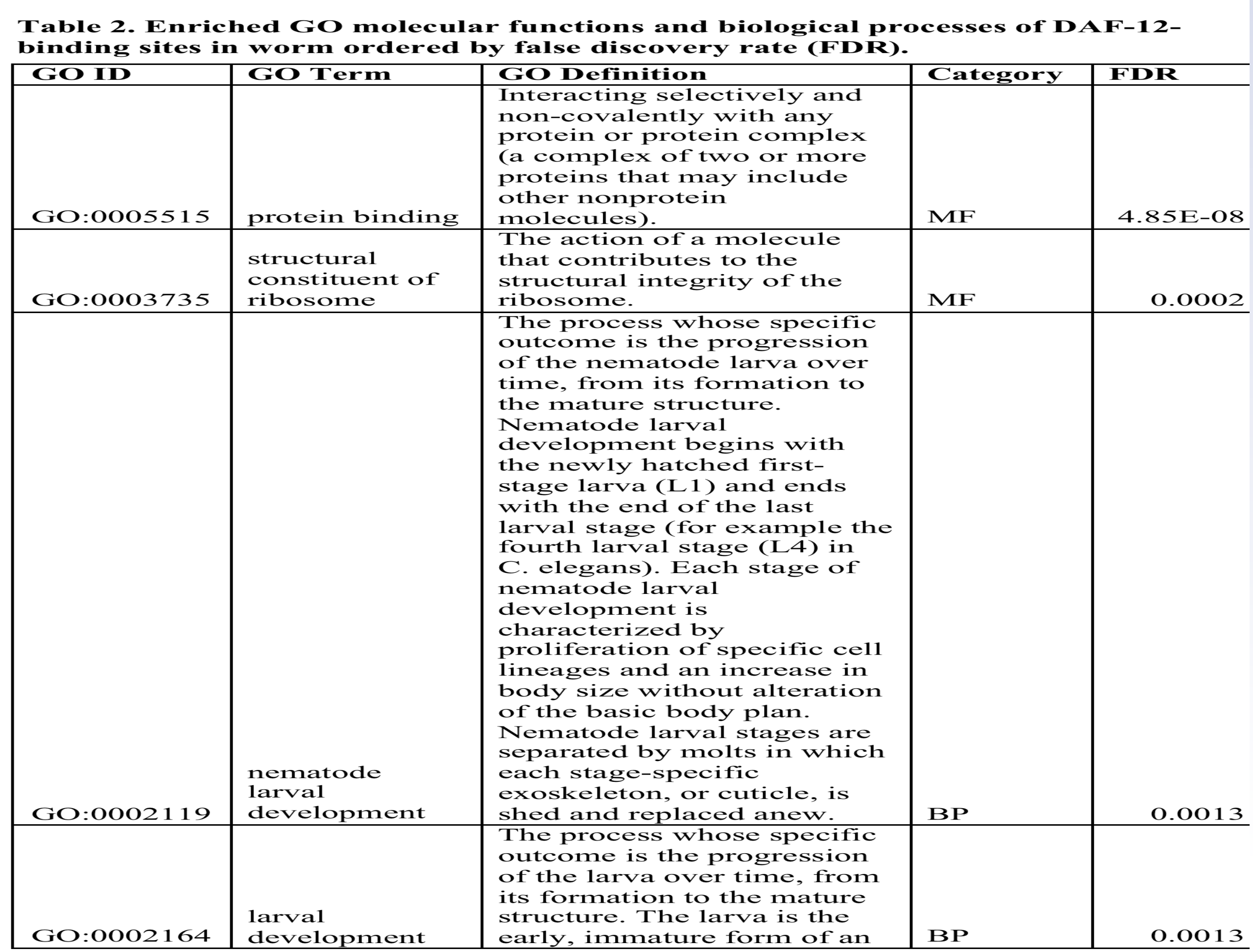


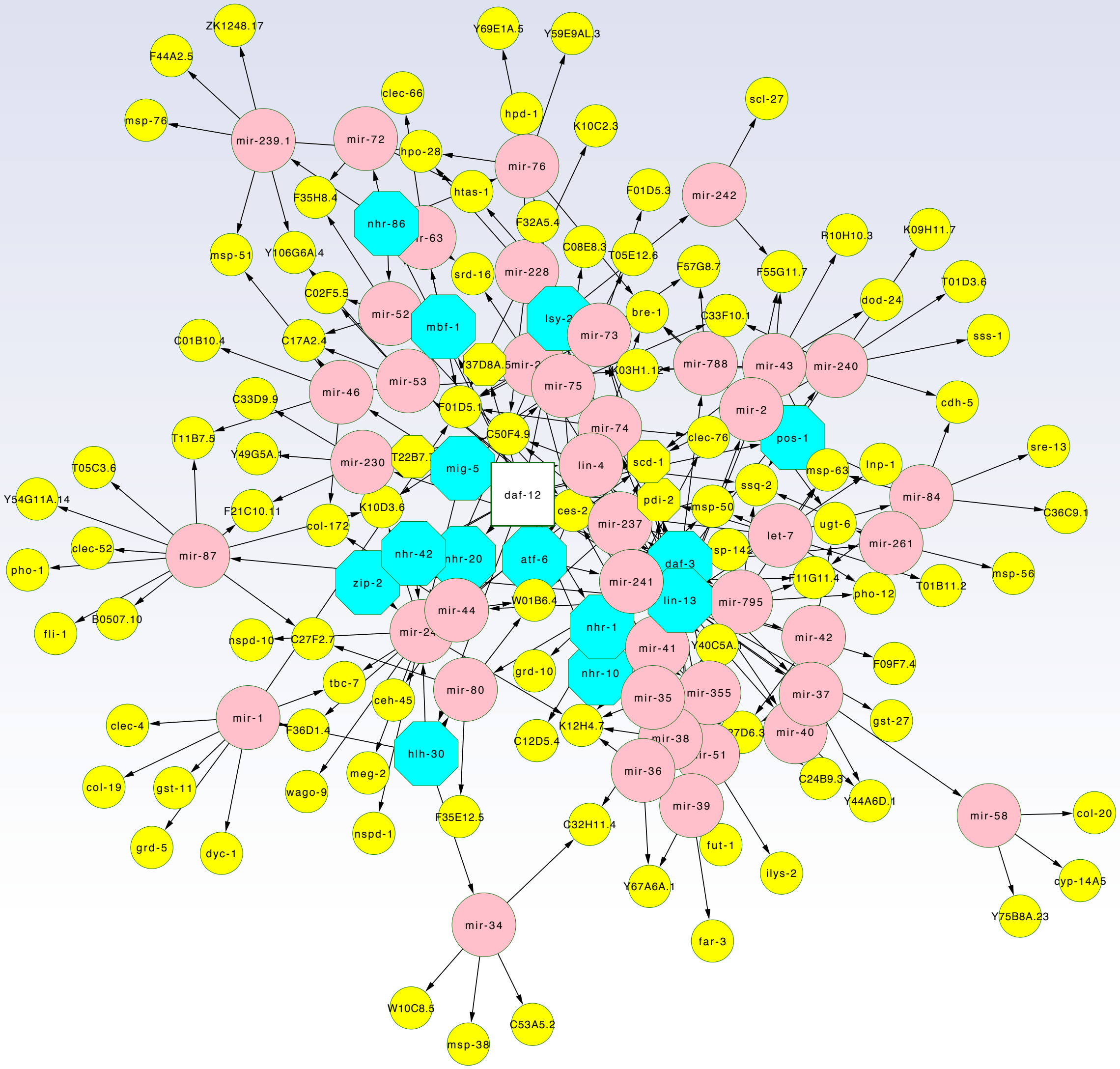
Table 2. Enriched GO molecular functions and biological processes of DAF-12-binding sites in worm ordered by false discovery rate (FDR).

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
UMASS. Medical School

SUMMARY

- Analysis of the DAF-12 example dataset shows that enriched GO terms and interaction pathways are consistent with the known functions of DAF-12.
- Network analysis, using GeneNetworkBuilder with ChIP data and expression data, generated a system-level view of the intertwined connections among the direct and indirect targets of DAF-12, which shows that DAF-12 is a master regulator.

ACKNOWLEDGEMENT

- Mark Robinson for the invitation and hospitality
- The Bioconductor package reviewers
  - Nishant Gopalak
  - Marc Carlson
  - Paul Shannon
- Coauthors
  - Jianhong Ou (Developer of GeneNetworkBuilder), Claude Gazin, Nathan Lawson, Hervé Pagès, Simon Lin, David Lapointe, Michael Green
- The users of the *ChIPpeakAnno*
- The Bioconductor core team, esp.,
  - Patrick Aboyoun
  - Vincent Carey
  - Martin Morgan
- Ivan Gregoretti
- Amy Molesworth
- Khademul Islam
- Hua Li
- Noah Dowell
- Yin Wu
- Zhiping Weng, Sara Evans , Alan Ritacco, Glenn Maston, Ping Wan, Ellen Kittler