# Insights from integrative analysis of the C. elegans genome:

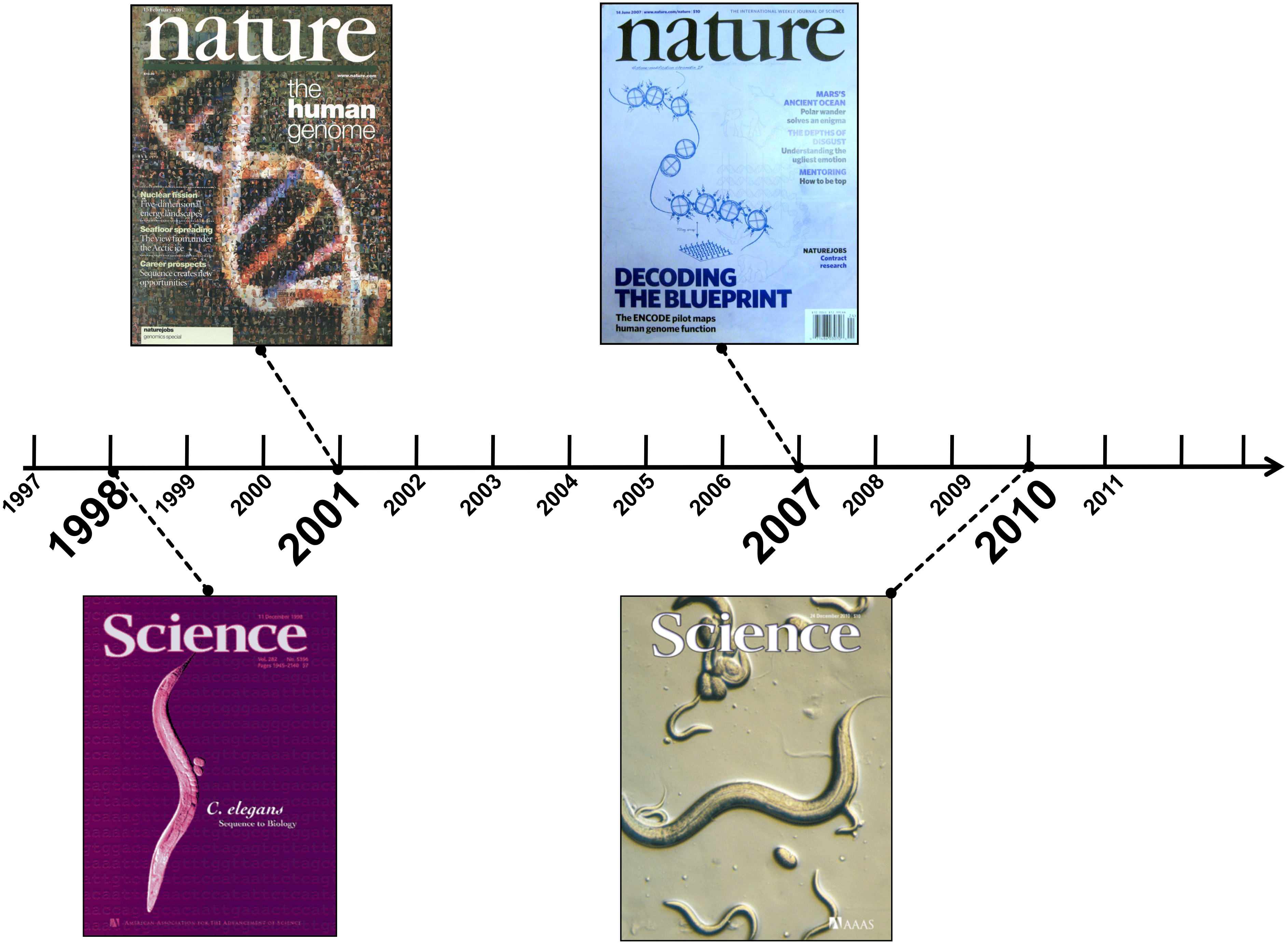
## What approaches we learned that were applicable to annotating the human genome

Mark B Gerstein
Yale

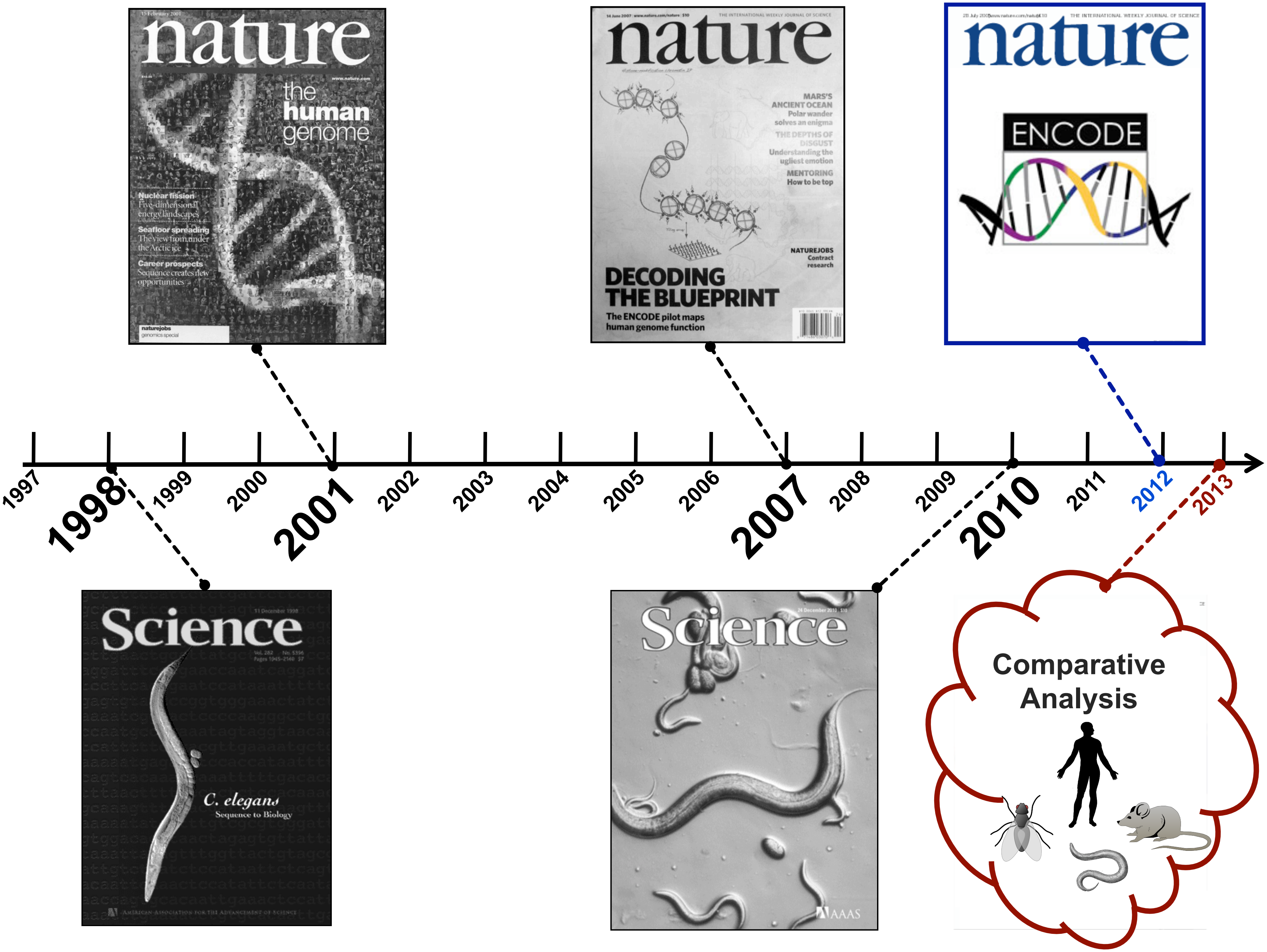
Slides at

Lectures.GersteinLab.org

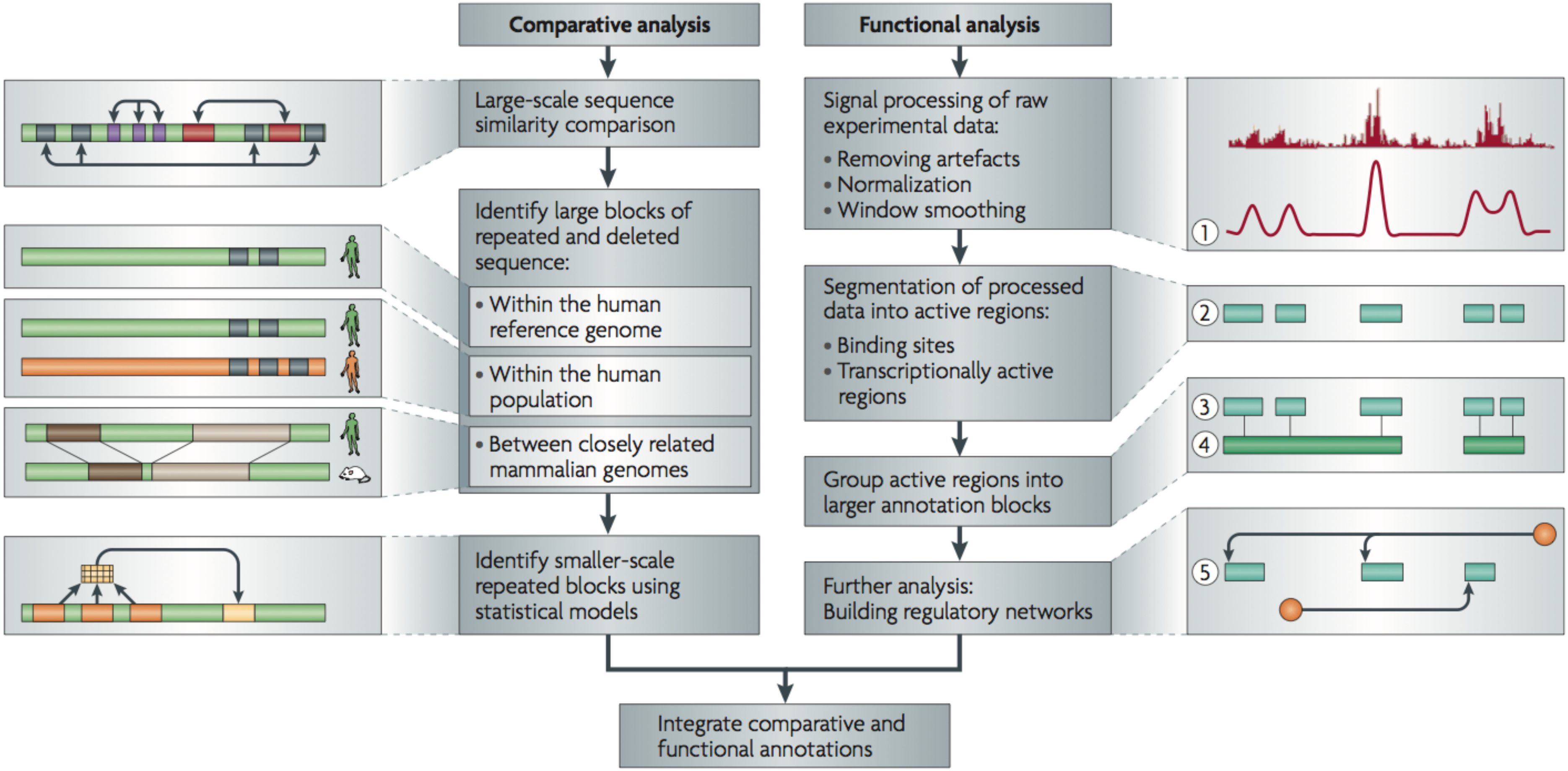
(See Last Slide for
References & More Info.)



1997 **1998** 1999 2000 **2001** 2002 2003 2004 2005 2006 **2007** 2008 2009 **2010** 2011


nature
the human genome
Nuclear fission
Five-dimensional energy landscapes
Seafloor spreading
The view from under the Arctic ice
Career prospects
Sequence creates new opportunities
naturejobs
genomics special
nature
MARS'S ANCIENT OCEAN
Polar wander solves an enigma
THE DEPTHS OF DISGUST
Understanding the ugliest emotion
MENTORING
How to be top
NATUREJOBS
Contract research
DECODING THE BLUEPRINT
The ENCODE pilot maps human genome function
nature
ENCODE
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
Science
C. elegans
Sequence to Biology
Science
AAAS
Comparative Analysis

# 2 Approaches to Genome Annotation



Comparative analysis

Large-scale sequence similarity comparison

Identify large blocks of repeated and deleted sequence:

- Within the human reference genome
- Within the human population
- Between closely related mammalian genomes

Identify smaller-scale repeated blocks using statistical models

Functional analysis

Signal processing of raw experimental data:

- Removing artefacts
- Normalization
- Window smoothing

Segmentation of processed data into active regions:

- Binding sites
- Transcriptionally active regions

Group active regions into larger annotation blocks

Further analysis: Building regulatory networks

Integrate comparative and functional annotations

[*Nat. Rev. Genet.* (2010) 11: 559]

# Importance of
# Dark Matter of the Genome

- Non-coding regions contain the control elements for coding regions.

- Some non-coding regions are functional & are pervasively transcribed.
- “Molecular Fossils” in the non-coding genome represent a historical record of the genome
- Most disease-associated mutations (e.g. GWAS hits) are in non-coding regions.

[Gravitational lensing by dark matter in Abell 1689 – HST (NASA, ESA)]

# Overview of the Data

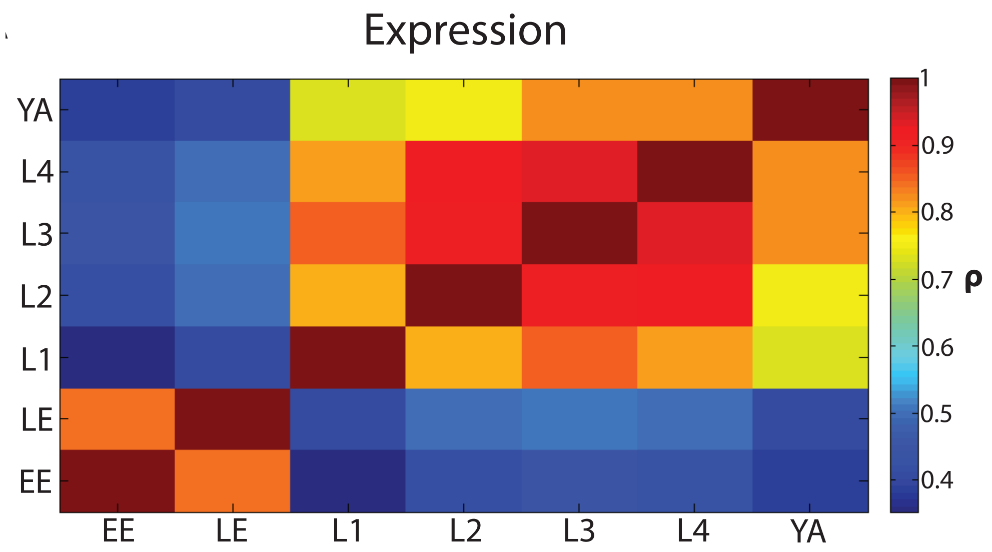
- **Worm**
  - Dev. Timecourse:
    E, L1, L2, L3, L4 + more
  - RNA-seq on timecourse + extra stages
    (polyA, small-RNA, 3' UTR selected)
  - Total RNA Tiling Arrays on timecourse + tissues
  - Chip-seq : 22 TFs + Pol2 in a variety of stages
  - Chromatin Chip-chip :
    >12 HMs mostly in EE & L3

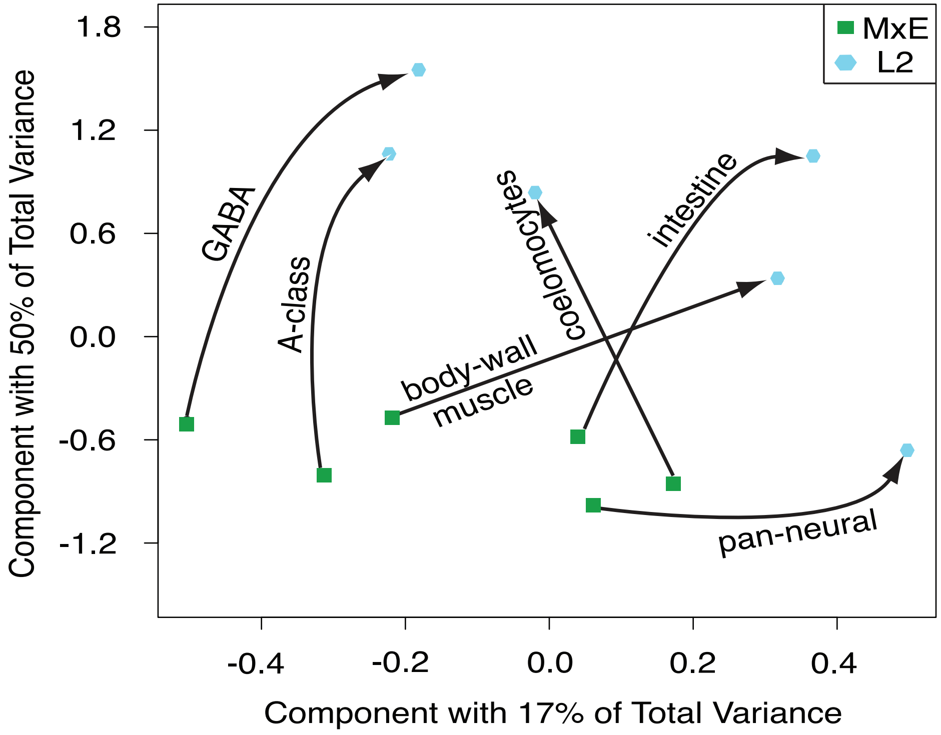
- **Human (very briefly!)**
  - ~200 tot. cell lines with lots on tier 1
    (GM12878, K562, H1)
  - ~120 TFs
  - deep RNA-seq
  - ~12 main HMs (chip-seq)

# Insights from worm modencode: Approaches useful for human annotation (outline)

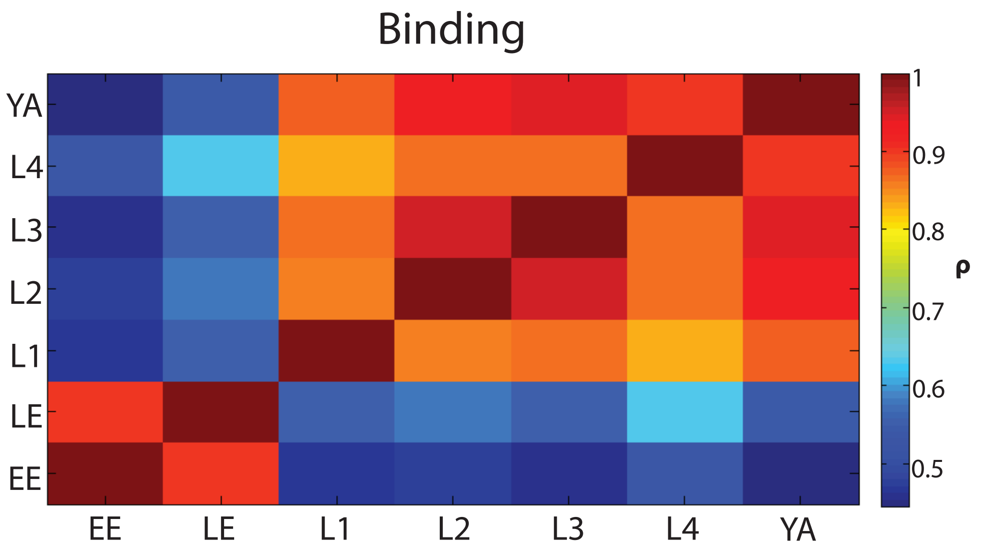
- **Expression Timecourse Analysis**
  - Coordinated binding & expression; E v L separation; ~280 large splicing changes
- **ncRNAs [Hum]**
  - Importance of evidence integration
  - Large numbers of transcribed pseudogenes (8-15%)
- **Chromosomal activity distribution [Hum]**
  - Most constrained regions active
  - Repressed arms & binding HOT spots.
- **Regulatory Net [Hum]**
  - Arranging TF binding into a hierarchy with differences betw. levels. Integration with miRNA regulation (more at top).
  - Network motifs & prevalence of FFLs
- **Stat Models relating HMs, TFs & Expression [Hum]**
  - HMs statistically predict expression for protein-coding genes and miRNAs
  - Similar results for TFs, highlighting predictive power of a few TFs.
  - Chromatin model (+ PWM) effective in predicting TF sites -- useful in identifying enhancers.

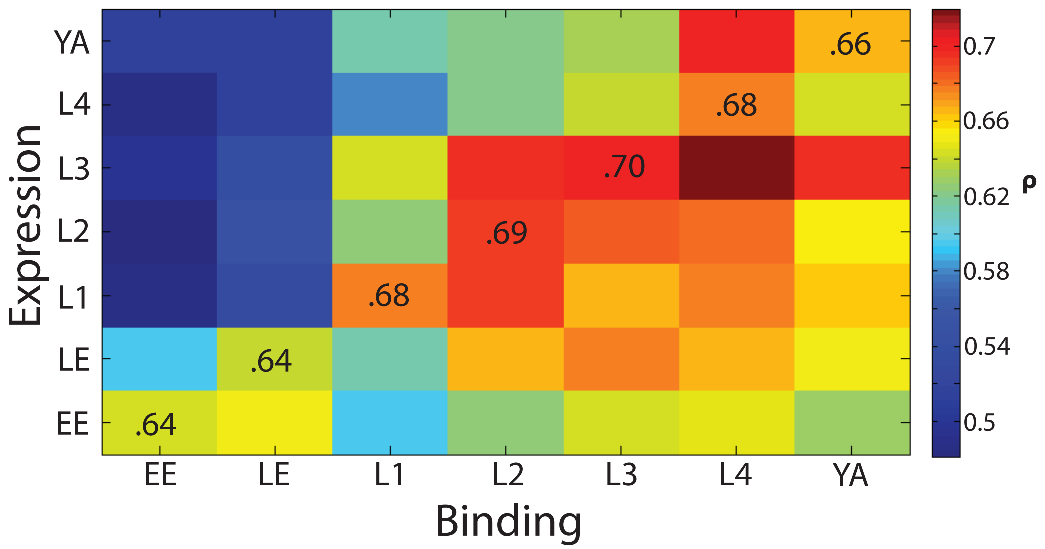
# Dynamics of Correlated Expression Changes Changes over Timecourse



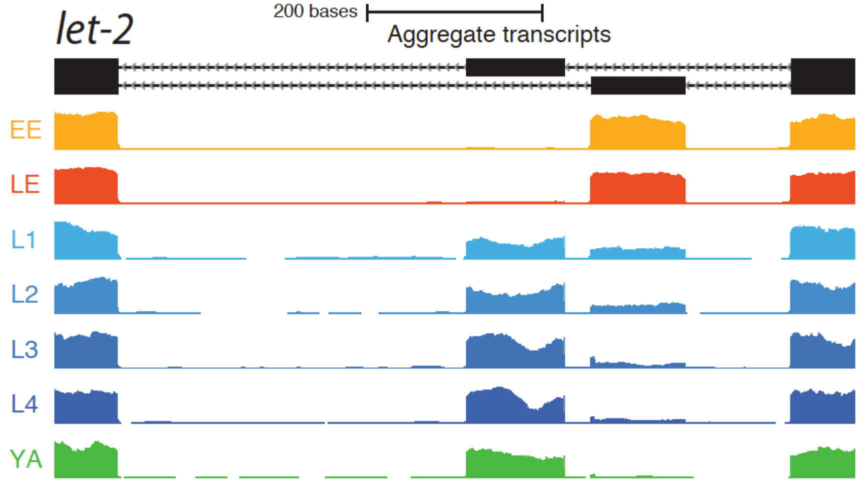


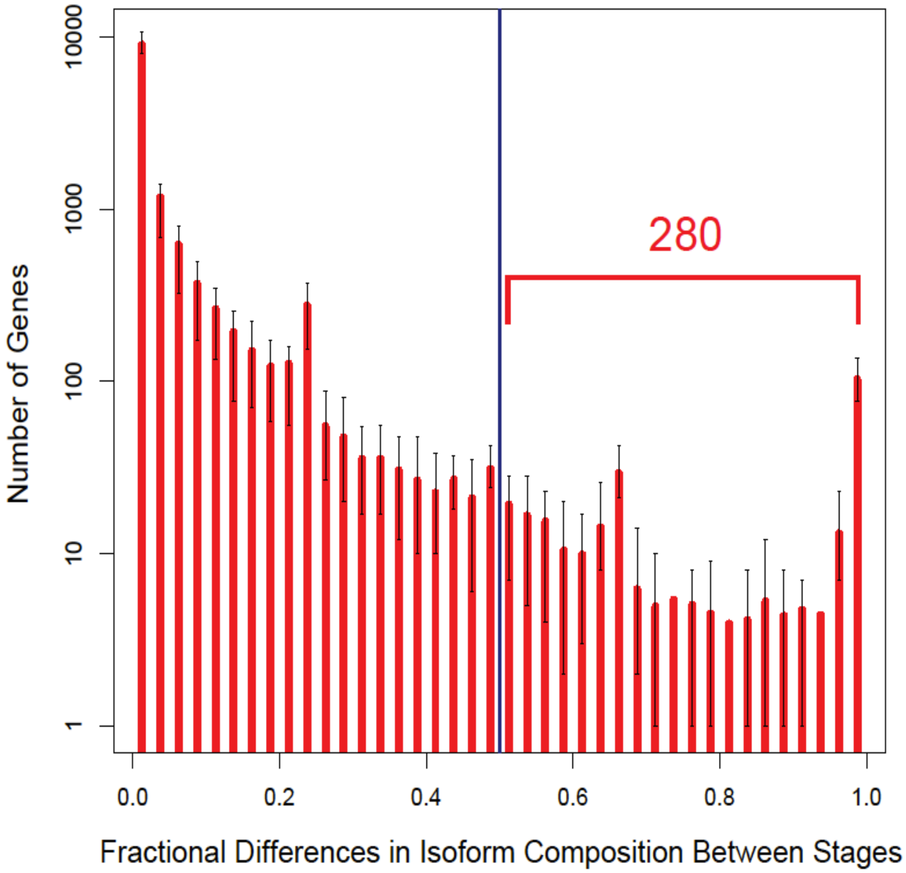
# Dynamics of Correlated Expression & Pol2 Binding Changes Changes over Timecourse





# Splicing Changes over the Timecourse (~280 changes/pair-of-stages)





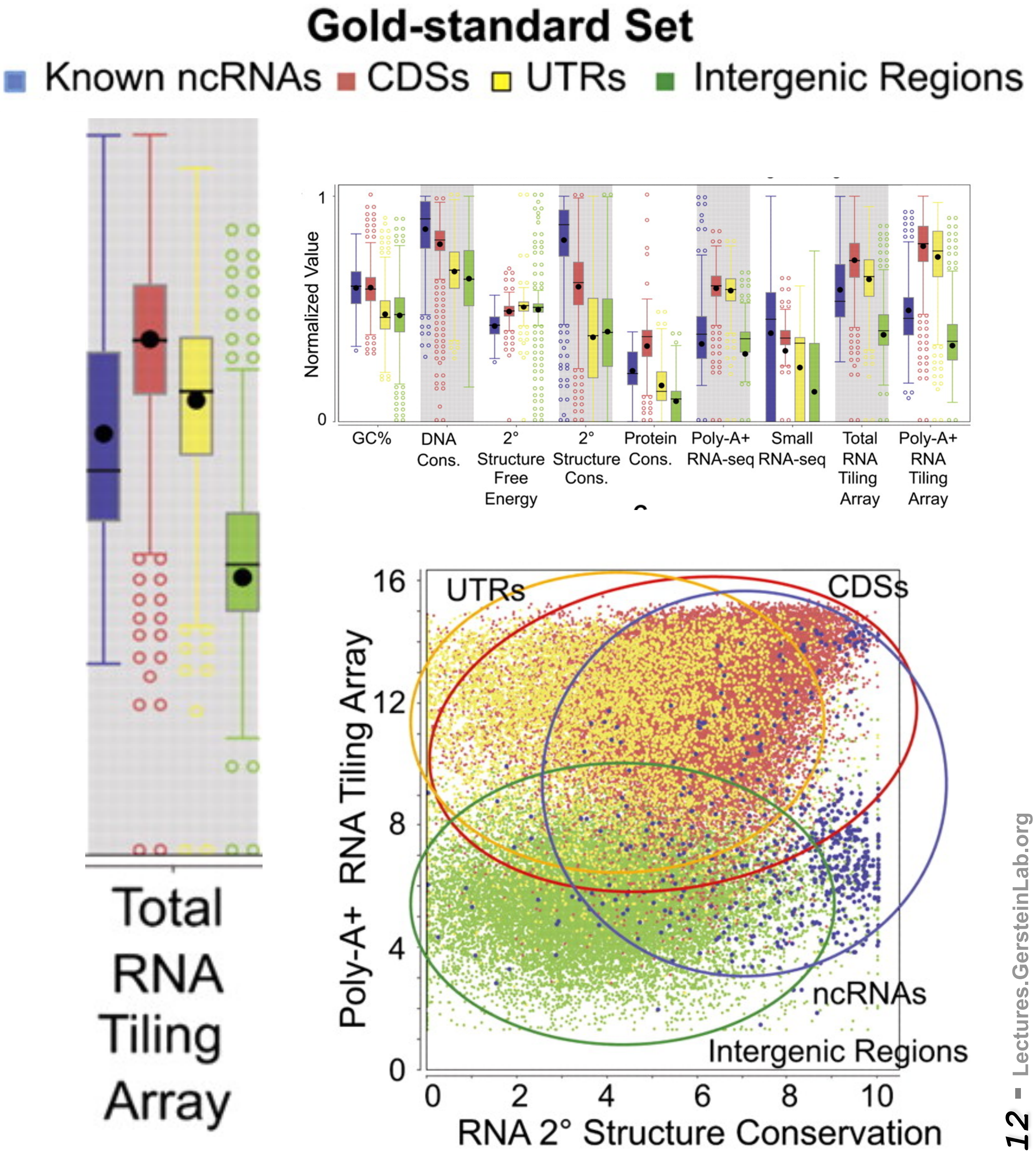
$$Diff_{gene_i}(\Theta^{(Stage1)}, \Theta^{(Stage2)}) = \frac{\sum_{k=1}^{K} (\theta_k^{(Stage1)} - \theta_k^{(Stage2)})^2}{K}$$

# Insights from worm modencode: Approaches useful for human annotation (outline)

- **Expression Timecourse Analysis**
  - Coordinated binding & expression; E v L separation; ~280 large splicing changes
- **ncRNAs** [Hum]
  - Importance of evidence integration
  - Large numbers of transcribed pseudogenes (8-15%)
- **Chromosomal activity distribution [Hum]**
  - Most constrained regions active
  - Repressed arms & binding HOT spots.
- **Regulatory Net [Hum]**
  - Arranging TF binding into a hierarchy with differences betw. levels. Integration with miRNA regulation (more at top).
  - Network motifs & prevalence of FFLs
- **Stat Models relating HMs, TFs & Expression [Hum]**
  - HMs statistically predict expression for protein-coding genes and miRNAs
  - Similar results for TFs, highlighting predictive power of a few TFs.
  - Chromatin model (+ PWM) effective in predicting TF sites  
    -- useful in identifying enhancers.

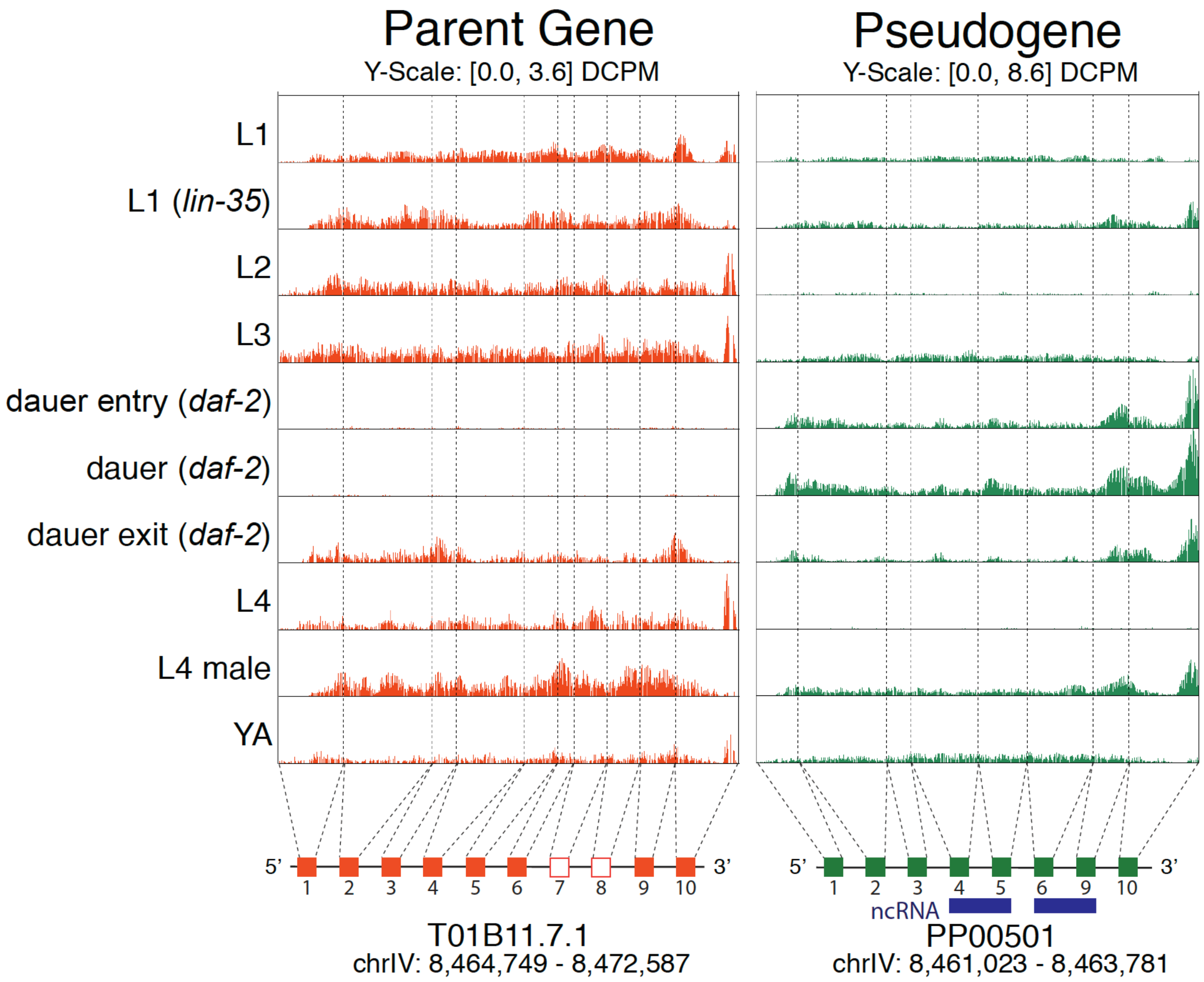
# Identification of many candidate ncRNAs through evidence integration

- ~7k candidates
- No single feature (e.g. expr. expts., conservation, or sec. struc.) finds all known ncRNAs => combine features in stat. model
- 90% PPV, 13 of 15 tested validate



[Lu et al. Genome Res. 2011;21:276-285]

# One type of ncRNA: Transcribed Pseudogenes

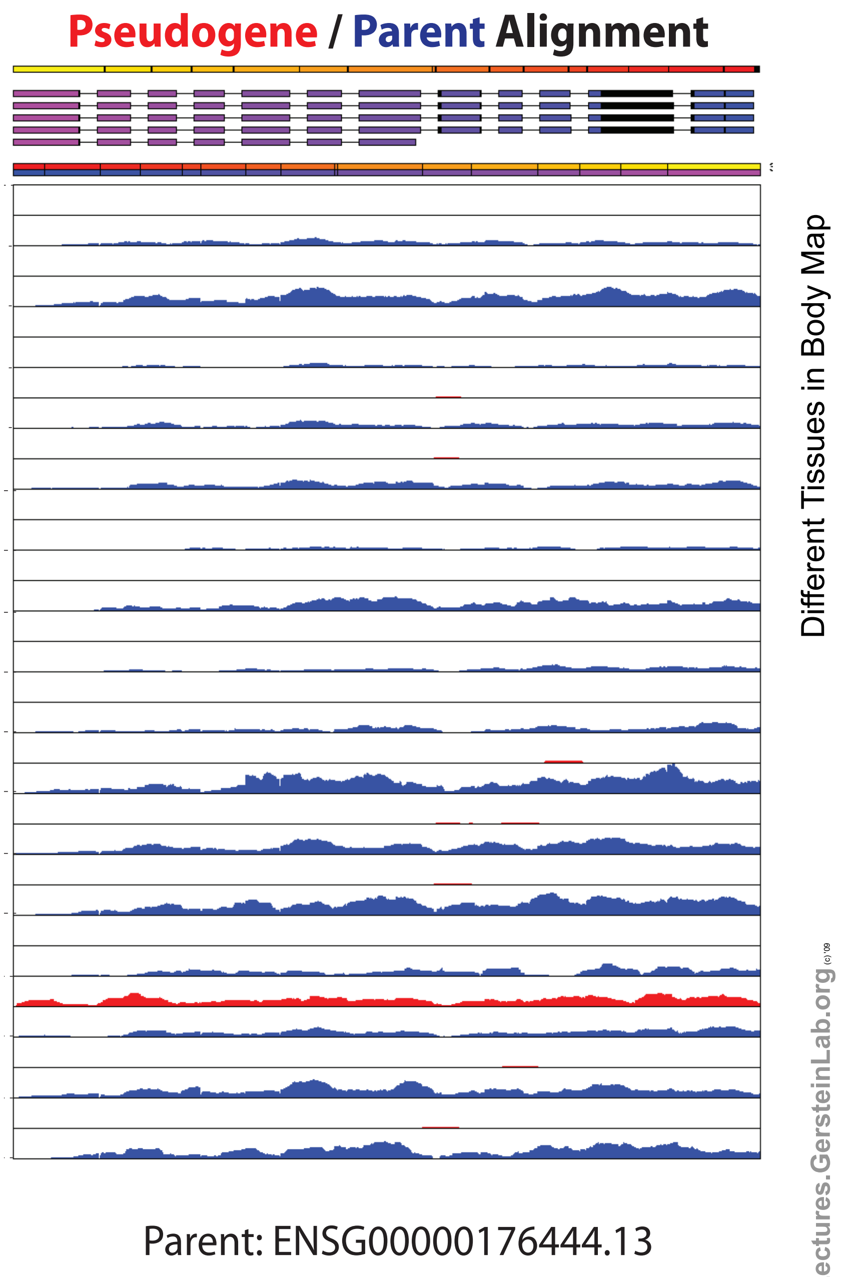


- 1198 total pseudogenes
- 194 (16%) have strong evidence of independent transcription

[*Science* 330:6012]

# Human ncRNAs and Pseudogene Transcription

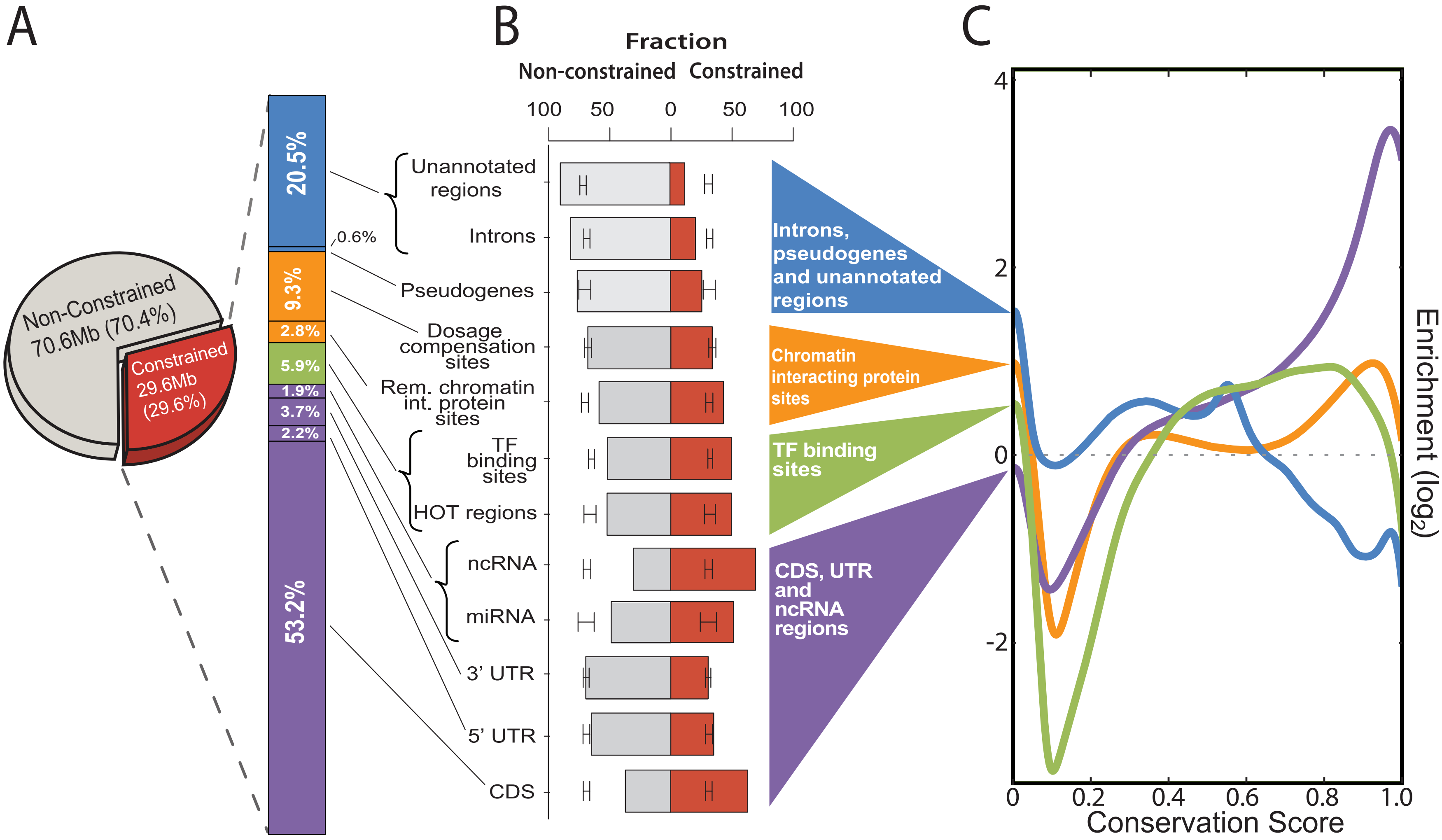
- Gencode 10 :
  manual annotation
  + a variety of pipelines
  - ~5500 lincRNAs
  - 11216 high-qual. pseudogenes (from ~14K total)
    - Total transcribed pseudogene: 876 (RTPCR validated: 57 of 76)



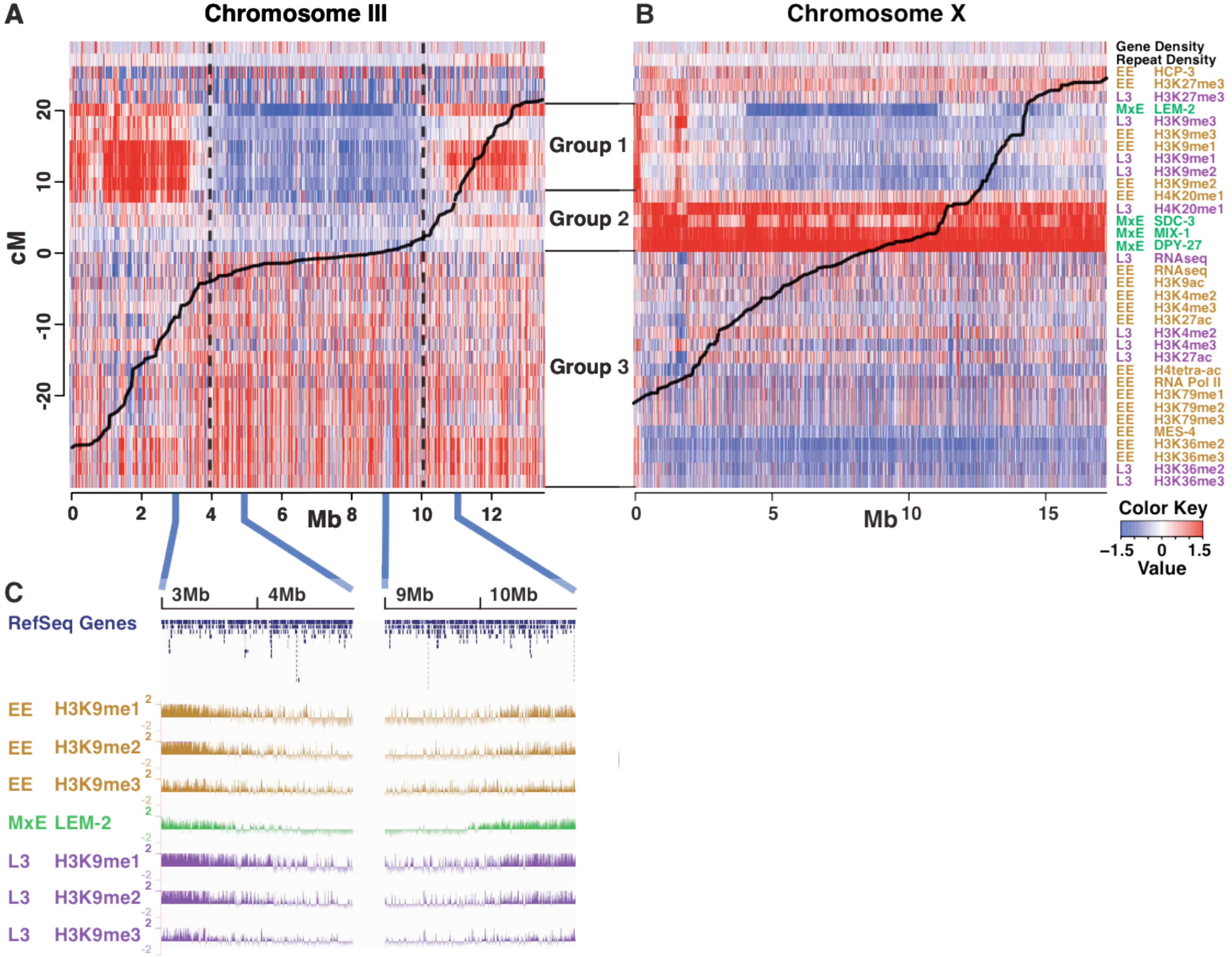
# Insights from worm modencode: Approaches useful for human annotation (outline)

- **Expression Timecourse Analysis**
  - Coordinated binding & expression; E v L separation; ~280 large splicing changes
- **ncRNAs [Hum]**
  - Importance of evidence integration
  - Large numbers of transcribed pseudogenes (8-15%)
- **Chromosomal activity distribution** [Hum]
  - Most constrained regions active
  - Repressed arms & binding HOT spots.
- **Regulatory Net [Hum]**
  - Arranging TF binding into a hierarchy with differences betw. levels. Integration with miRNA regulation (more at top).
  - Network motifs & prevalence of FFLs
- **Stat Models relating HMs, TFs & Expression [Hum]**
  - HMs statistically predict expression for protein-coding genes and miRNAs
  - Similar results for TFs, highlighting predictive power of a few TFs.
  - Chromatin model (+ PWM) effective in predicting TF sites
    -- useful in identifying enhancers.

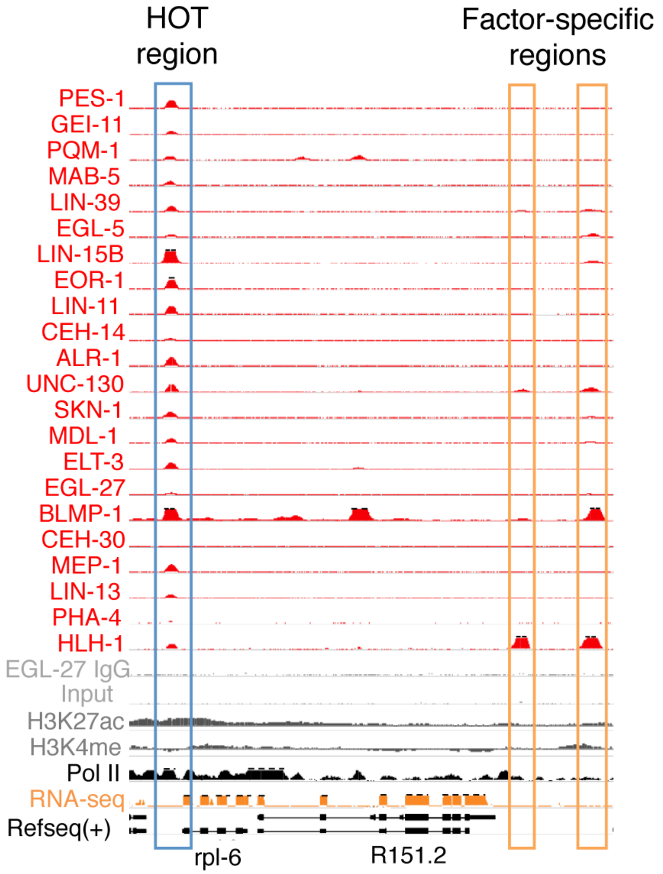
# Conservation of Functional Elements: Most Constrained Bases are Annotated

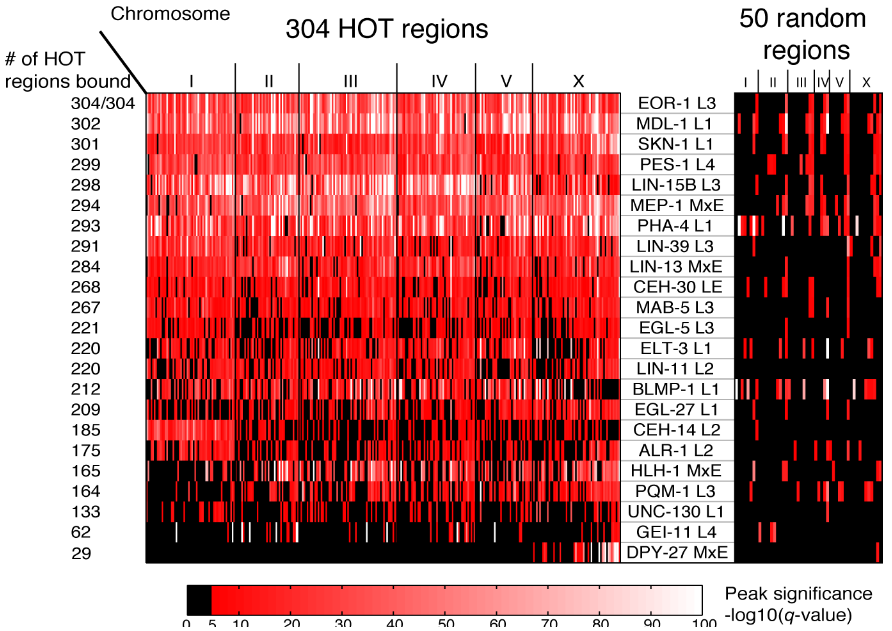


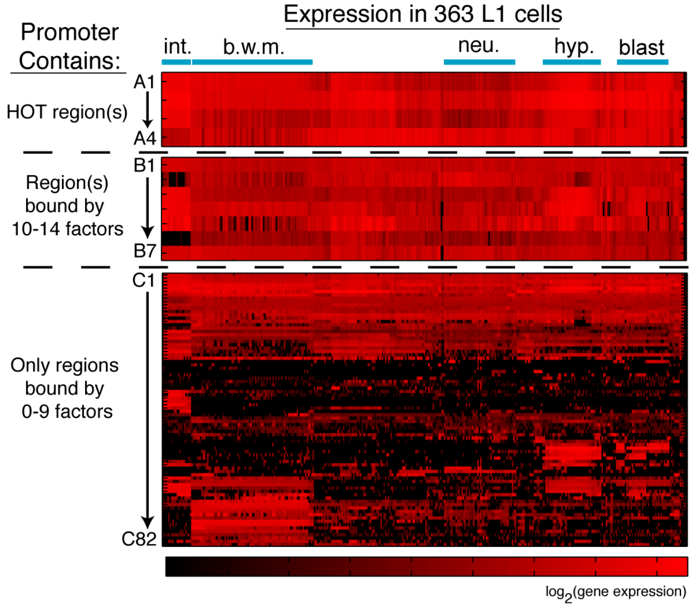
# Large-scale Chromatin Structure

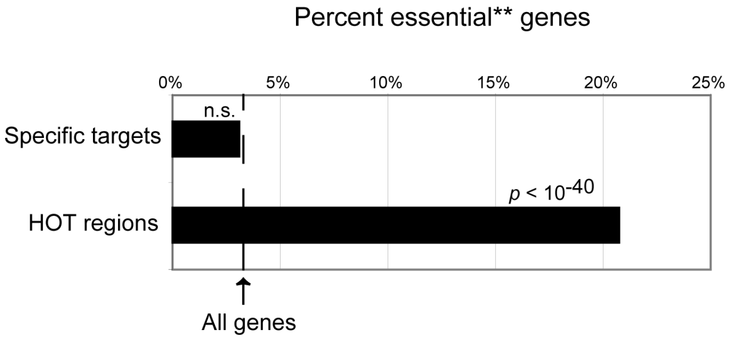


# HOT regions of clustered binding







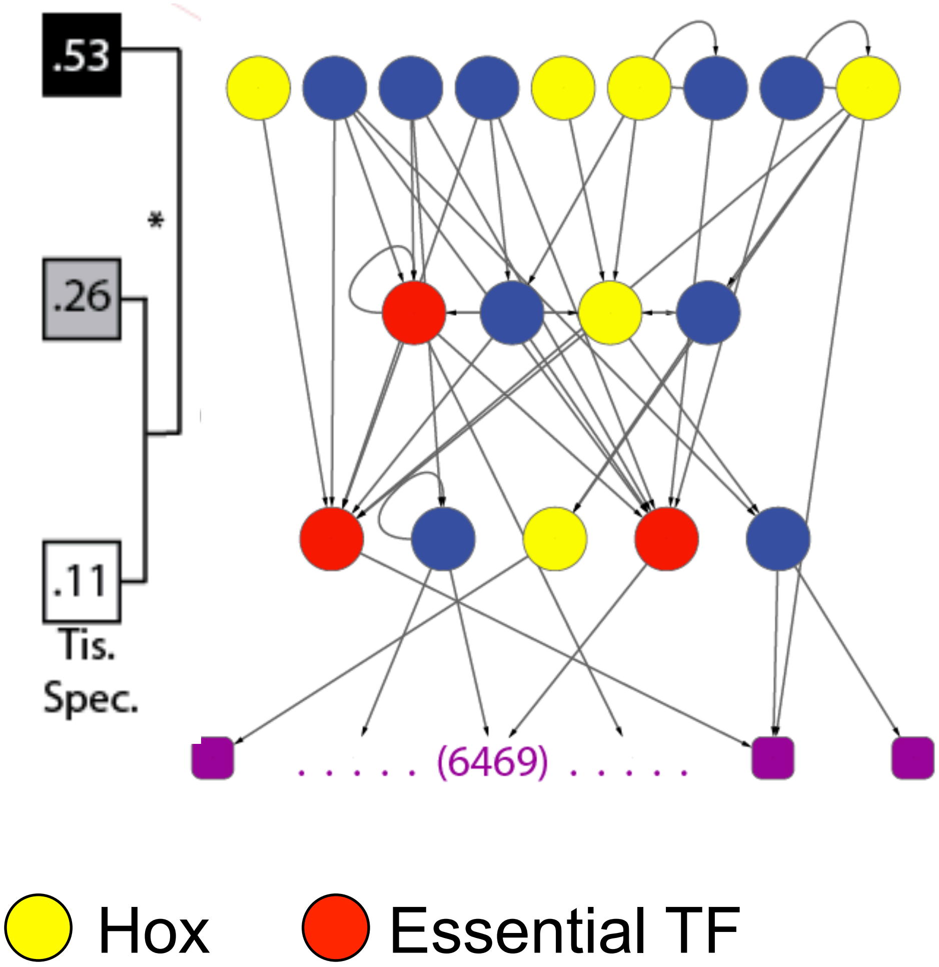


**Similar approach in human ~81K HOT Regions (K5672)**

# Insights from worm modencode: Approaches useful for human annotation (outline)

- **Expression Timecourse Analysis**
  - Coordinated binding & expression; E v L separation; ~280 large splicing changes
- **ncRNAs [Hum]**
  - Importance of evidence integration
  - Large numbers of transcribed pseudogenes (8-15%)
- **Chromosomal activity distribution [Hum]**
  - Most constrained regions active
  - Repressed arms & binding HOT spots.
- **Regulatory Net** [Hum]
  - Arranging TF binding into a hierarchy with differences betw. levels. Integration with miRNA regulation (more at top).
  - Network motifs & prevalence of FFLs
- **Stat Models relating HMs, TFs & Expression [Hum]**
  - HMs statistically predict expression for protein-coding genes and miRNAs
  - Similar results for TFs, highlighting predictive power of a few TFs.
  - Chromatin model (+ PWM) effective in predicting TF sites
    -- useful in identifying enhancers.

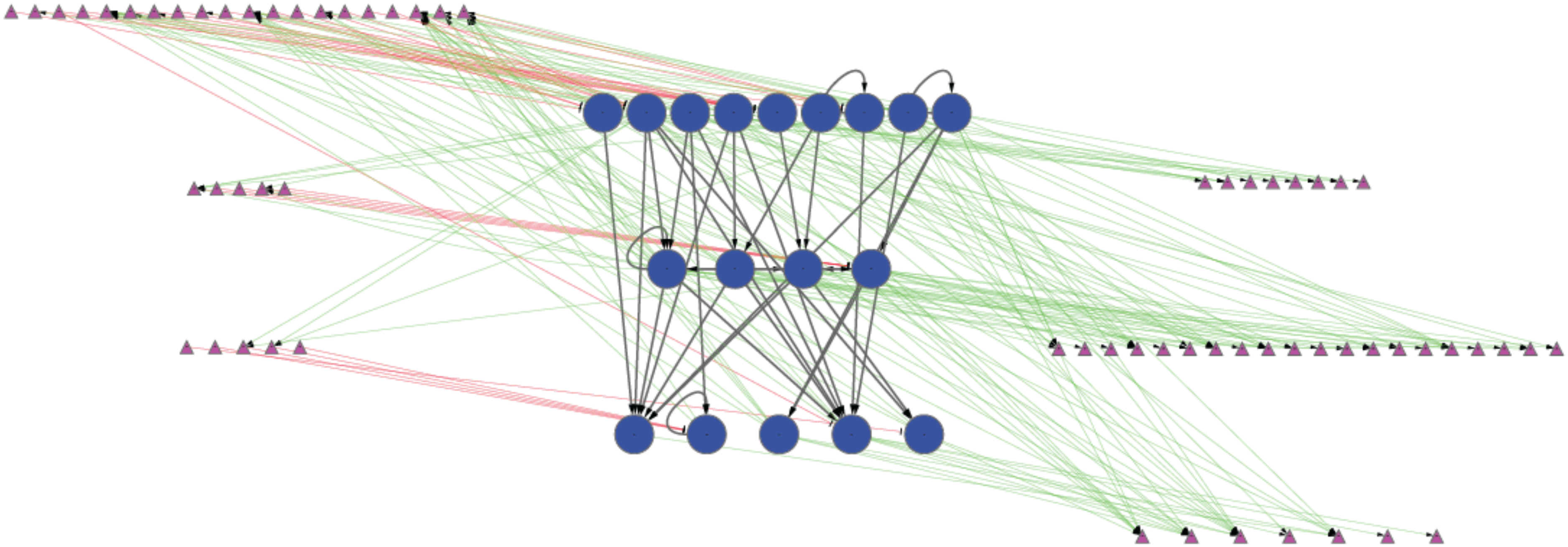
# Worm TF Hierarchy & Gene Properties



- ~25K edges
- Top:
  more tissue specific & HOX (& more miRNA reg.)
- Bottom:
  more essential
- Stats weak but pattern consistent with that in yeast, human…

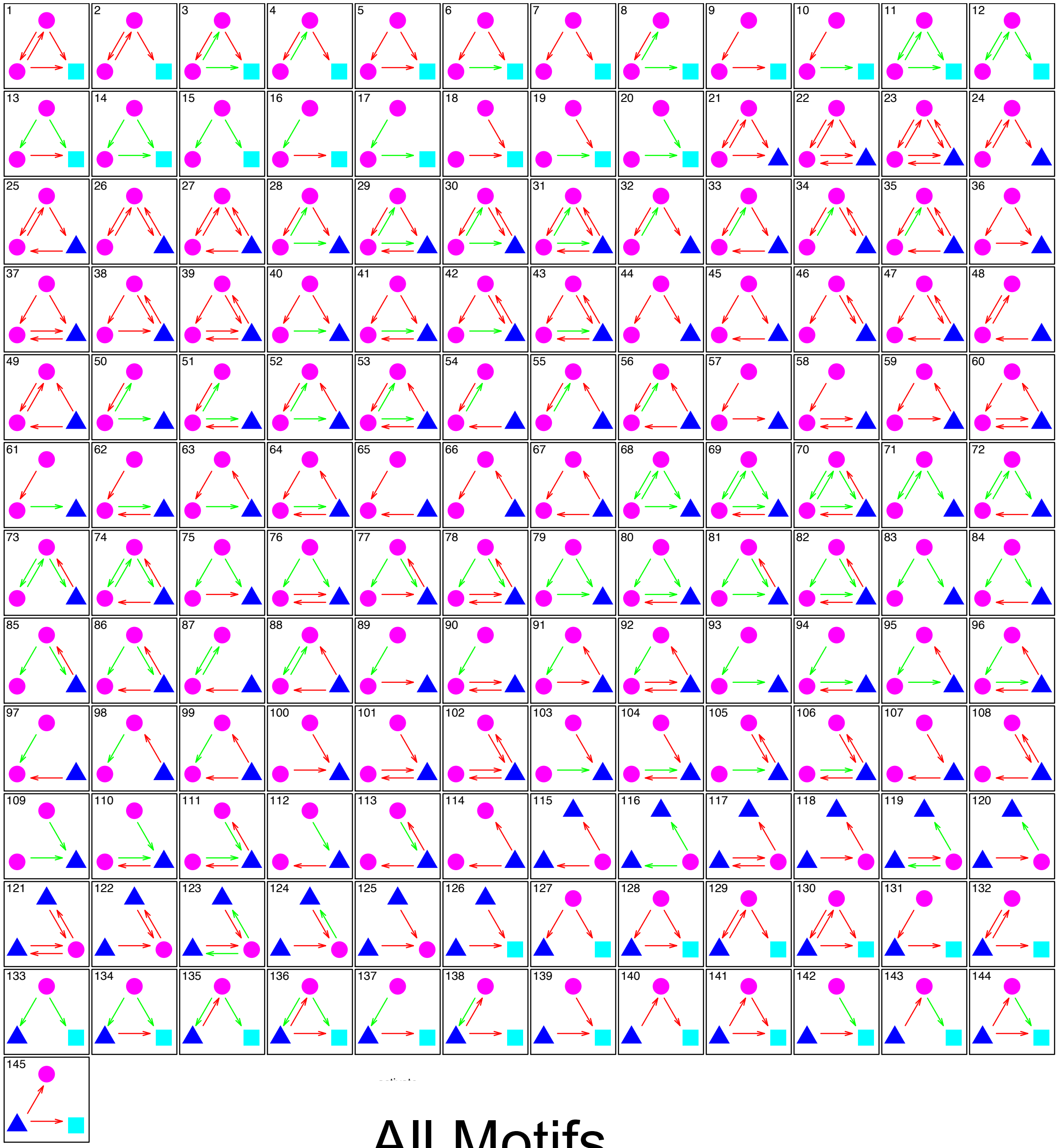
[*Science* 330:6012]

# Relating Worm TF Hierarchy with miRNAs



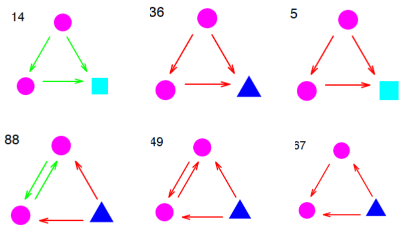
[*Science* 330:6012]

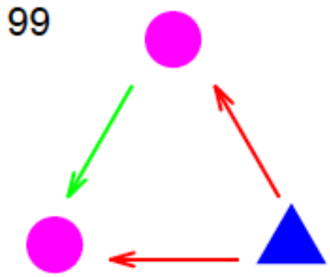
# Network Motifs



All Motifs

7 Motifs
Over-represented



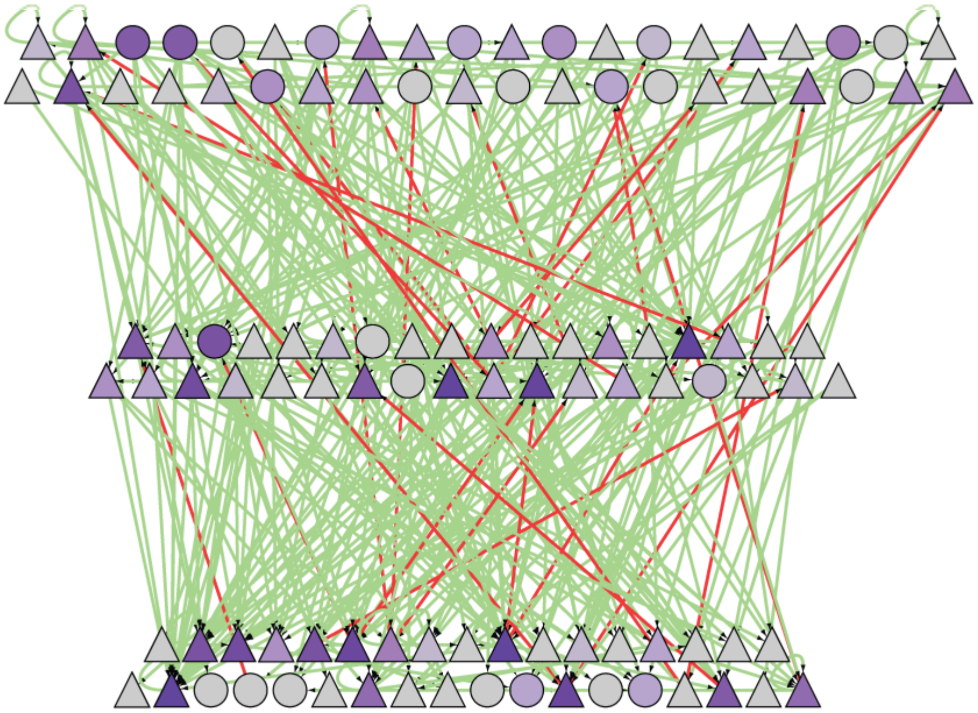


FFL involving
miRNA & 2 TFs

● TF ▲ MIR ■ GENE

activate →
repress →

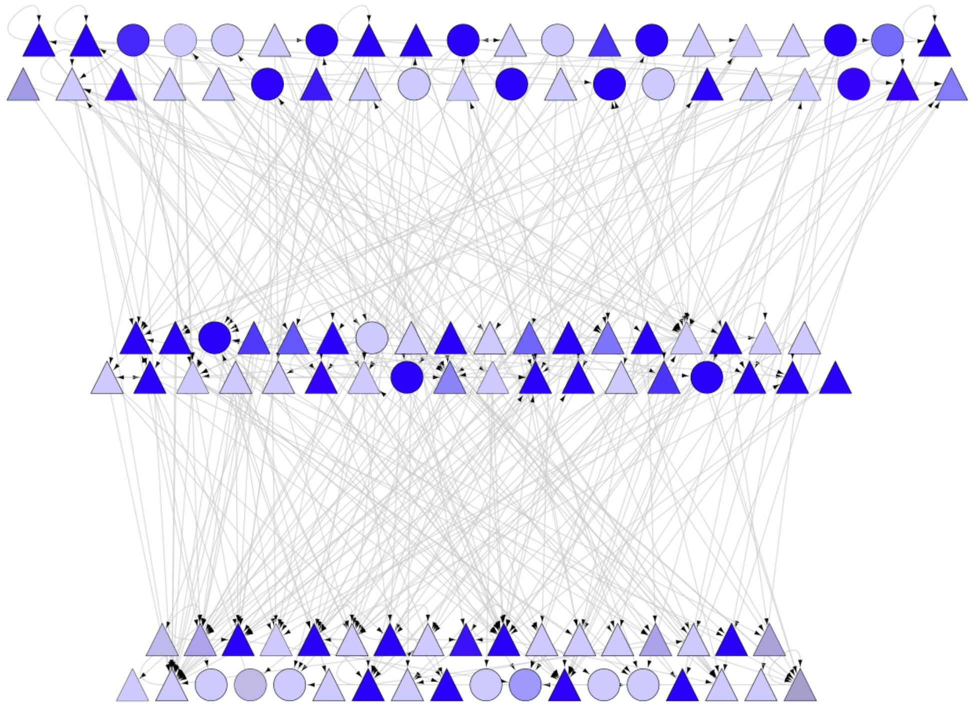
[*Science* 330:6012]

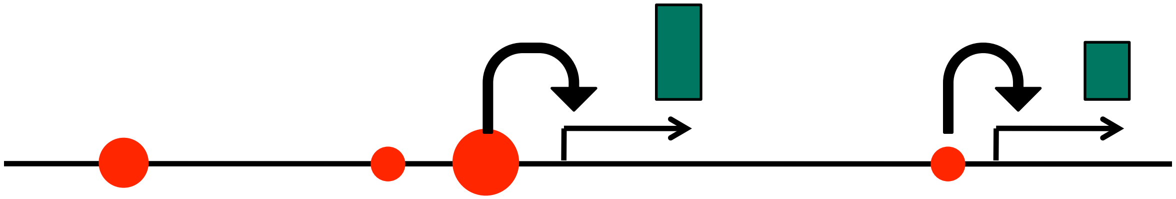


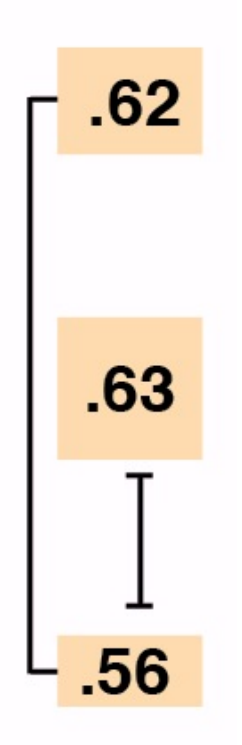
# Human: Strongest Proximal Regulatory Edges Can be Arranged into a Hierarchy

Optimally arrange TFs into 3 levels by sim. annealing, maximizing downward-pointing edges

[ ]



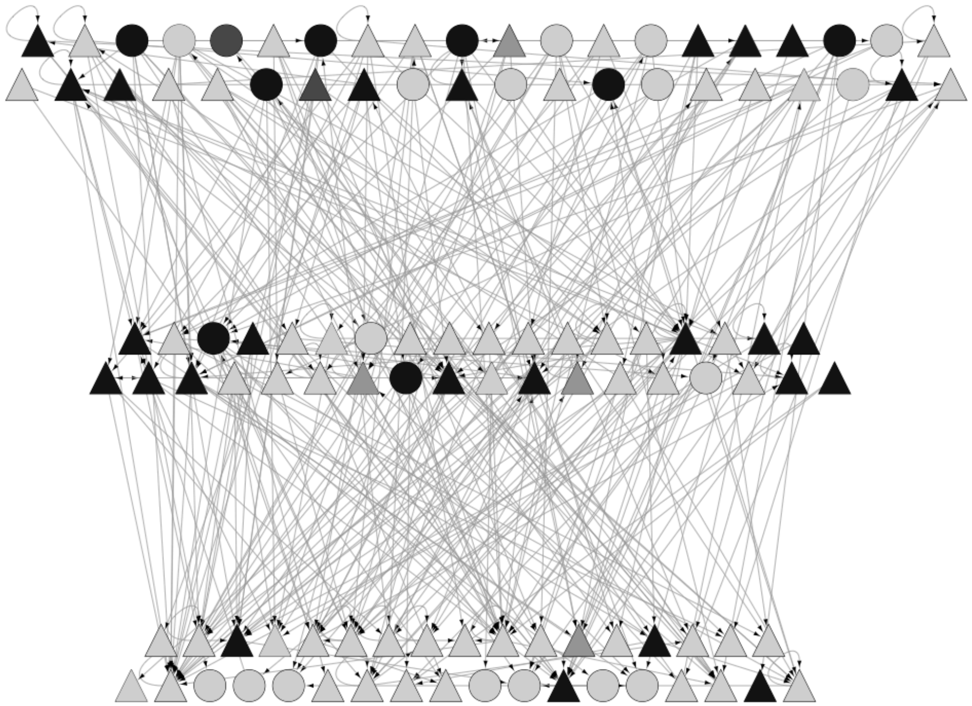


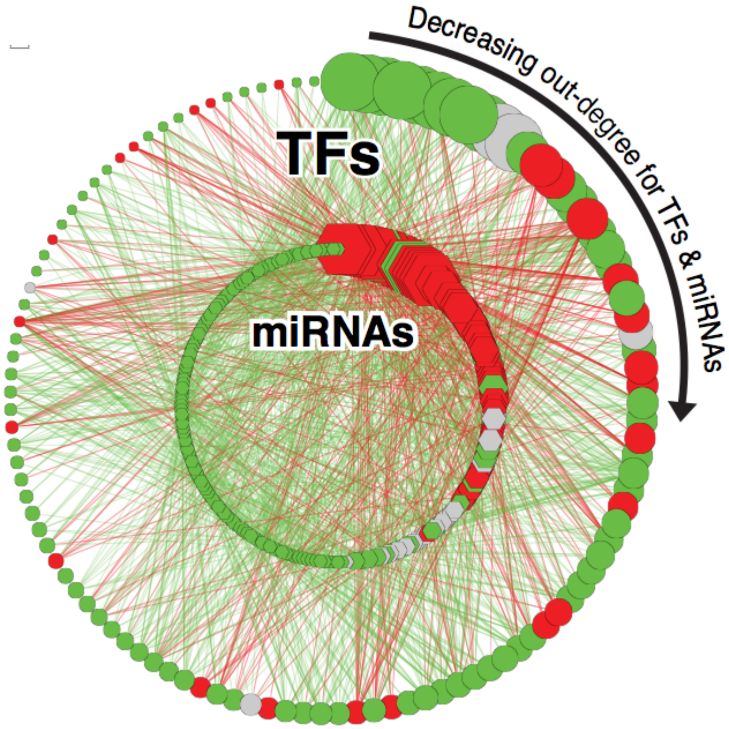


Avg. correlation betw. binding signal of TF & gene expr. of its target

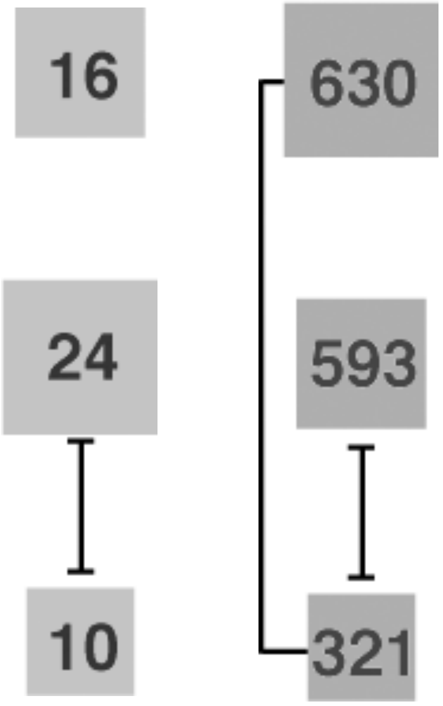
# Integration of TF hierarchy with other ‘omic information :

**more influential & connected TFs on the top**





Sig. corr. w/ TF hubbiness
(.24 & .62)



Avg. values

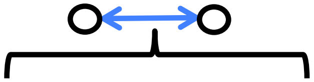
# regulating miRNAs & # regulated miRNAs

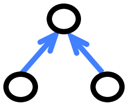
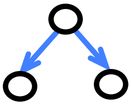
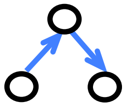
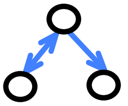
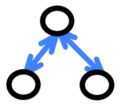
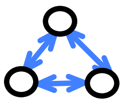
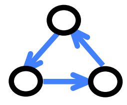
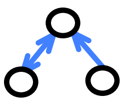
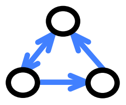
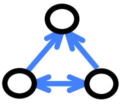
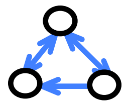
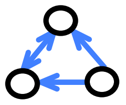
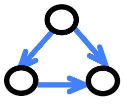
## Integration of TF hierarchy with other ‘omic information :

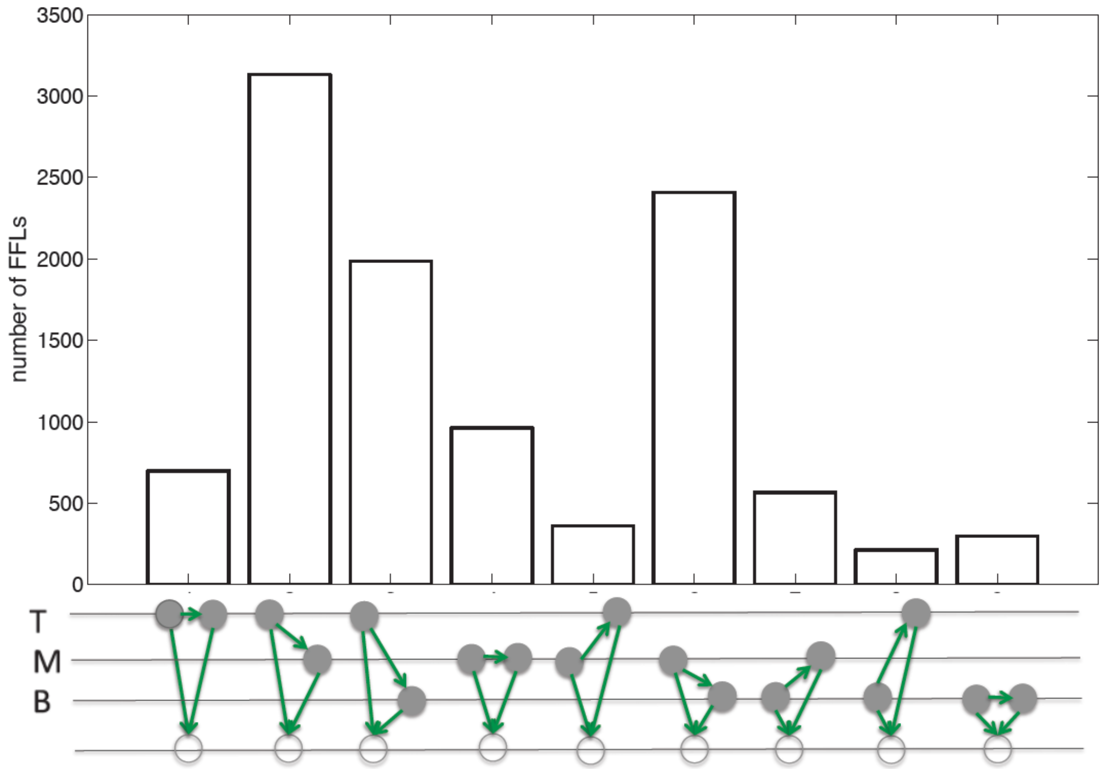
more influential & connected TFs on the top

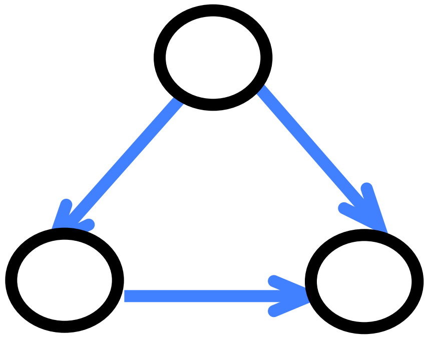
# Network Motif Analysis: Enrichment of FFLs

6 pairs of toggle switches



| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3-node motifs |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Freq. N | 868<br>(0.84) | 490<br>(0.81) | 729<br>(0.72) | 26<br>(0.62) | 0<br>(0) | 0<br>(0) | 8<br>(0.70) | 64<br>(0.91) | 6<br>(1.8) | 6<br>(2.9) | 2<br>(5.6) | 16<br>(4.8) | 122<br>(1.4) |





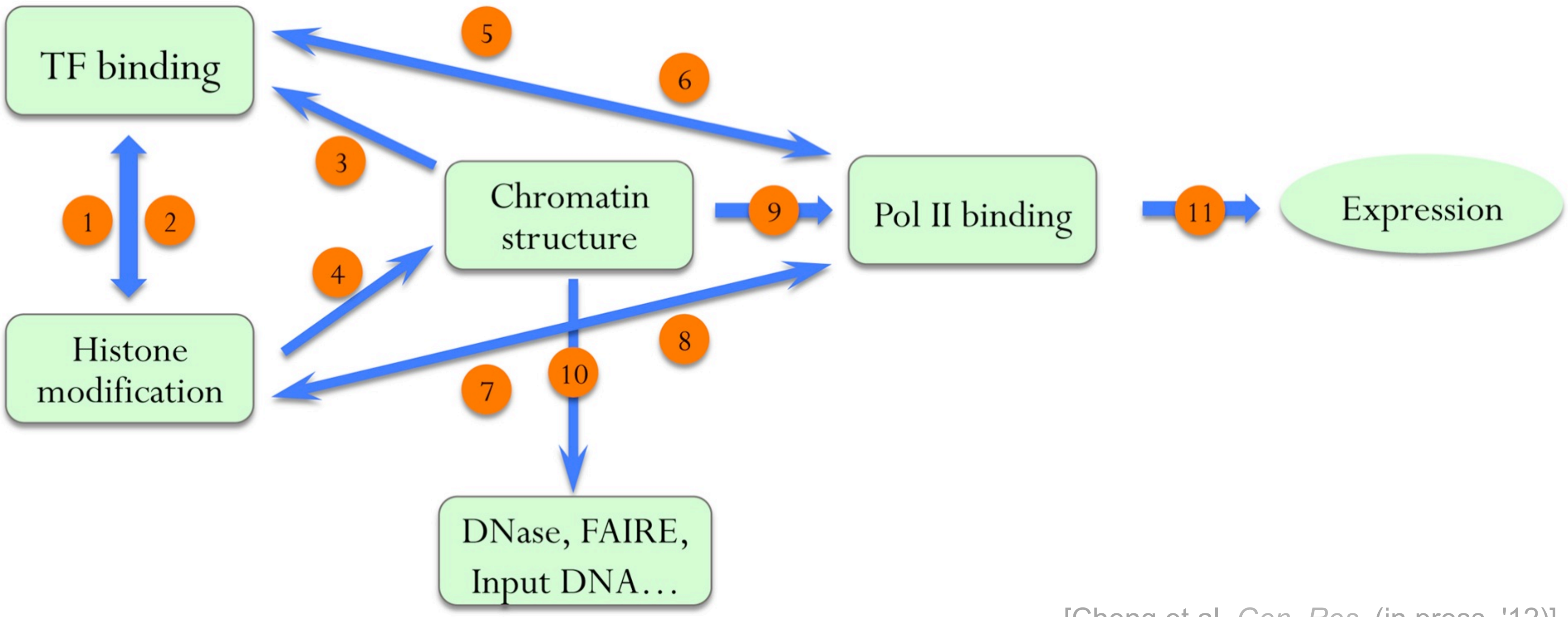
[ ]

# Insights from worm modencode: Approaches useful for human annotation (outline)

- **Expression Timecourse Analysis**
  - Coordinated binding & expression; E v L separation; ~280 large splicing changes
- **ncRNAs [Hum]**
  - Importance of evidence integration
  - Large numbers of transcribed pseudogenes (8-15%)
- **Chromosomal activity distribution [Hum]**
  - Most constrained regions active
  - Repressed arms & binding HOT spots.
- **Regulatory Net [Hum]**
  - Arranging TF binding into a hierarchy with differences betw. levels. Integration with miRNA regulation (more at top).
  - Network motifs & prevalence of FFLs
- **Stat Models relating HMs, TFs & Expression [Hum]**
  - **HMs statistically predict expression** for protein-coding genes and miRNAs
  - Similar results for TFs, highlighting predictive power of a few TFs.
  - Chromatin model (+ PWM) effective in predicting TF sites
    -- useful in identifying enhancers.

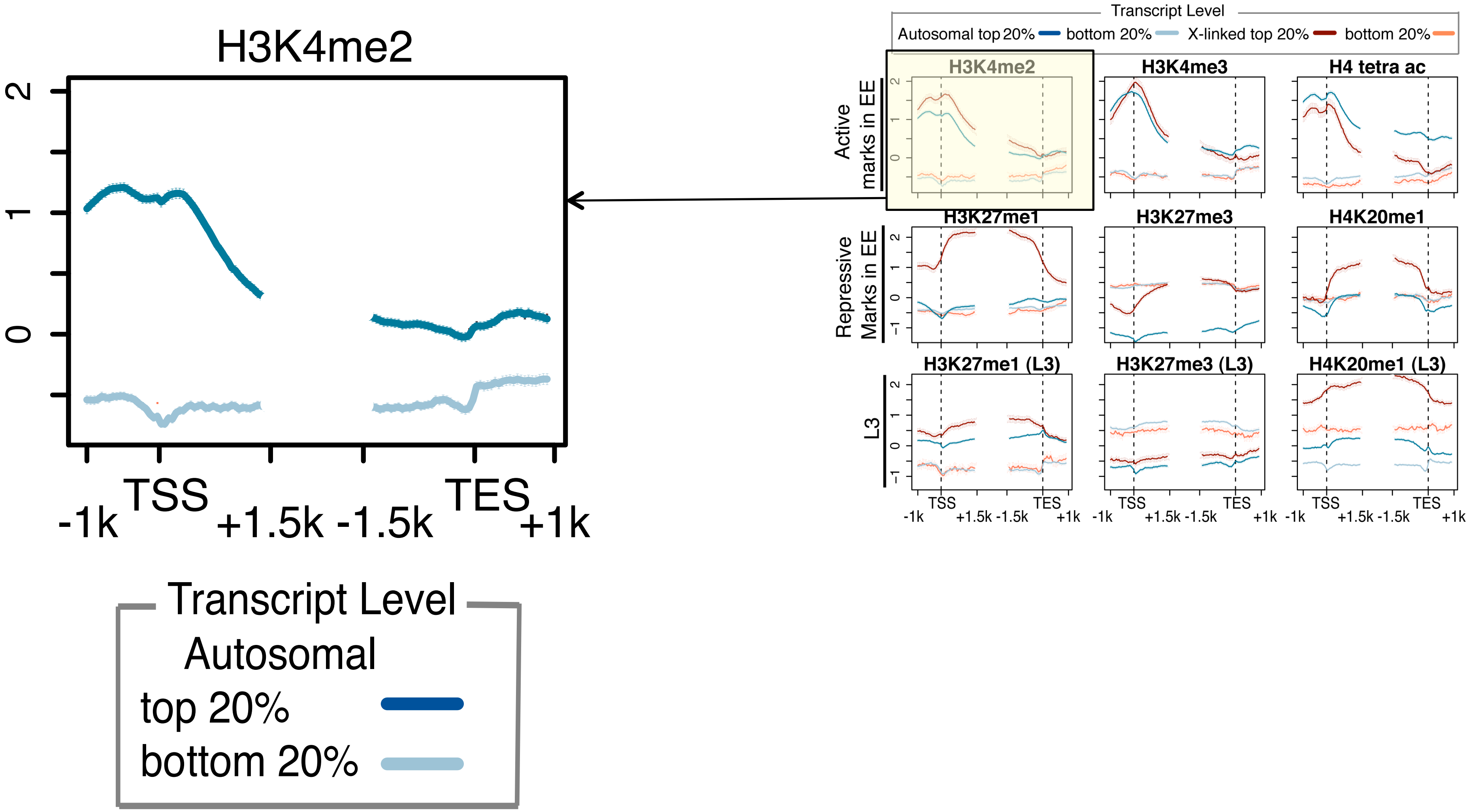
# Modeling Transcription: Connecting Inputs & Outputs

- Models
  - HMs+TFs => gene expression
  - HMs => TFs



[Cheng et al. *Gen. Res.* (in press, '12)]

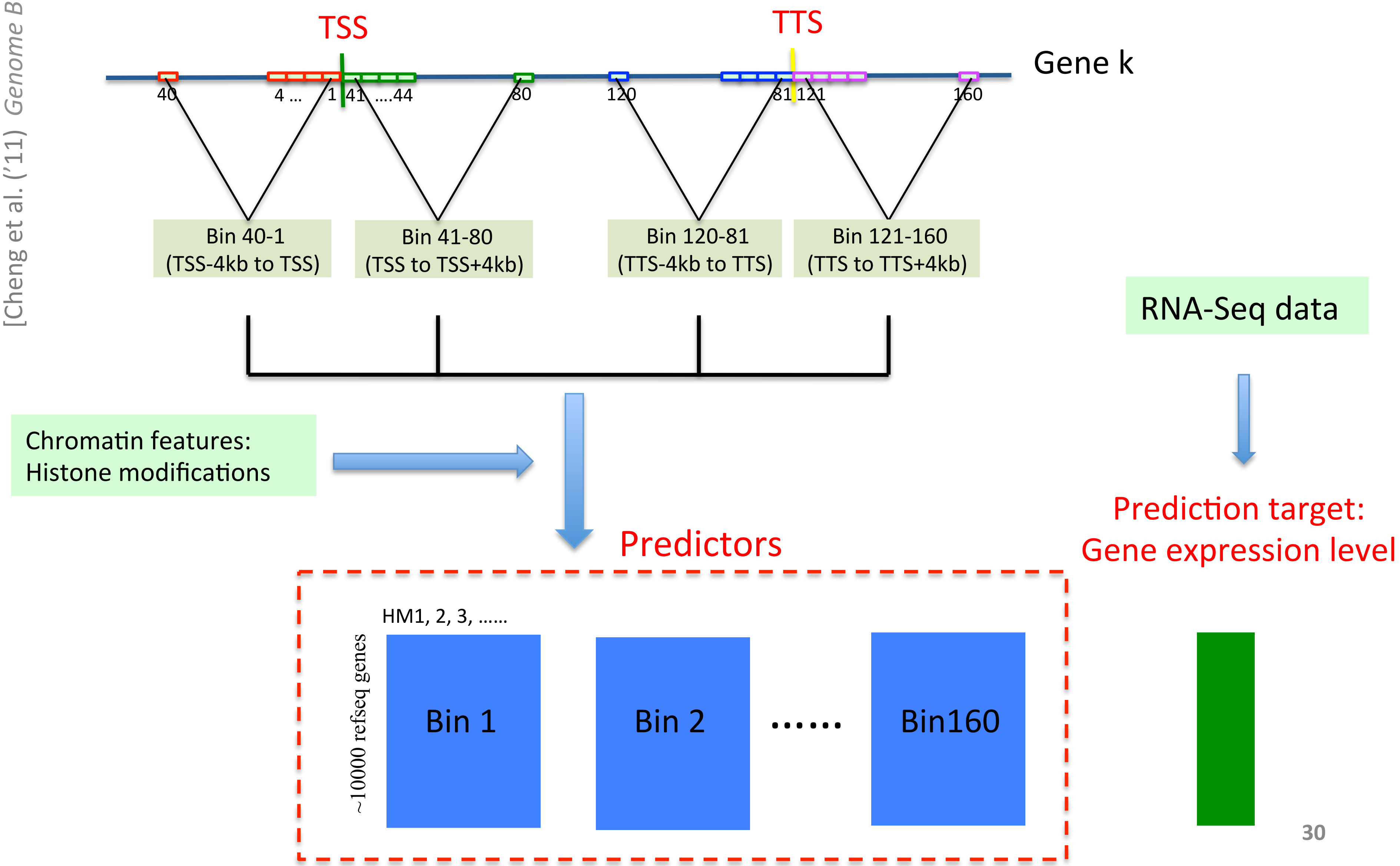
# His. mods around TSS are related to level of gene expression



[*Science* 330:6012]

[Cheng et al. ('11) *Genome Biol.* 12: R15]

# Histone Modification model



Bin 40-1 (TSS-4kb to TSS)

Bin 41-80 (TSS to TSS+4kb)

Bin 120-81 (TTS-4kb to TTS)

Bin 121-160 (TTS to TTS+4kb)

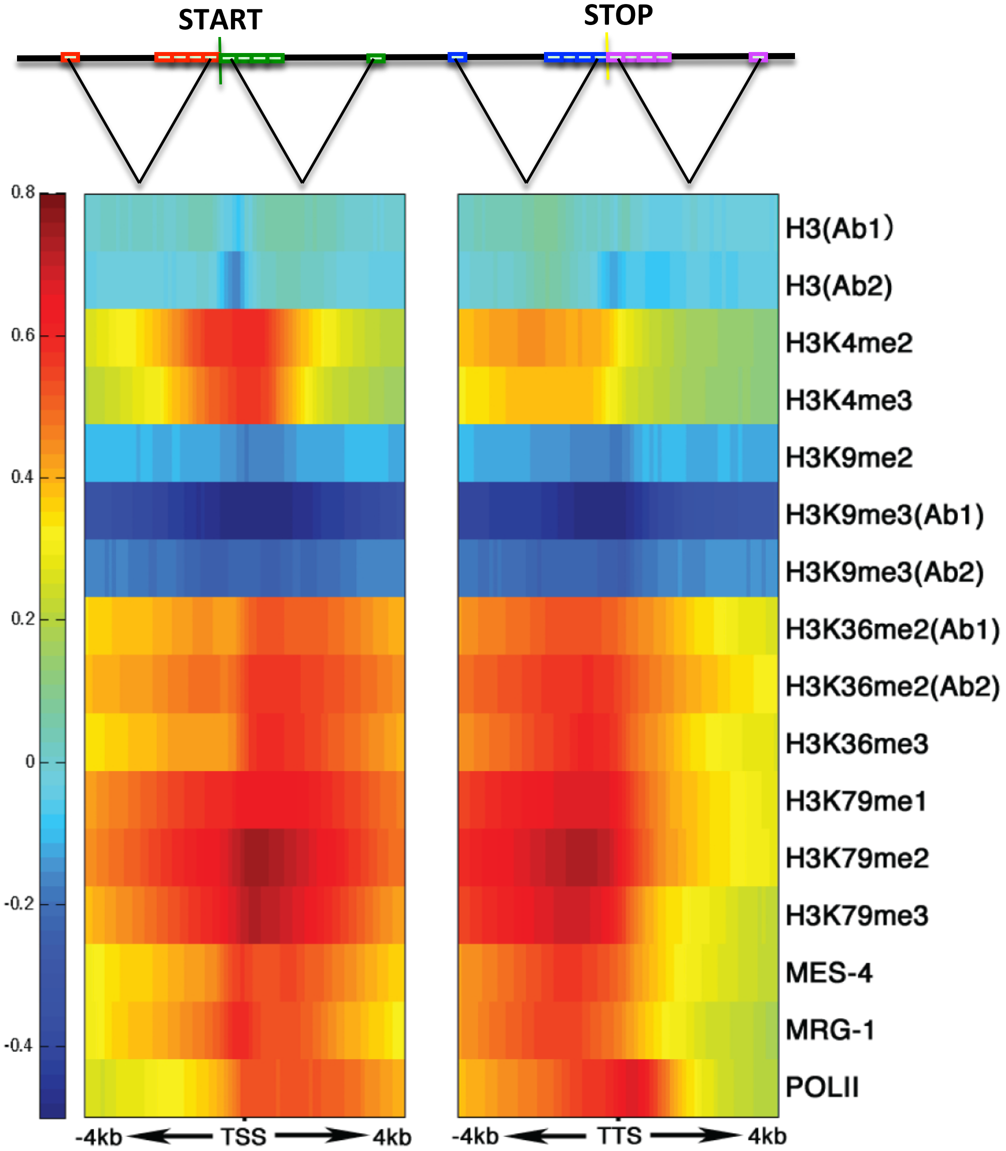
RNA-Seq data

Chromatin features: Histone modifications

Predictors

Prediction target: Gene expression level

# His. mods around TSS & TTS are clearly related to level of gene expression, in a position-dependent fashion

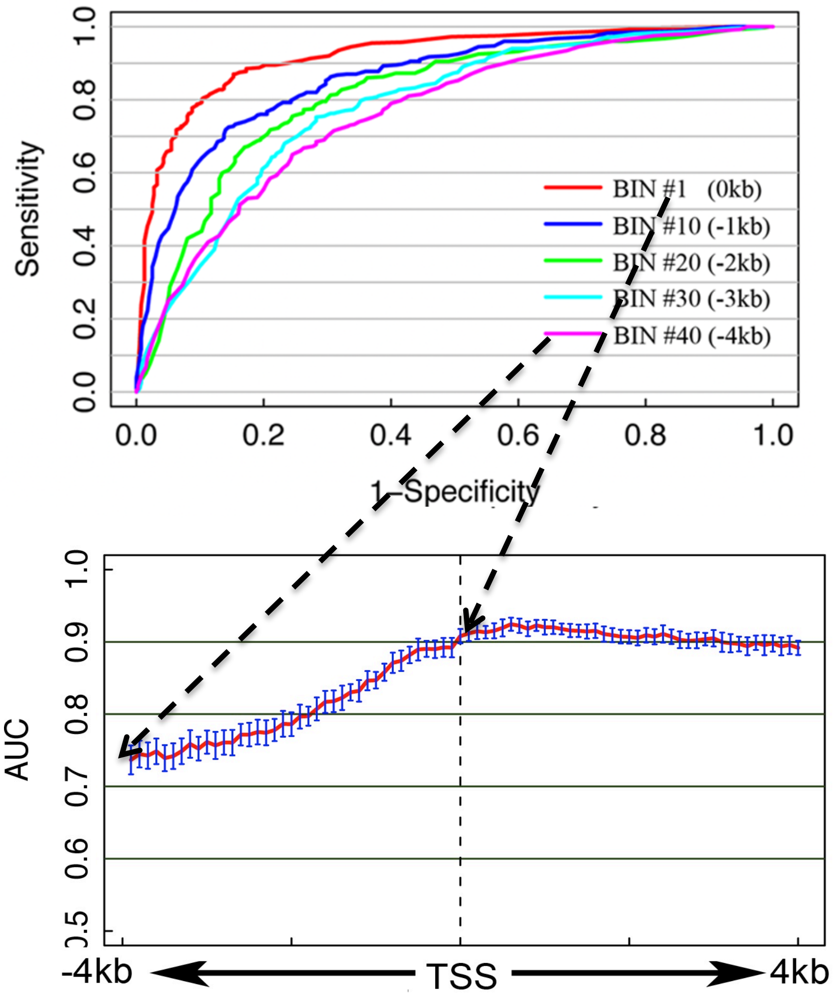


[*Science* 330:6012] [Related work: Ouyang et al. ('09) *PNAS*; Karlic et al. ('10) *PNAS*]

[Cheng et al. ('11) *Genome Biol.* 12: R15]

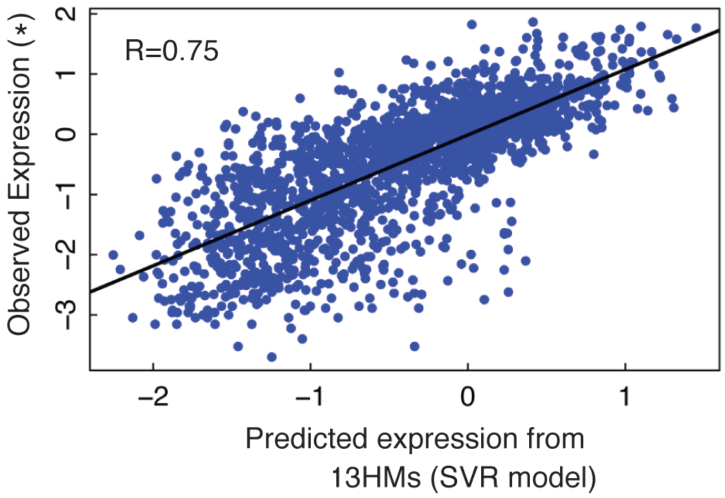
# Integrate all histone modifications to predict gene expression levels

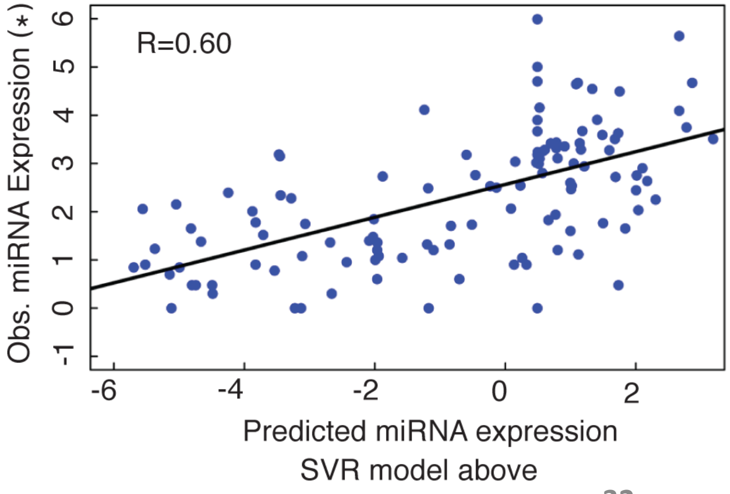
### Classify H/L genes (SVM)



Magnitude of Prediction from a "bin" around the TSS

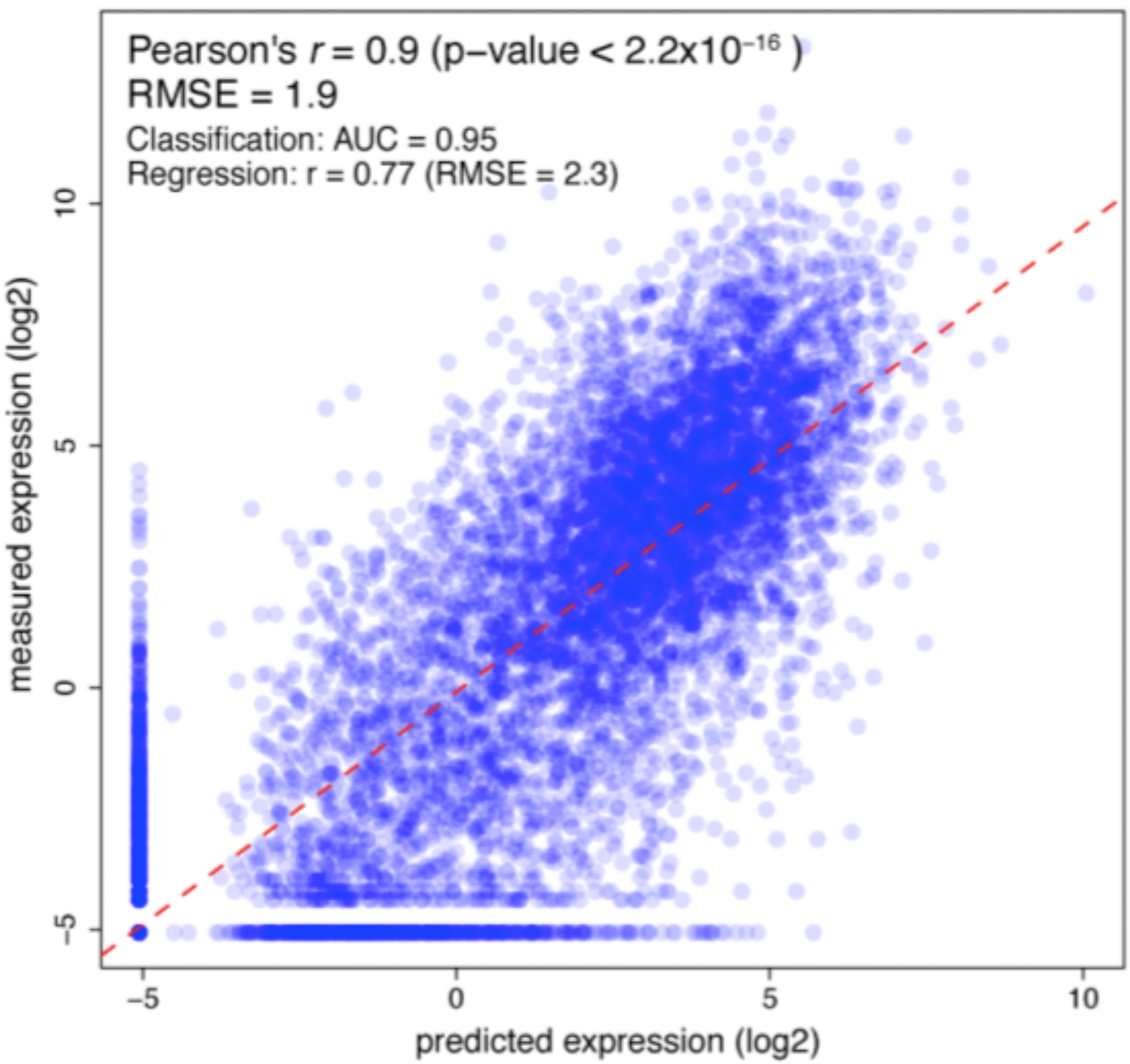
### Predict expression values

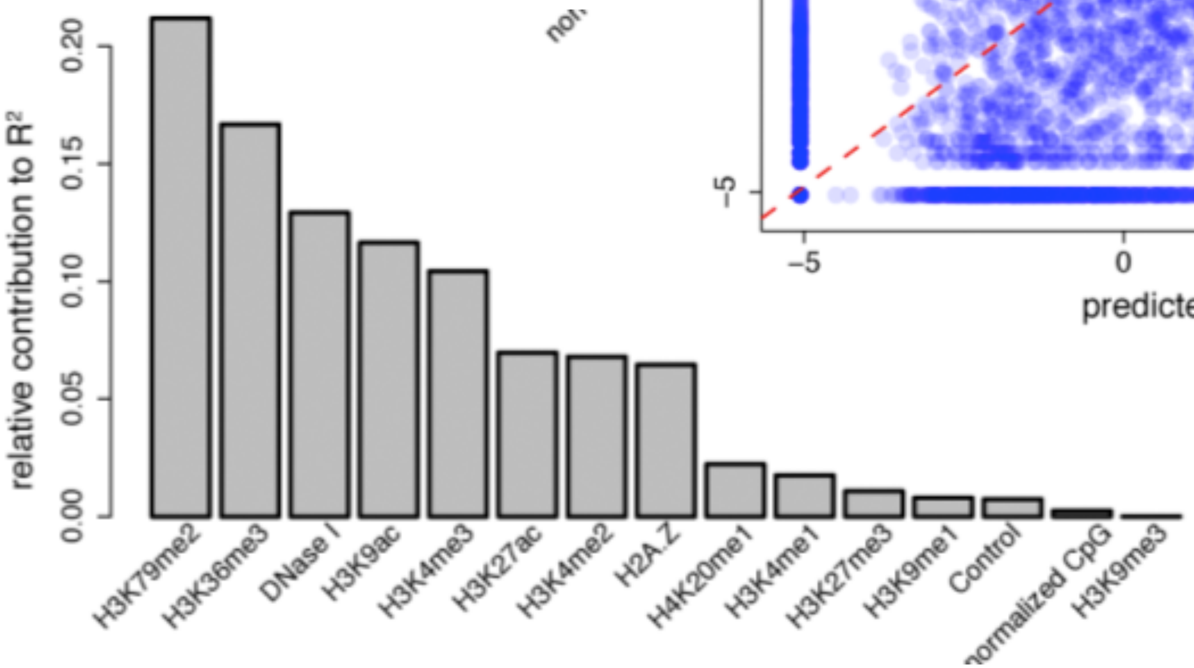




$*$ = LOG$_{10}$RPKM

# Scale up to Human & Mouse

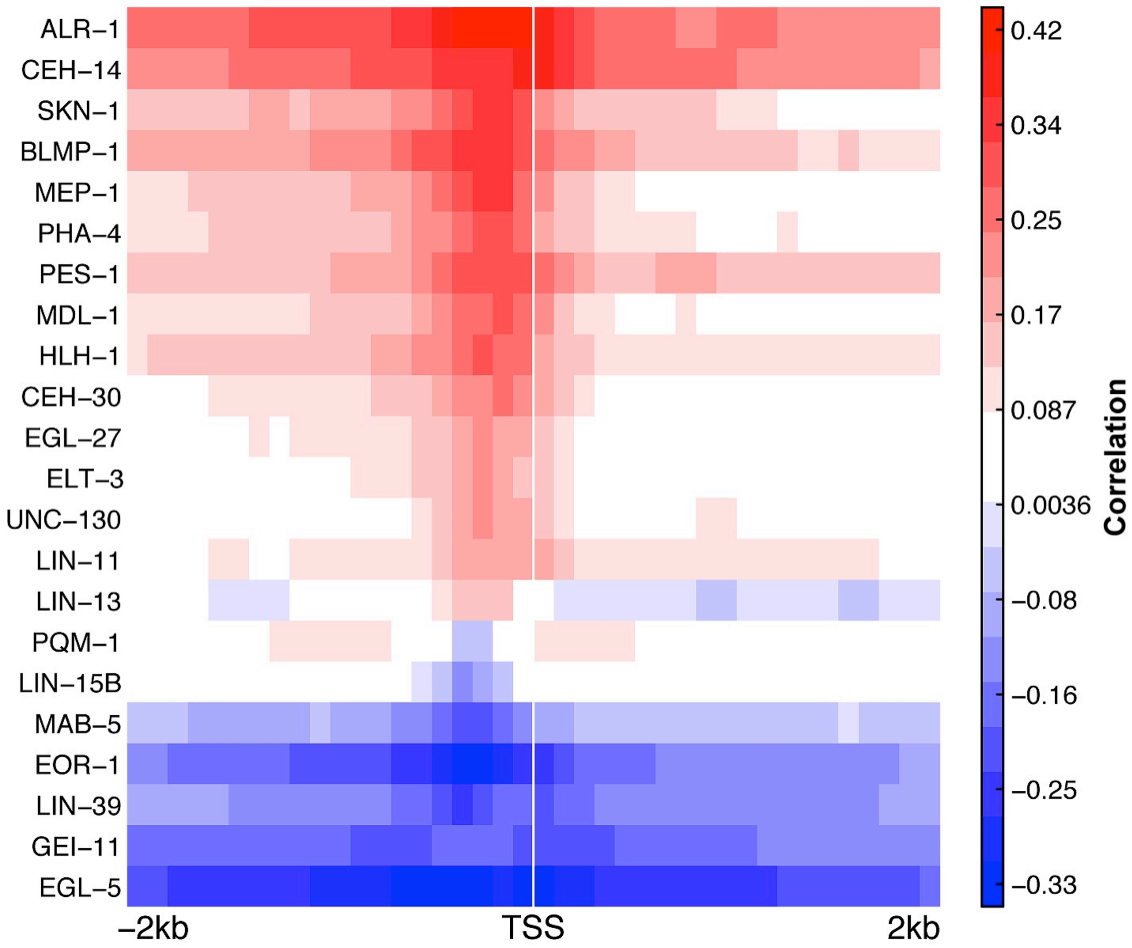




# Insights from worm modencode: Approaches useful for human annotation (outline)

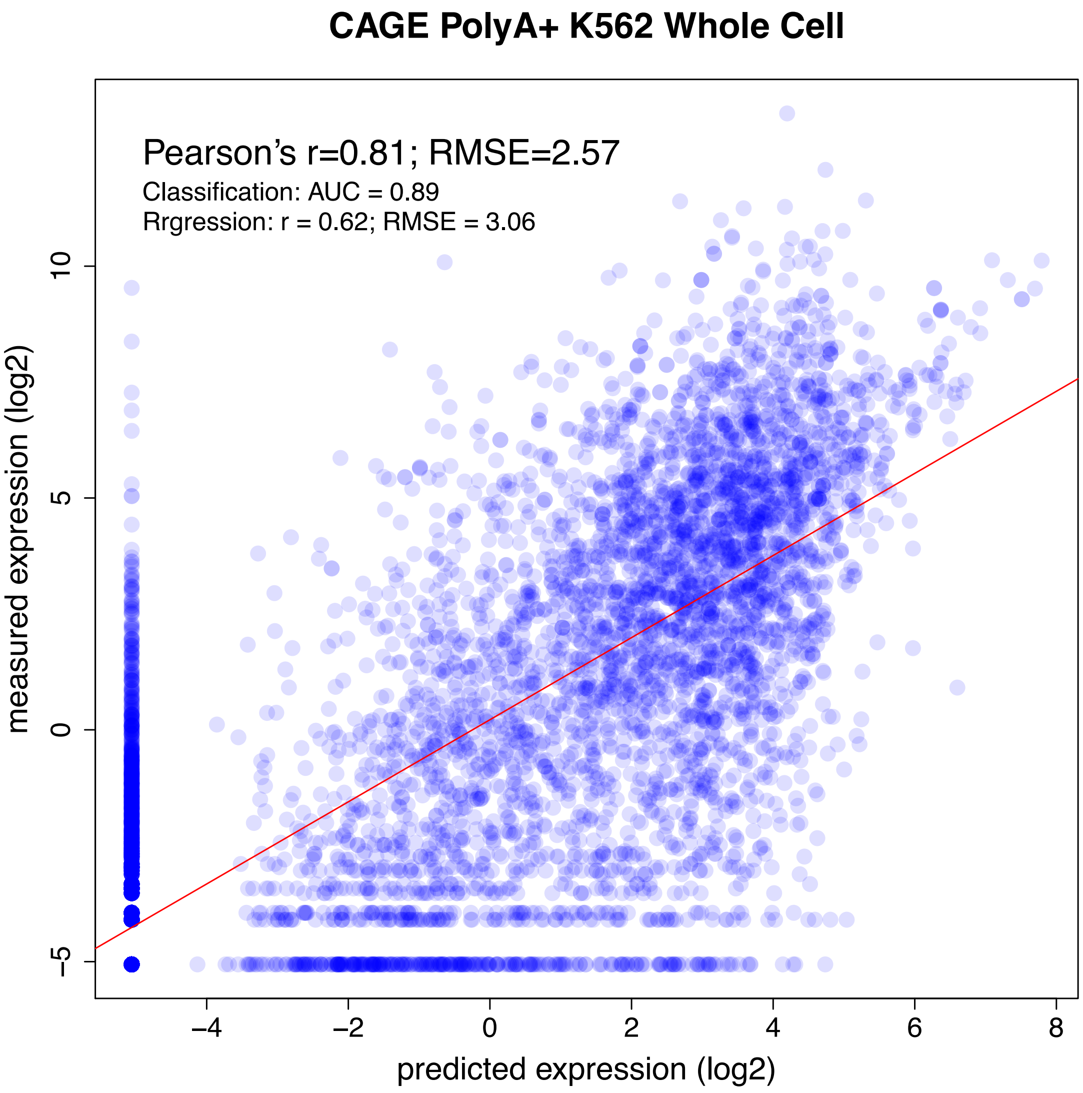
- **Expression Timecourse Analysis**
  - Coordinated binding & expression; E v L separation; ~280 large splicing changes
- **ncRNAs [Hum]**
  - Importance of evidence integration
  - Large numbers of transcribed pseudogenes (8-15%)
- **Chromosomal activity distribution [Hum]**
  - Most constrained regions active
  - Repressed arms & binding HOT spots.
- **Regulatory Net [Hum]**
  - Arranging TF binding into a hierarchy with differences betw. levels. Integration with miRNA regulation (more at top).
  - Network motifs & prevalence of FFLs
- **Stat Models relating HMs, TFs & Expression [Hum]**
  - HMs statistically predict expression for protein-coding genes and miRNAs
  - **Similar results for TFs**, highlighting predictive power of a few TFs.
  - Chromatin model (+ PWM) effective in predicting TF sites -- useful in identifying enhancers.

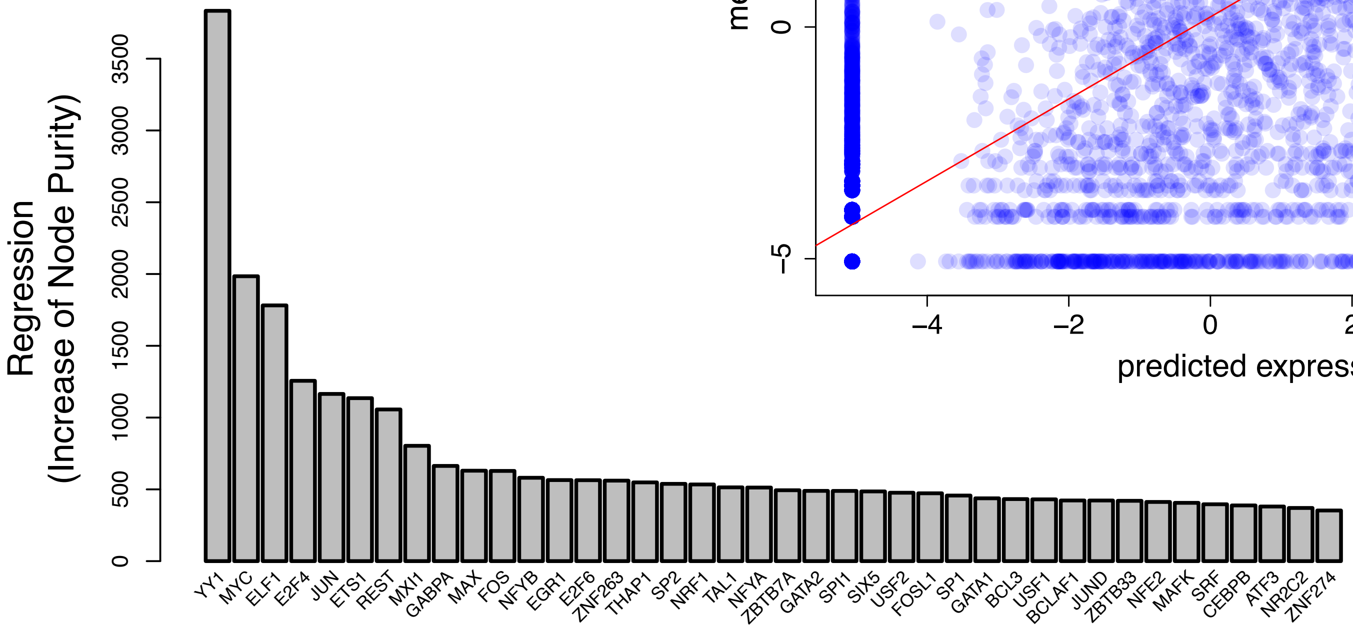
# Modeling with Worm TFs: Positive and negative regulators from correlating TF signal at TSS with gene expression



[ ]

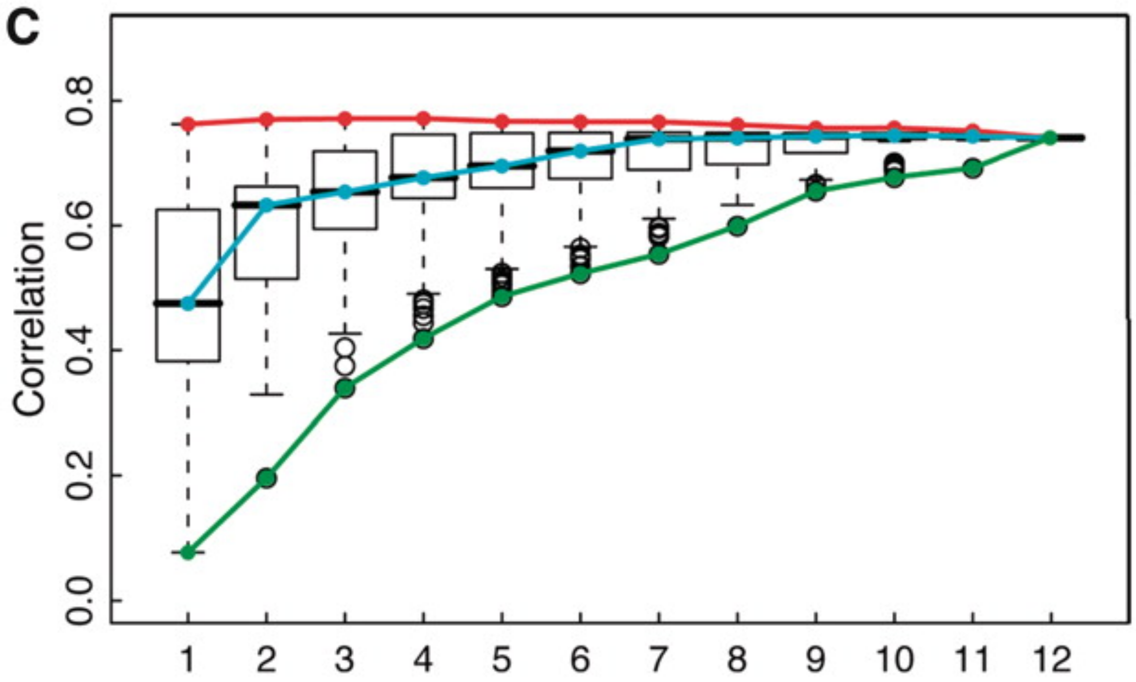
# Scale up to Human: TFs



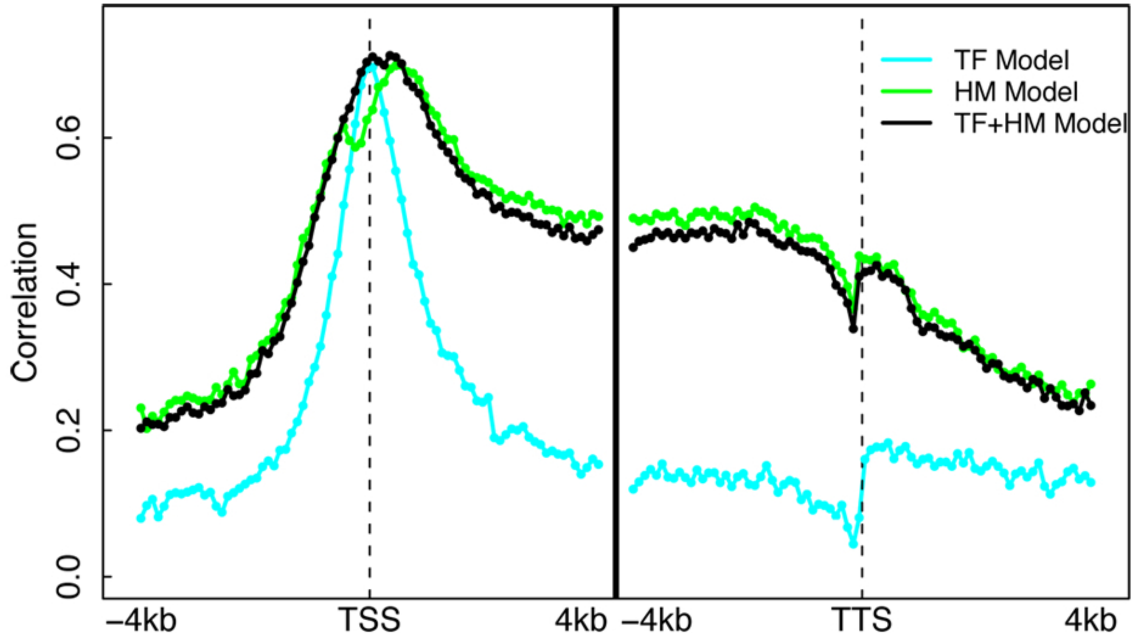


# Aspects of Mammalian (Human & Mouse) TF Model

- Model with only a few of the 1000s of total TFs is able to predict well



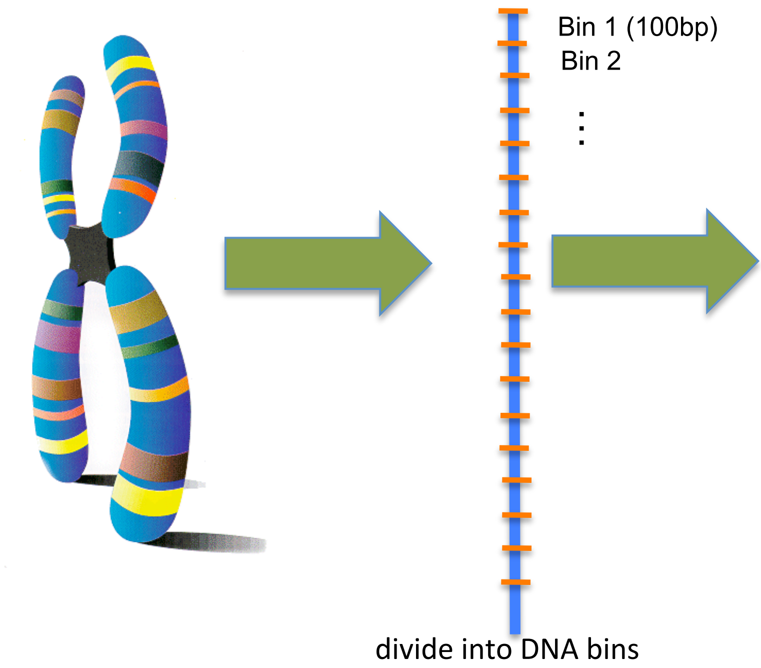
- Different Regions of Influence for TFs vs HMs



# Insights from worm modencode: Approaches useful for human annotation (outline)

- **Expression Timecourse Analysis**
  - Coordinated binding & expression; E v L separation; ~280 large splicing changes
- **ncRNAs [Hum]**
  - Importance of evidence integration
  - Large numbers of transcribed pseudogenes (8-15%)
- **Chromosomal activity distribution [Hum]**
  - Most constrained regions active
  - Repressed arms & binding HOT spots.
- **Regulatory Net [Hum]**
  - Arranging TF binding into a hierarchy with differences betw. levels. Integration with miRNA regulation (more at top).
  - Network motifs & prevalence of FFLs
- **Stat Models relating HMs, TFs & Expression [Hum]**
  - HMs statistically predict expression for protein-coding genes and miRNAs
  - Similar results for TFs, highlighting predictive power of a few TFs.
  - **Chromatin model (+ PWM) effective in predicting TF sites**
    -- useful in identifying enhancers.

# Chromatin model: link histone modification patterns to TF binding



ChIP-seq Data

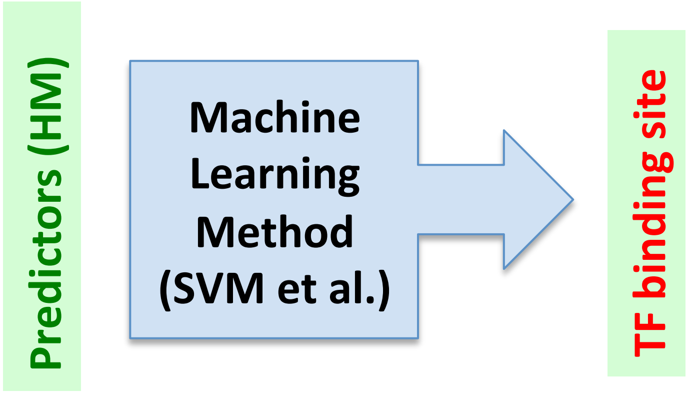
Predictors (HM)

|  | HM1 | HM2 | HM3 | …… |
|---|---|---|---|---|
| **Bin1** | 0.2 | 0.4 | 0.6 | … |
| **Bin2** | 0.3 | 0.3 | 0.2 | … |
| **Bin3** | 0 | 0.2 | 2.1 | … |
| **Bin4** | 0.4 | 0.4 | 0 | … |
| **Bin5** | 0.3 | 1.2 | 0.5 | … |
| **Bin6** | 1.2 | 3.1 | 2.1 | … |
| **Bin7** | 3.4 | 2.4 | 0.8 | … |
| **Bin8** | 1.5 | 1.2 | 0.9 | … |
| **…….** | … | … | … | … |

+

TF binding site?

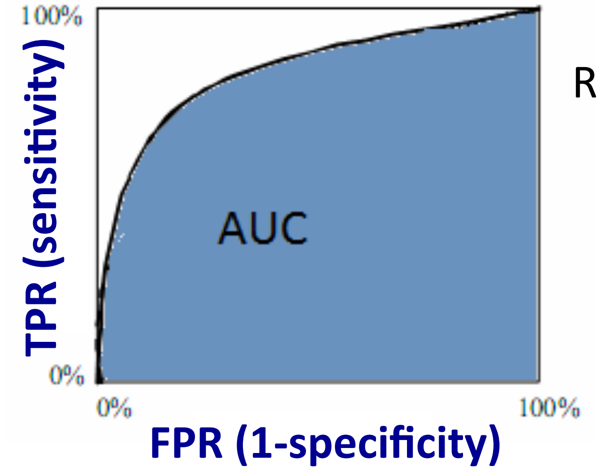
|  | TF1 | TF2 | …… |
|---|---|---|---|
| **Bin1** | 0 | 0 | … |
| **Bin2** | 0 | 0 | … |
| **Bin3** | 1 | 0 | … |
| **Bin4** | 1 | 0 | … |
| **Bin5** | 0 | 1 | … |
| **Bin6** | 0 | 0 | … |
| **Bin7** | 0 | 1 | … |
| **Bin8** | 1 | 0 | … |
| **…….** | … | … | … |



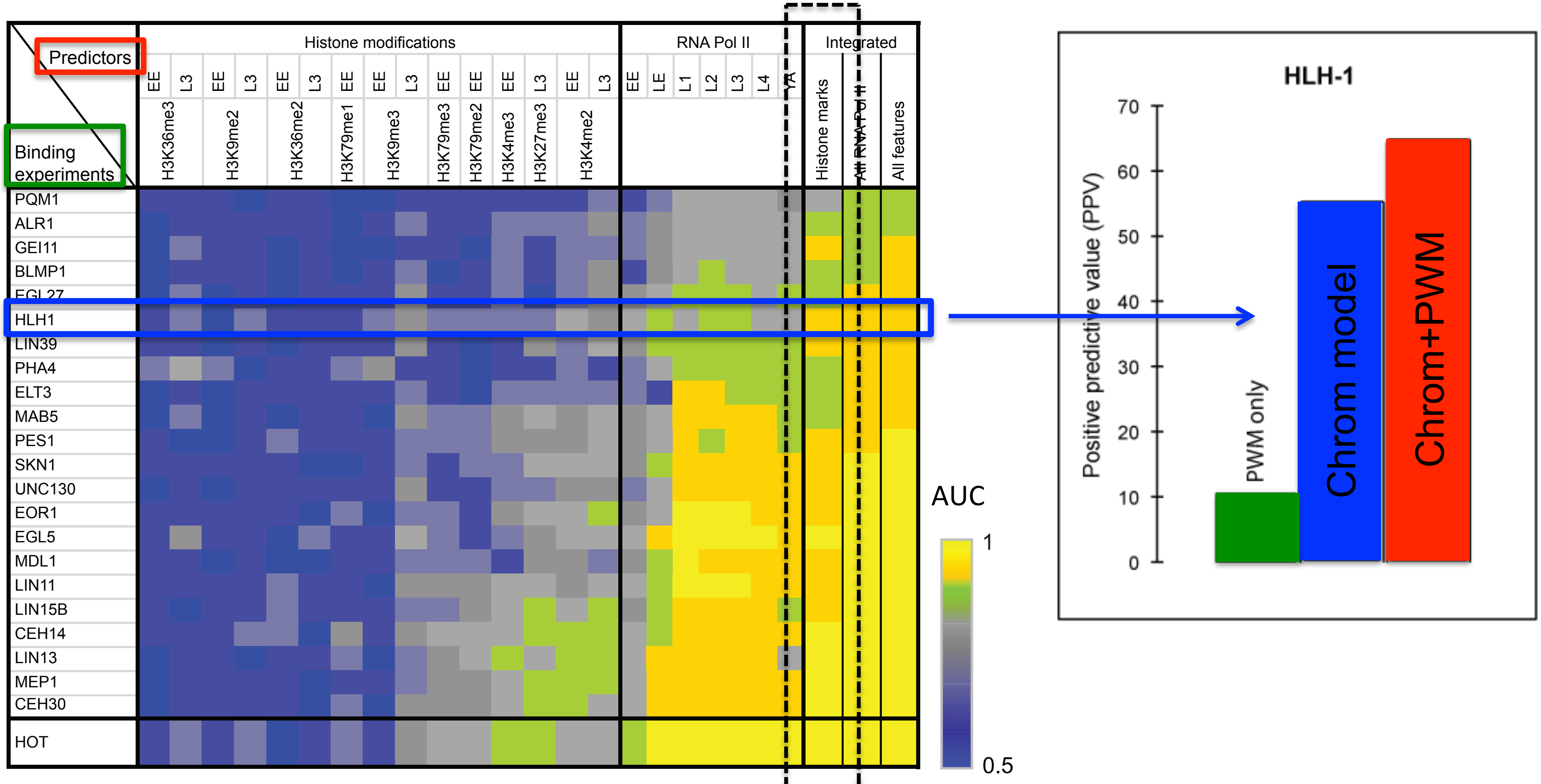
Predictors (HM)

Machine Learning Method (SVM et al.)

TF binding site



ROC curve

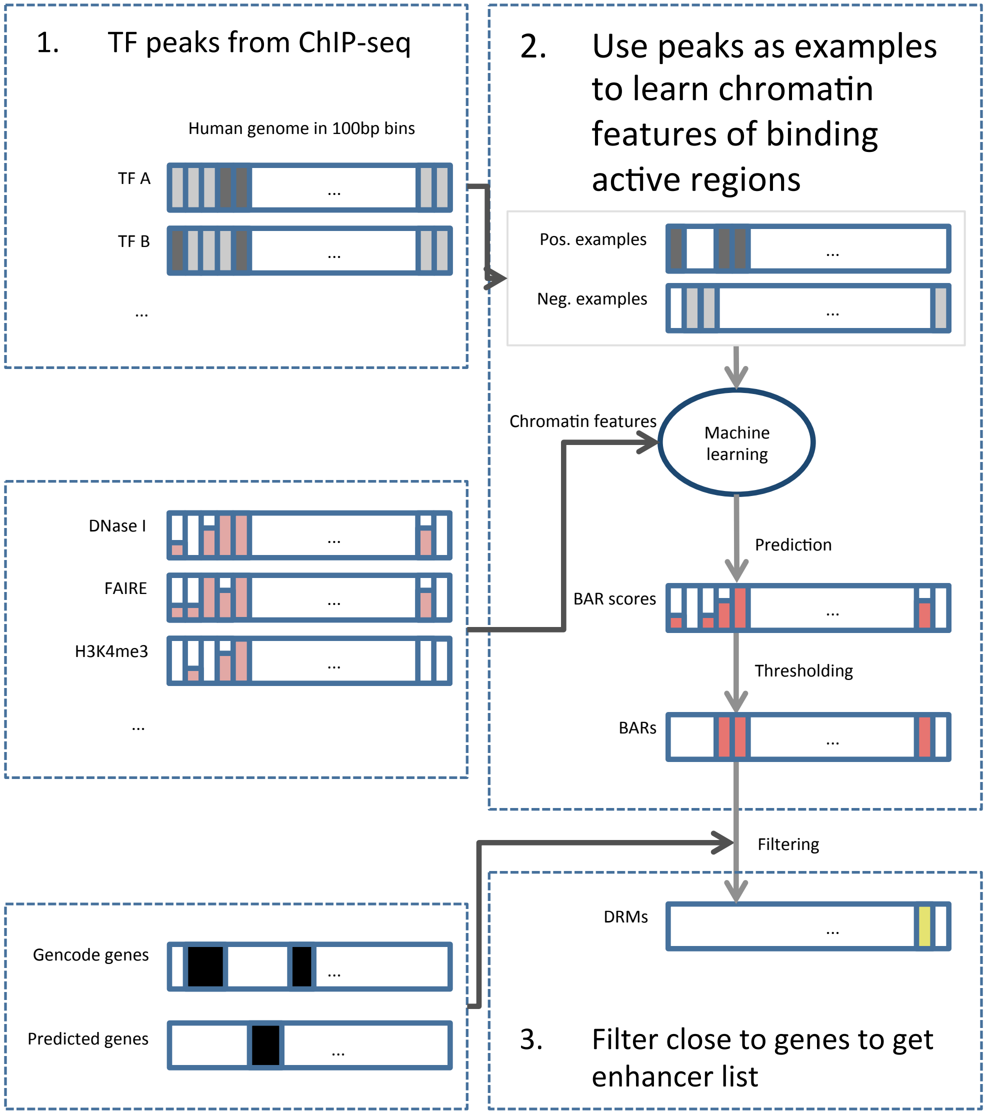


# Relating the Chromatin Model to TF Binding Sites in worm

- Model Predicts Sites Fairly Accuracy
- Accuracy improved when coupled with PWM

[ ]

## Identifying potential enhancers

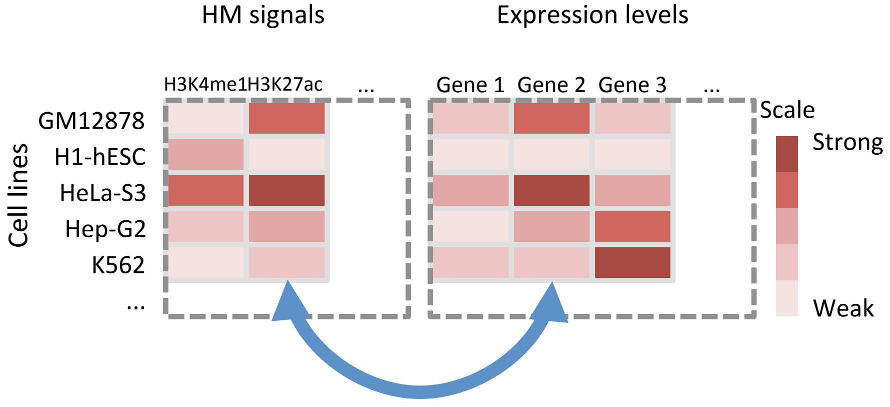


1. TF peaks from ChIP-seq

2. Use peaks as examples to learn chromatin features of binding active regions

3. Filter close to genes to get enhancer list

## Finding potential target genes



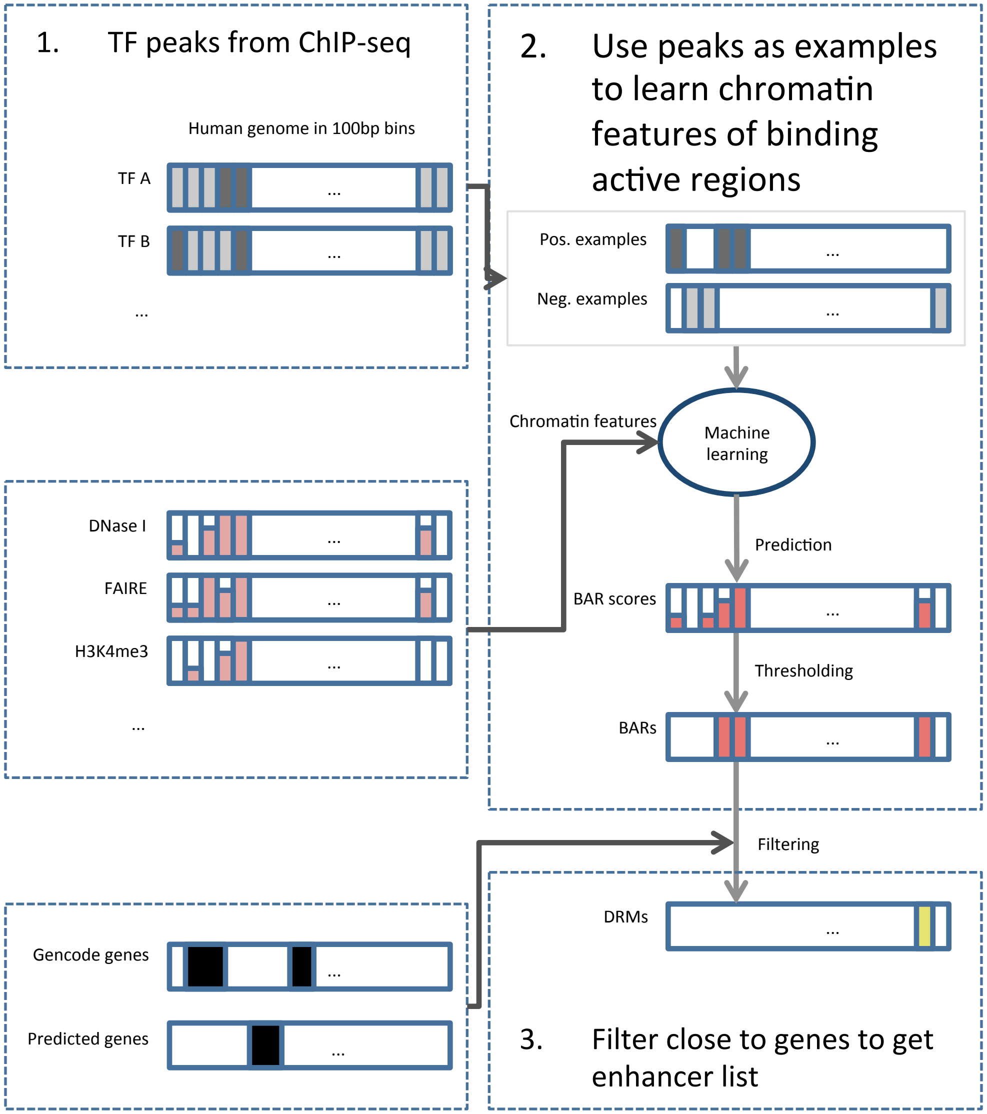
4. Find correlated enhancer-target pairs

~20K distal edges tot. from
~130K enhancer-like elements

(Related to but more “targeted” than enhancer “states” from unsupervised segmentation,
M Hoffman et al. & J Ernst et al.)

## Identifying Potential Enhancer-like Elements from Discriminative HM Model & then Linking these to Targets (via cell-line correlations) to Create Distal Edges

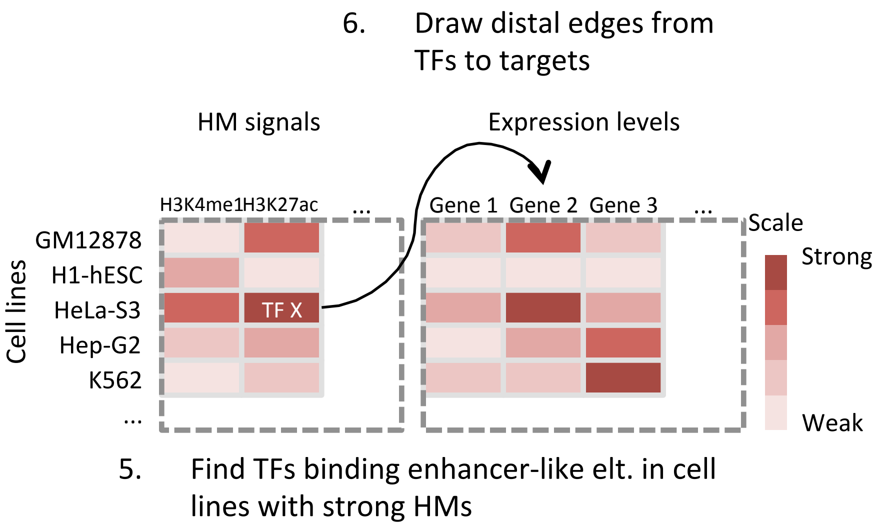
## Identifying potential enhancers



1. TF peaks from ChIP-seq

2. Use peaks as examples to learn chromatin features of binding active regions

3. Filter close to genes to get enhancer list



6. Draw distal edges from TFs to targets

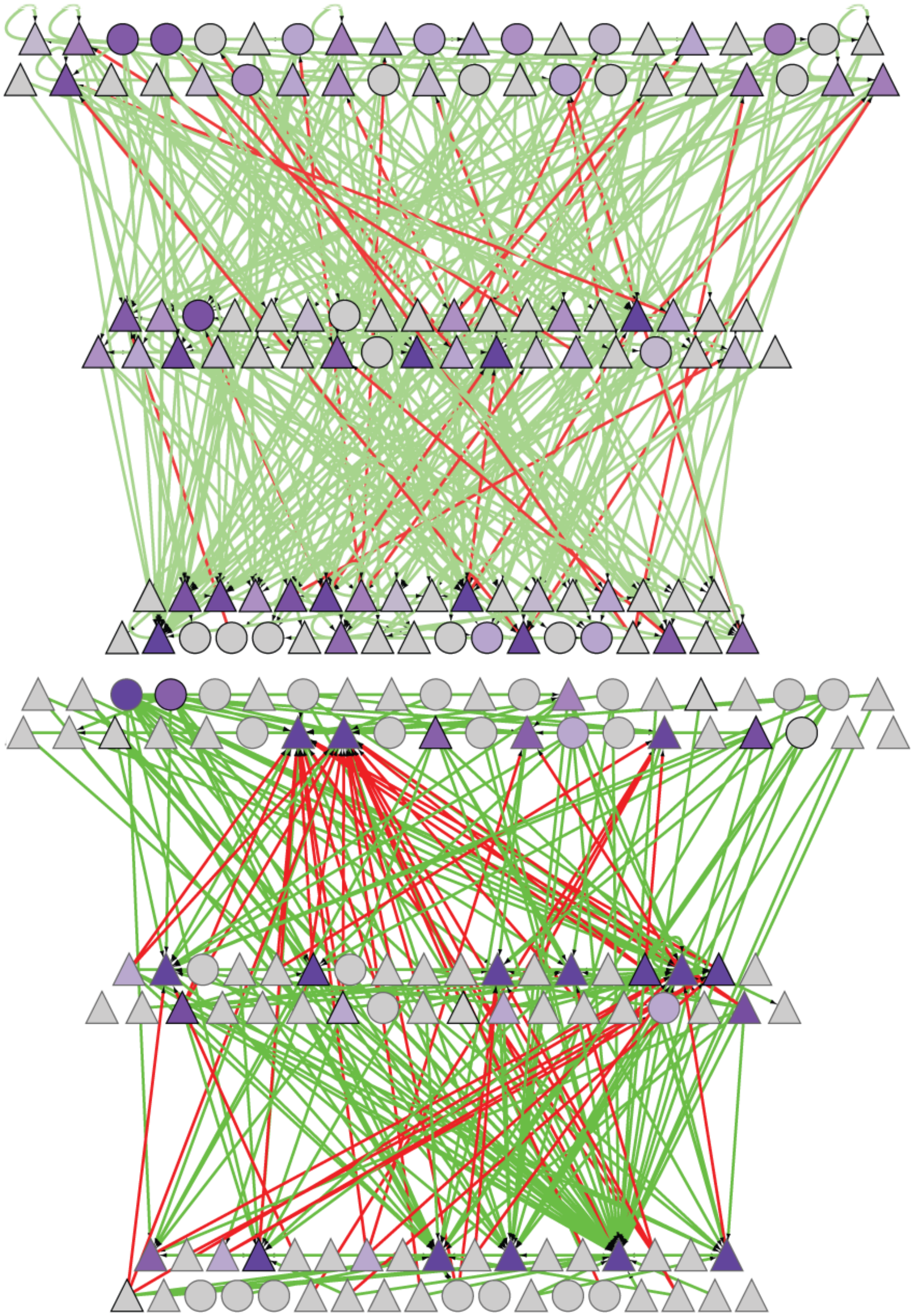
5. Find TFs binding enhancer-like elt. in cell lines with strong HMs

~20K distal edges tot. from ~130K enhancer-like elements

(Related to but more “targeted” than enhancer “states” from unsupervised segmentation,
M Hoffman et al. & J Ernst et al.)

## Identifying Potential Enhancer-like Elements from Discriminative HM Model & then Linking these to Targets (via cell-line correlations) to Create Distal Edges

[ ]



# Comparing

# Proximal

# &

# Distal

# Networks

[ ]

# Insights from worm modencode: Approaches useful for human annotation (outline)

- **Expression Timecourse Analysis**
  - Coordinated binding & expression; E v L separation; ~280 large splicing changes
- **ncRNAs [Hum]**
  - Importance of evidence integration
  - Large numbers of transcribed pseudogenes (8-15%)
- **Chromosomal activity distribution [Hum]**
  - Most constrained regions active
  - Repressed arms & binding HOT spots.
- **Regulatory Net [Hum]**
  - Arranging TF binding into a hierarchy with differences betw. levels. Integration with miRNA regulation (more at top).
  - Network motifs & prevalence of FFLs
- **Stat Models relating HMs, TFs & Expression [Hum]**
  - HMs statistically predict expression for protein-coding genes and miRNAs
  - Similar results for TFs, highlighting predictive power of a few TFs.
  - Chromatin model (+ PWM) effective in predicting TF sites
    -- useful in identifying enhancers.

ENCODE

Acknowledgements

(11 Main Projects, ~50 labs, >350 substantial contributors + NHGRI)

Nets/ Elements

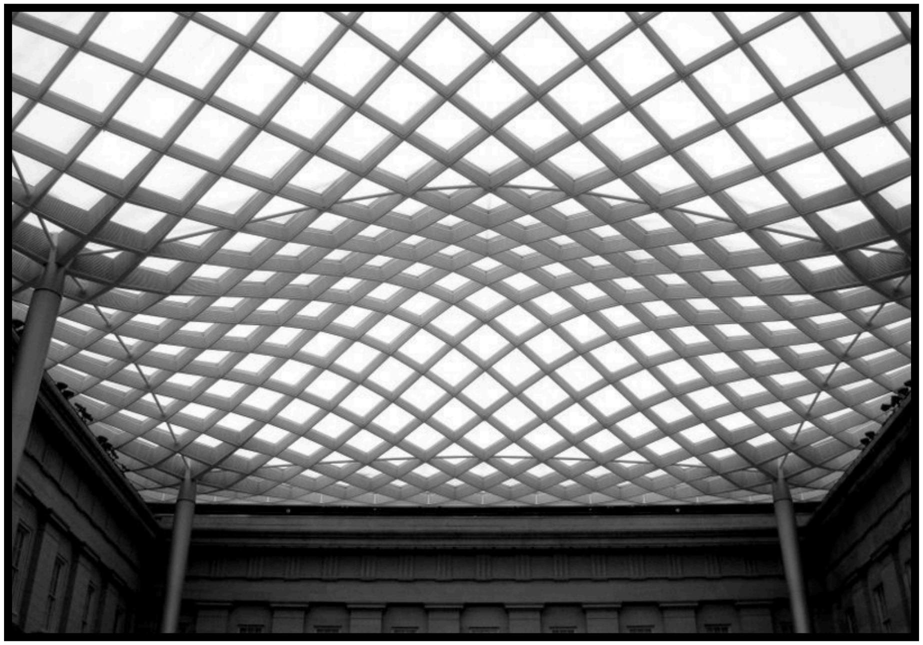
GENCODE

**A Kundaje, M Hariharan, S Landt, K Yan, C Cheng, X Mu, E Khurana, J Rozowsky, R Alexander, R Min, P Alves,** A Abyzov, N Addleman, N Bhardwaj, A Boyle, P Cayting, A Charos, D Chen, Y Cheng, D Clarke, C Eastman, G Euskirchen, S Frietze, Y Fu, J Gertz, F Grubert, A Harmanci, P Jain, M Kasowski, P Lacroute, J Leng, J Lian, H Monahan, H O'Geen, Z Ouyang, E Partridge, D Patacsil, F Pauli, D Raha, L Ramirez, T Reddy, B Reed, M Shi, T Slifer, J Wang, L Wu, X Yang, K Yip, G Zilberman-Schapira, S Batzoglou, A Sidow, P Farnham, R Myers, S Weissman, **M Snyder**

+ (for enhancers) E Birney, P Bickel, & B Brown

**Jennifer Harrow, Adam Frankish,** Jose M. Gonzalez , Electra Tapanari , Mark Diekhans, Felix Kokocinski, Bronwen Aken, Daniel Barrell, Amonida Zadissa, Stephen Searle, If Barnes, Alexandra Bignell, Veronika Boychenko, Toby Hunt, Mike Kay, Gaurab Mukherjee, Jeena Rajan, Gloria Despacio-Reyes, Gary Saunders, Charles Steward, Rachel Harte, Michael Lin, Cédric Howald, Andrea Tanzer, Thomas Derrien, Jacqueline Chrast, Nathalie Walters , **Suganthi Balasubramanian, Baikang Pei, Cristina Sisu,** Michael Tress, Jose Manuel Rodriguez, Iakes Ezkurdia, Jeltje van Baren , Michael Brent, David Haussler, Manolis Kellis, Alfonso Valencia, Alexandre Reymond, Mark Gerstein, Roderic Guigo and **Tim J. Hubbard.**

# ModENCODE Acknowlegements



**Zhi Lu, Eric L. Van Nostrand, Chao Cheng, Bradley I. Arshinoff, Tao Liu, K Yip,**

R Robilotto, Andreas Rechtsteiner, Kohta Ikegami, Pedro Alves, Aurelien Chateigner, Marc Perry, Mitzi Morris, Raymond K. Auerbach, Xin Feng,

**Jing Leng,** Anne Vielle, Wei Niu, Kahn Rhrissorrakra, Ashish Agarwal, Roger P. Alexander, Galt Barber, Cathleen M. Brdlik, Jennifer Brenna, Jeremy Jean Brouillet, Adrian Carr, Ming-Sin Cheung, Hiram Clawson, Sergio Contrino, Luke O. Dannenberg, Abby F. Dernburg, Arshad Desai, Lindsay Dick, Andréa C. Dosé, Jiang Du, Thea Egelhofer, Sevinc Ercan, Ghia Euskirchen, Brent Ewing, **Elise A. Feingold,** Reto Gassman, **Peter J. Good,** Phil Green, Francois Gullier, Michelle Gutwein, Mark S. Guyer, **L Habegger,** Ting Han, Jorja G. Henikoff, Stefan R. Henz, Angie Hinrichs, Heather Holster, Tony Hyman, A. Leo Iniguez, Judith Janette, Morten Jensen, Masaomi Kato, W. James Kent, Ellen Kephart, Vishal Khivansara, E Khurana, John K. Kim, Paulina Kolasinska-Zwierz, Eric C. Lai, Isabel Latorre, Amber Leahey, Suzanna Lewis, Paul Lloyd, Lucas Lochovsky, Rebecca F. Lowdon, Yaniv Lubling, Rachel Lyne, Michael MacCoss, Sebastian D. Mackowiak, Marco Mangone, Sheldon McKay, Desirea Mecenas, Gennifer Merrihew, David M. Miller III, Andrew Muroyama, John I. Murray, Siew-Loon Ooi, Hoang Pham, Taryn Phippen, Elicia A. Preston, Nikolaus Rajewsky, Gunnar Rätsch, Heidi Rosenbaum, Joel Rozowsky, Kim Rutherford, Peter Ruzanov, Mihail Sarov, Rajkumar Sasidharan, Andrea Sboner, Paul Scheid, Eran Segal, Hyunjin Shin, Chong Shou, Frank J. Slack, Cindie Slightam, Richard Smith, William C. Spencer, E.O. Stinson, Scott Taing, Teruaki Takasaki, Dionne Vafeados, Ksenia Voronina, Guilin Wang, Nicole L. Washington, Christina Whittle, Beijing Wu, **K Yan,** Georg Zeller, Zheng Zha, Mei Zhong, Xingliang Zhou,

Julie Ahringer, Susan Strome, Kristin C. Gunsalus,

**Gos Micklem, X. Shirley Liu, Valerie Reinke, Stuart K. Kim, LaDeana W. Hillier,**
Steven Henikoff, Fabio Piano,

**M Snyder, L Stein, J Lieb, R Waterston**