Transcription Factors Across Life: Control of Gene Expression

Matt Weirauch
Center for Autoimmune Genomics and Etiology (CAGE)
Biomedical Informatics, Dev Bio
Cincinnati Children's Hospital

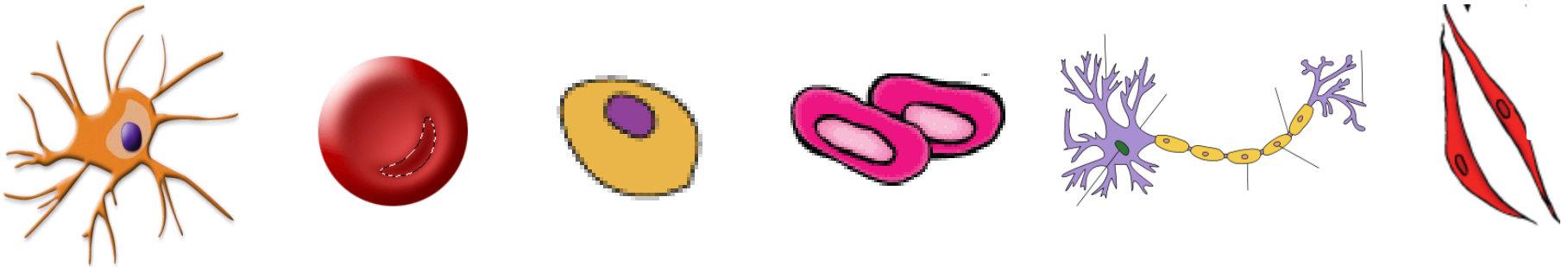
Some perspective…

- You are composed of ~100 trillion cells
- Each cell has an identical copy of your genome
- Within your genome, there are genes
  - Stretches of DNA that encode blueprints for making proteins
  - Humans have ~20,000 of them

Same genome/different cells

- Every cell has the same genomic DNA

- How can you explain the wide range of cellular form and function?

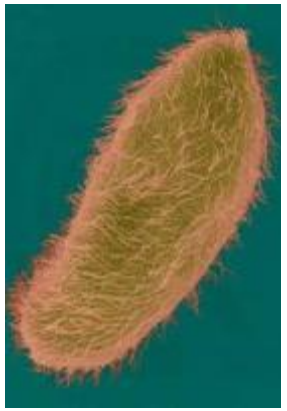


Same genes/different forms

- Life is incredibly diverse
- Organisms with different forms and lifestyles use highly similar genes
  - Human/Chimp DNA is >99% identical
  - Human and Mouse have the same basic set of genes







How is this so?
It's the utilization of genes

- Genes aren't all used at all times
- Certain genes function only in certain
  - Developmental stages
  - Environments
  - Tissue types
  - Cellular states
- How does the cell "know" whether or not to use a certain gene? – Gene regulation

Transcription Factors (TFs)

- Proteins that “turn on” and “turn off” the synthesis of genes in response to specific cues

Transcription Factors (TFs)

- Proteins that “turn on” and “turn off” the synthesis of genes in response to specific cues
  - Humans have ~1700 TFs

Transcription Factors (TFs)

- Proteins that “turn on” and “turn off” the synthesis of genes in response to specific cues
  - Humans have ~1700 TFs
  - Brain-specific TFs
  - Immune-responsive TFs
  - Developmental stage-specific TFs

Transcription Factors (TFs)

- Proteins that “turn on” and “turn off” the synthesis of genes in response to specific cues
  - Humans have ~1700 TFs
  - Brain-specific TFs
  - Immune-responsive TFs
  - Developmental stage-specific TFs

- Interact with the genome by binding to short sequences

Transcription Factors (TFs)

- Proteins that “turn on” and “turn off” the synthesis of genes in response to specific cues
  - Humans have ~1700 TFs
  - Brain-specific TFs
  - Immune-responsive TFs
  - Developmental stage-specific TFs

- Interact with the genome by binding to short sequences

- Each TF recognizes a specific DNA motif in the genome:

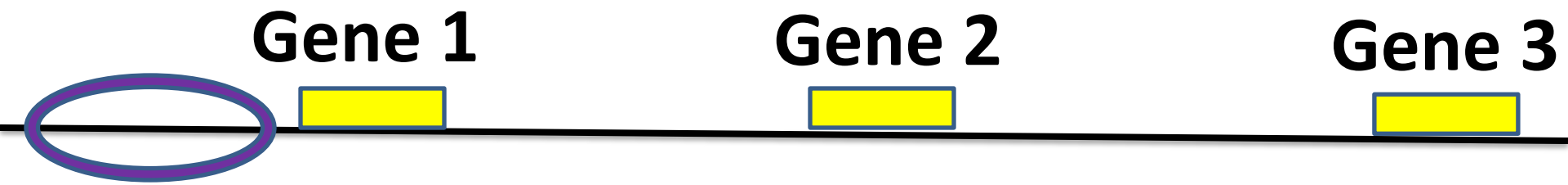



TF regulatory elements

- Usually located near the genes they control
- Humans have ~20,000 genes, all of which are controlled by multiple regulatory elements (Dozens? Hundreds?)
- In order to understand how/when/why genes are turned on and off, we need to know what regulatory elements TFs can recognize (motifs)

Quick summary

Quick summary (cont'd)



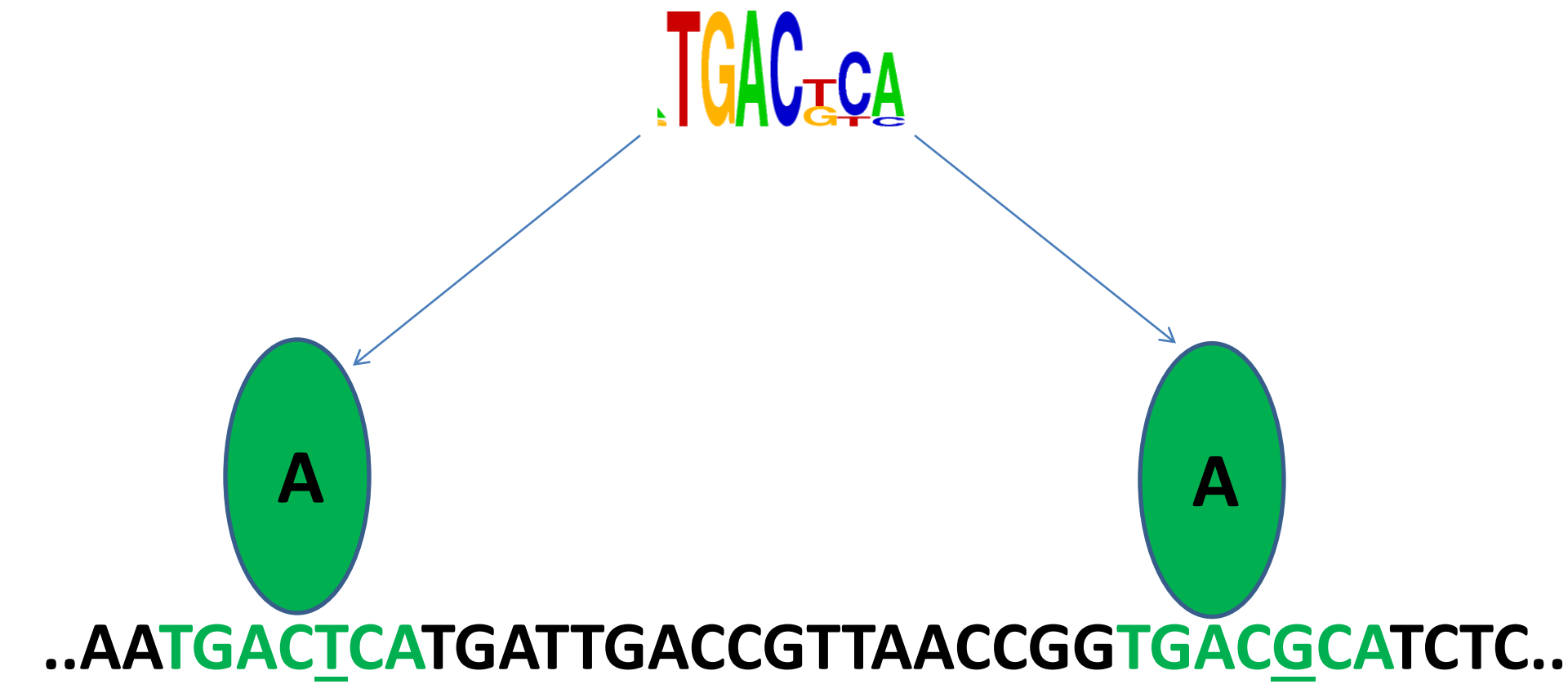
Quick summary (cont'd)

..AATGACTCATGATTGACCGTTAACCGGTGACGCATCTC..

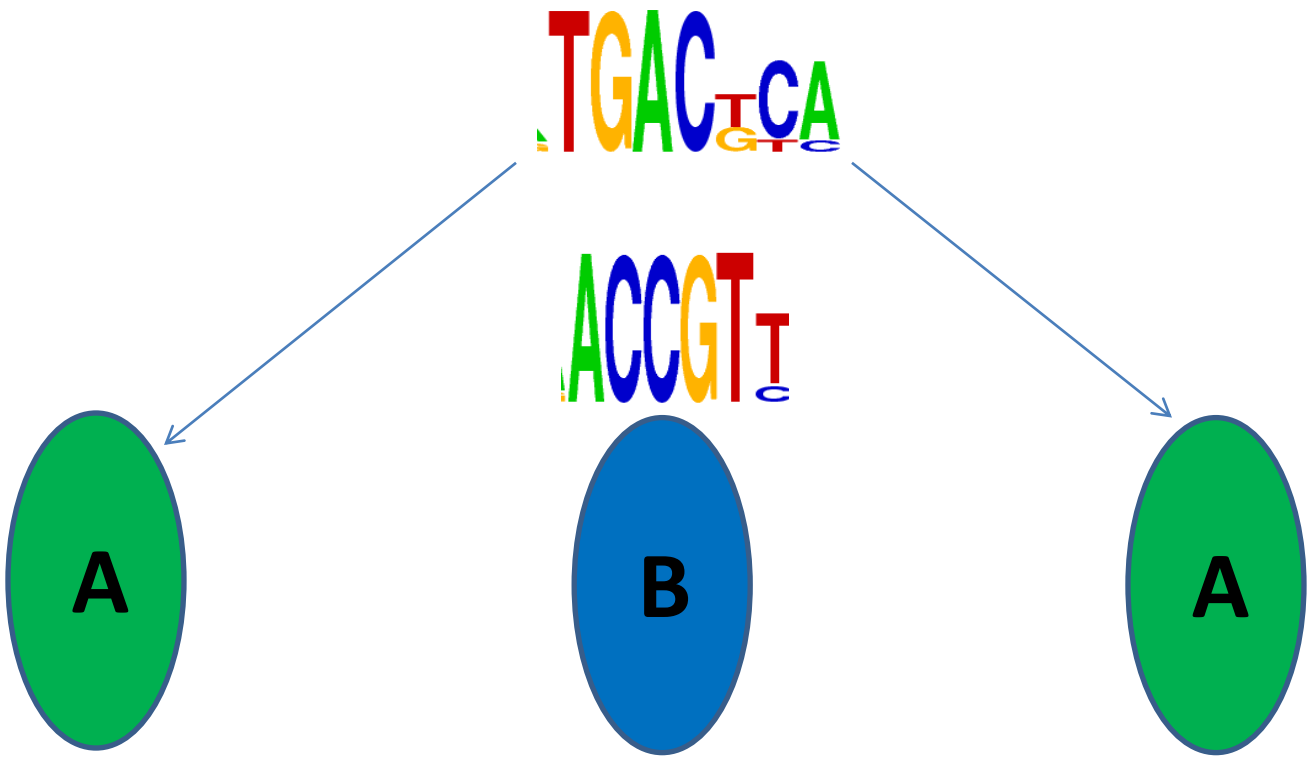
Quick summary (cont’d)

..AATGACTCATGATTGACCGTTAACCGGTGACGCATCTC..

Quick summary (cont'd)

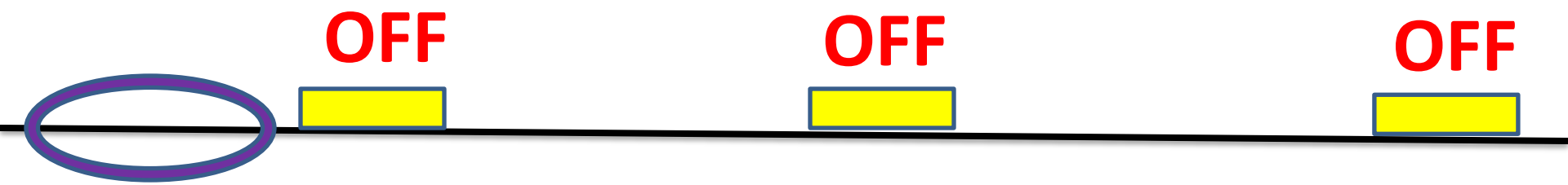


Quick summary (cont'd)

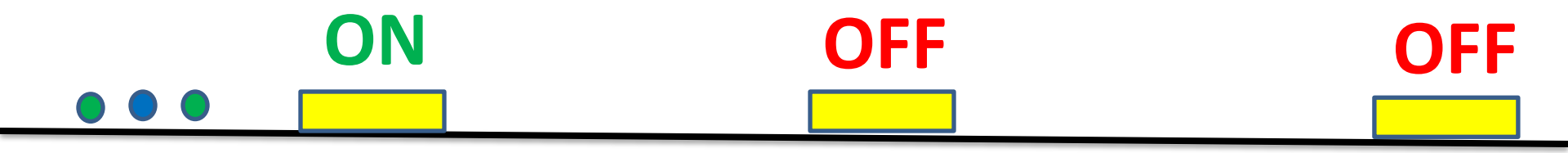


..AATGACTCATGATTGACCGTTAACCGGTGACGCATCTC..

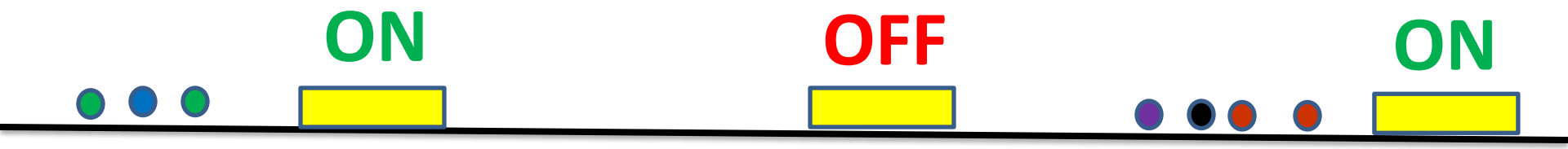
Quick summary (cont'd)

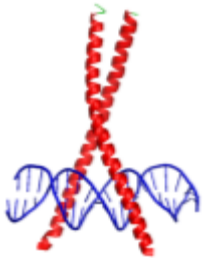


Quick summary (cont'd)

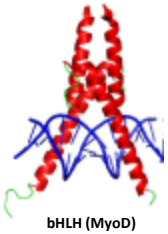


Quick summary (cont'd)

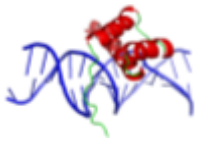




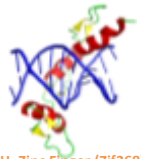
bZIP (C/EBPalpha)
1NWQ [12578822]



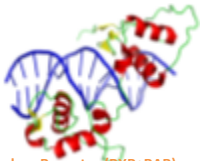
bHLH (MyoD)
1MDY [8181063]



Homeodomain (VND/NK-2)
1NK3 [9154919]

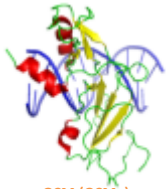


C_2H_2 Zinc Finger (Zif268)
1A1L [9562555]



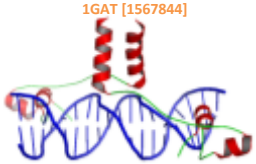
Nuclear Receptor (RXR+RAR)
1DSZ [10698945]

GATA (GATA-1)
1GAT [1567844]



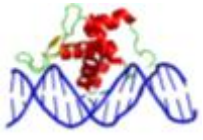
GCM (GCMa)
1ODH [12682016]

THAP (THAP)
3KDE [20010837]

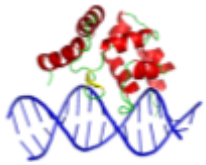


Zinc cluster (GAL4)
1D66 [1557122]

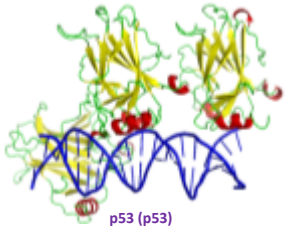
Myb/SANT (c-Myb)
1MSF [7954830]



Forkhead (Genesis)
2HDC [10369754]

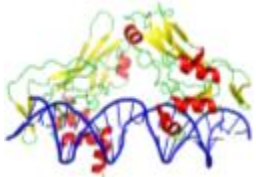


ARID (Dead ringer)
1KQQ [11867548]

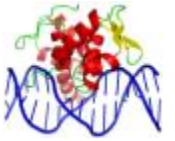


p53 (p53)
1TSR [8023157]

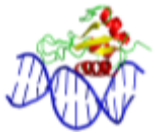
MADS box (SRF)
1SRS [7637780]



T-box (Bracyury/T)
1XBR [9349824]

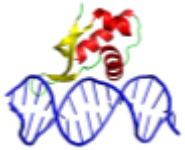


E2F+DP (E2F4+DP2)
1CF7 [10090723]

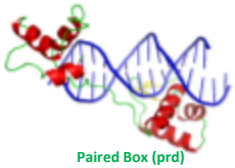


IRF (IRF-1)
1IF1 [9422515]

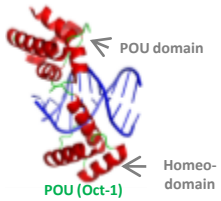
Ets (SAP-1)
1BC8 [9734357]



RFX (RFX1)
1DP7 [10706293]



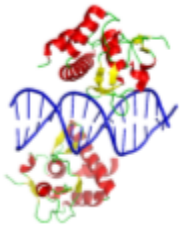
Paired Box (prd)
1PDN [7867071]



POU (Oct-1)
1E3O [11583619]

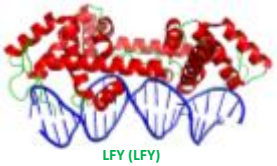
RHD (NFκB)
1VKX [9450761]

STAT (STAT1)
1BF5 [9630226]



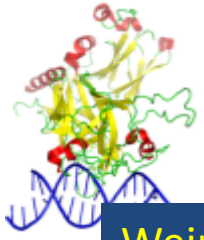
SMAD (Smad3)
1OZJ [12686552]

IBD (IBP39)
1PP7 [14622596]

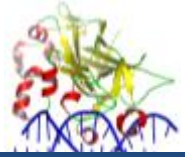


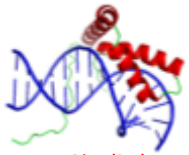
LFY (LFY)
2VY1 [18784751]

CUT (SATB1)
2O49 [17652321]

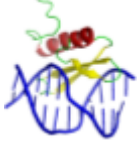


CSL (C
1TTU [152

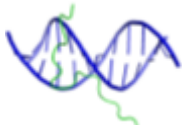




HMG box (SRY)
1J46 [11563911]

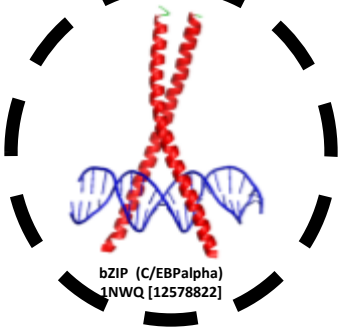


AP2 (ATERF1)
1GCC [9736626]

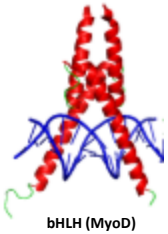


AT hook (HMG-I(Y))
2EZD [9253416]

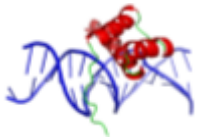
Weirauch and Hughes, 2011
In: A Handbook of Transcription Factors



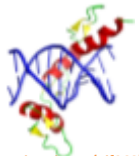
bZIP (C/EBPalpha)
1NWQ [12578822]



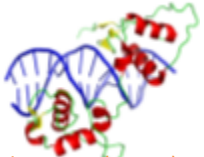
bHLH (MyoD)
1MDY [8181063]



Homeodomain (VND/NK-2)
1NK3 [9154919]

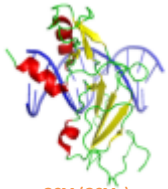


C_2H_2 Zinc Finger (Zif268)
1A1L [9562555]



Nuclear Receptor (RXR+RAR)
1DSZ [10698945]

GATA (GATA-1)
1GAT [1567844]

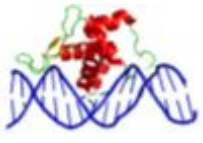


GCM (GCMa)
1ODH [12682016]

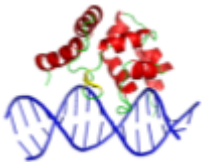
THAP (THAP)
3KDE [20010837]

Zinc cluster (GAL4)
1D66 [1557122]

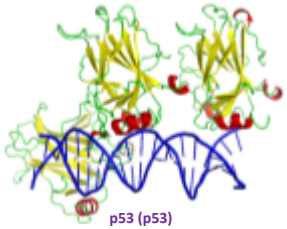
Myb/SANT (c-Myb)
1MSF [7954830]



Forkhead (Genesis)
2HDC [10369754]

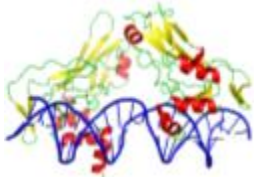


ARID (Dead ringer)
1KQQ [11867548]

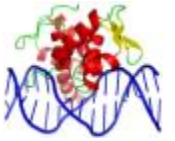


p53 (p53)
1TSR [8023157]

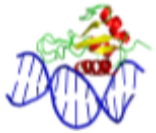
MADS box (SRF)
1SRS [7637780]



T-box (Bracyury/T)
1XBR [9349824]

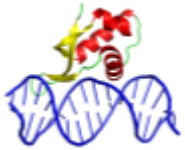


E2F+DP (E2F4+DP2)
1CF7 [10090723]

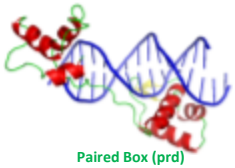


IRF (IRF-1)
1IF1 [9422515]

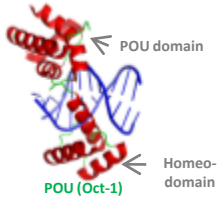
Ets (SAP-1)
1BC8 [9734357]



RFX (RFX1)
1DP7 [10706293]

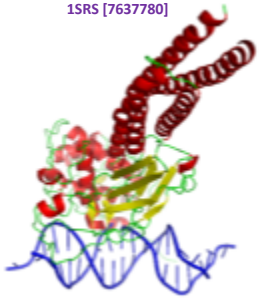


Paired Box (prd)
1PDN [7867071]

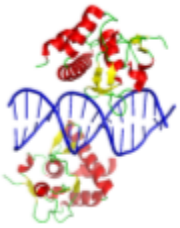


POU (Oct-1)
1E3O [11583619]

RHD (NFκB)
1VKX [9450761]

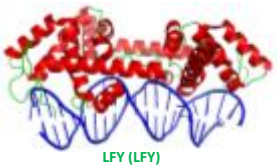


STAT (STAT1)
1BF5 [9630226]



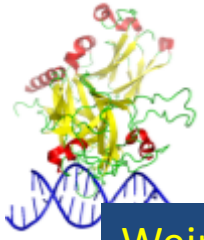
SMAD (Smad3)
1OZJ [12686552]

IBD (IBP39)
1PP7 [14622596]

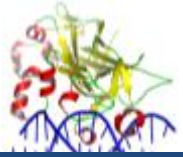


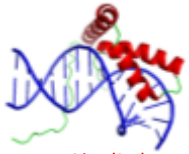
LFY (LFY)
2VY1 [18784751]

CUT (SATB1)
2O49 [17652321]

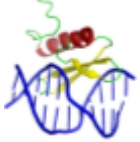


CSL (C
1TTU [152

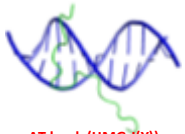




HMG box (SRY)
1J46 [11563911]

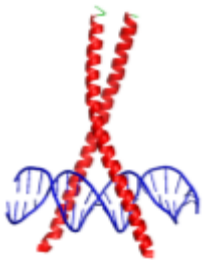


AP2 (ATERF1)
1GCC [9736626]

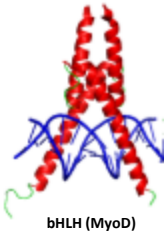


AT hook (HMG-I(Y))
2EZD [9253416]

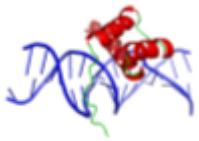
Weirauch and Hughes, 2011
In: A Handbook of Transcription Factors



bZIP (C/EBPalpha)
1NWQ [12578822]

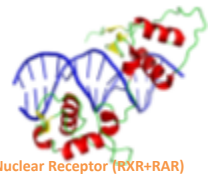


bHLH (MyoD)
1MDY [8181063]



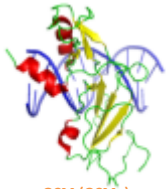
Homeodomain (VND/NK-2)
1NK3 [9154919]

C_2H_2 Zinc Finger (Zif268)
1A1L [9562555]



Nuclear Receptor (RXR+RAR)
1DSZ [10698945]

GATA (GATA-1)
1GAT [1567844]

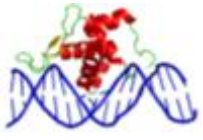


GCM (GCMa)
1ODH [12682016]

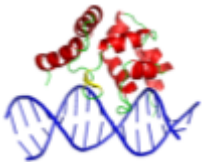
THAP (THAP)
3KDE [20010837]

Zinc cluster (GAL4)
1D66 [1557122]

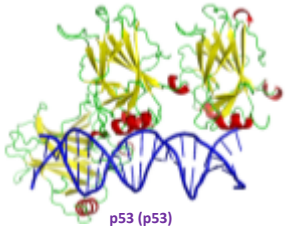
Myb/SANT (c-Myb)
1MSF [7954830]



Forkhead (Genesis)
2HDC [10369754]

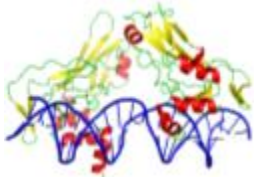


ARID (Dead ringer)
1KQQ [11867548]

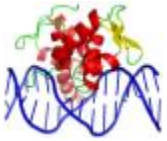


p53 (p53)
1TSR [8023157]

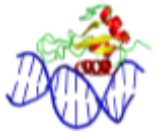
MADS box (SRF)
1SRS [7637780]



T-box (Bracyury/T)
1XBR [9349824]

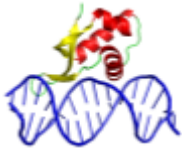


E2F+DP (E2F4+DP2)
1CF7 [10090723]

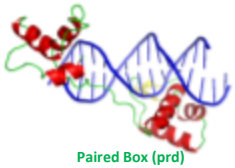


IRF (IRF-1)
1IF1 [9422515]

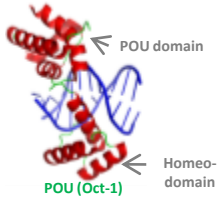
Ets (SAP-1)
1BC8 [9734357]



RFX (RFX1)
1DP7 [10706293]



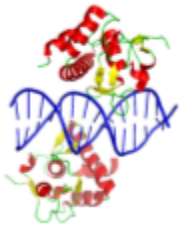
Paired Box (prd)
1PDN [7867071]



POU (Oct-1)
1E3O [11583619]

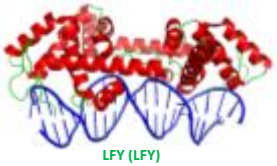
RHD (NFκB)
1VKX [9450761]

STAT (STAT1)
1BF5 [9630226]



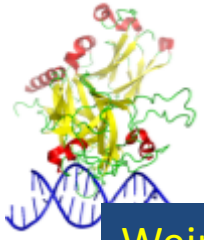
SMAD (Smad3)
1OZJ [12686552]

IBD (IBP39)
1PP7 [14622596]

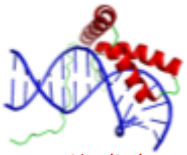


LFY (LFY)
2VY1 [18784751]

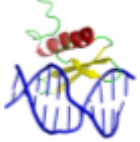
SATB1
2O49 [17652321]



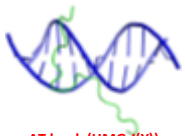
CSL (C
1TTU [152



HMG box (SRY)
1J46 [11563911]

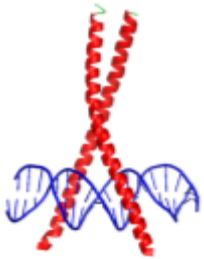


AP2 (ATERF1)
1GCC [9736626]

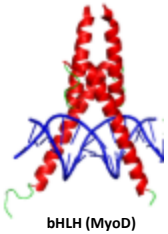


AT hook (HMG-I(Y))
2EZD [9253416]

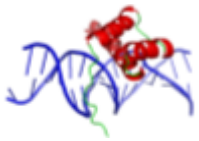
Weirauch and Hughes, 2011
In: A Handbook of Transcription Factors



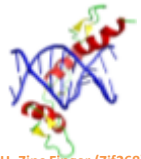
bZIP (C/EBPalpha)
1NWQ [12578822]



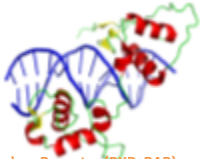
bHLH (MyoD)
1MDY [8181063]



Homeodomain (VND/NK-2)
1NK3 [9154919]



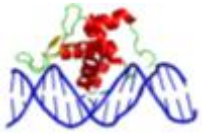
C_2H_2 Zinc Finger (Zif268)
1A1L [9562555]



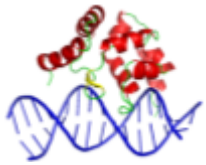
Nuclear Receptor (RXR+RAR)
1DSZ [10698945]

GATA (GATA-1)
1GAT [1567844]

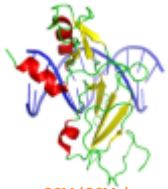
Myb/SANT (c-Myb)
1MSF [7954830]



Forkhead (Genesis)
2HDC [10369754]



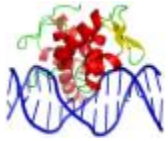
ARID (Dead ringer)
1KQQ [11867548]



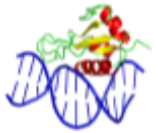
GCM (GCMa)
1ODH [12682016]

THAP (THAP)
3KDE [20010837]

Zinc cluster (GAL4)
1D66 [1557122]

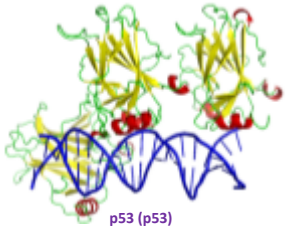


E2F+DP (E2F4+DP2)
1CF7 [10090723]



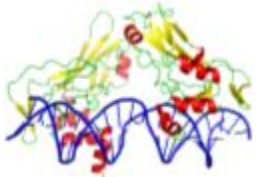
IRF (IRF-1)
1IF1 [9422515]

Ets (SAP-1)
1BC8 [9734357]

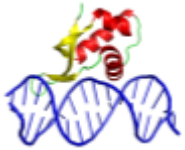


p53 (p53)
1TSR [8023157]

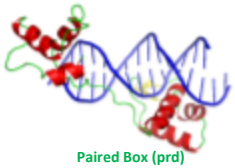
MADS box (SRF)
1SRS [7637780]



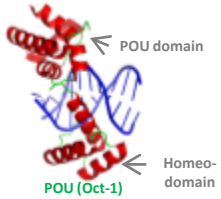
T-box (Bracyury/T)
1XBR [9349824]



RFX (RFX1)
1DP7 [10706293]

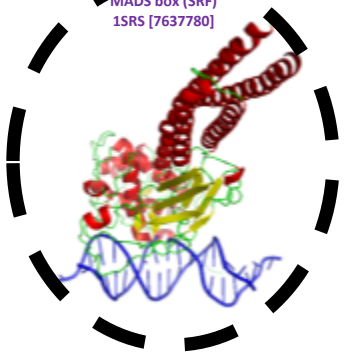


Paired Box (prd)
1PDN [7867071]

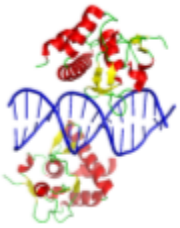


POU (Oct-1)
1E3O [11583619]

RHD (NFκB)
1VKX [9450761]

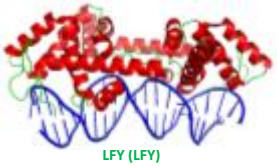


STAT (STAT1)
1BF5 [9630226]



SMAD (Smad3)
1OZJ [12686552]

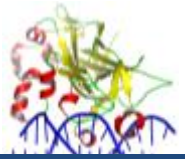
IBD (IBP39)
1PP7 [14622596]

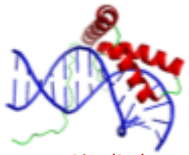


LFY (LFY)
2VY1 [18784751]

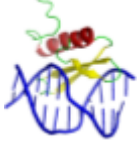
CUT (SATB1)
2O49 [17652321]

CSL (C
1TTU [152

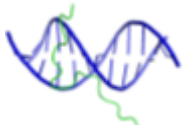




HMG box (SRY)
1J46 [11563911]



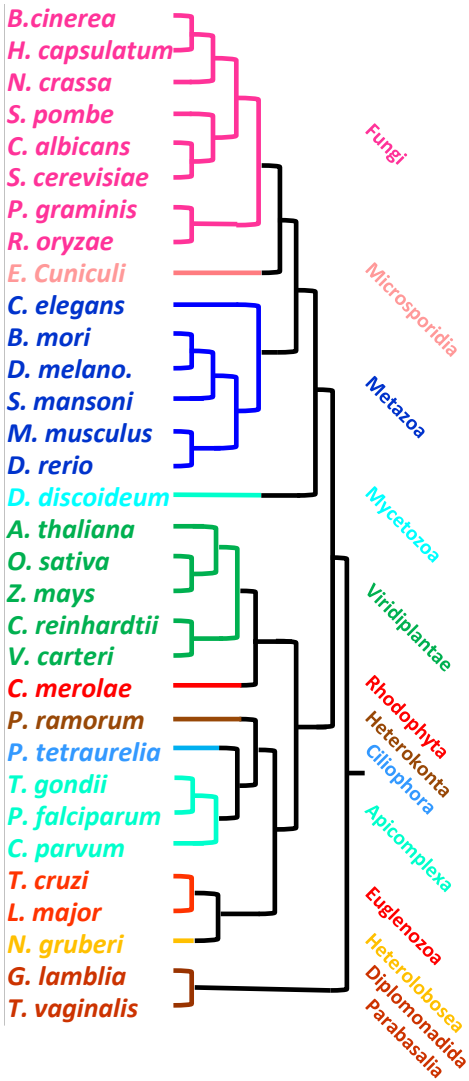
AP2 (ATERF1)
1GCC [9736626]



AT hook (HMG-I(Y))
2EZD [9253416]

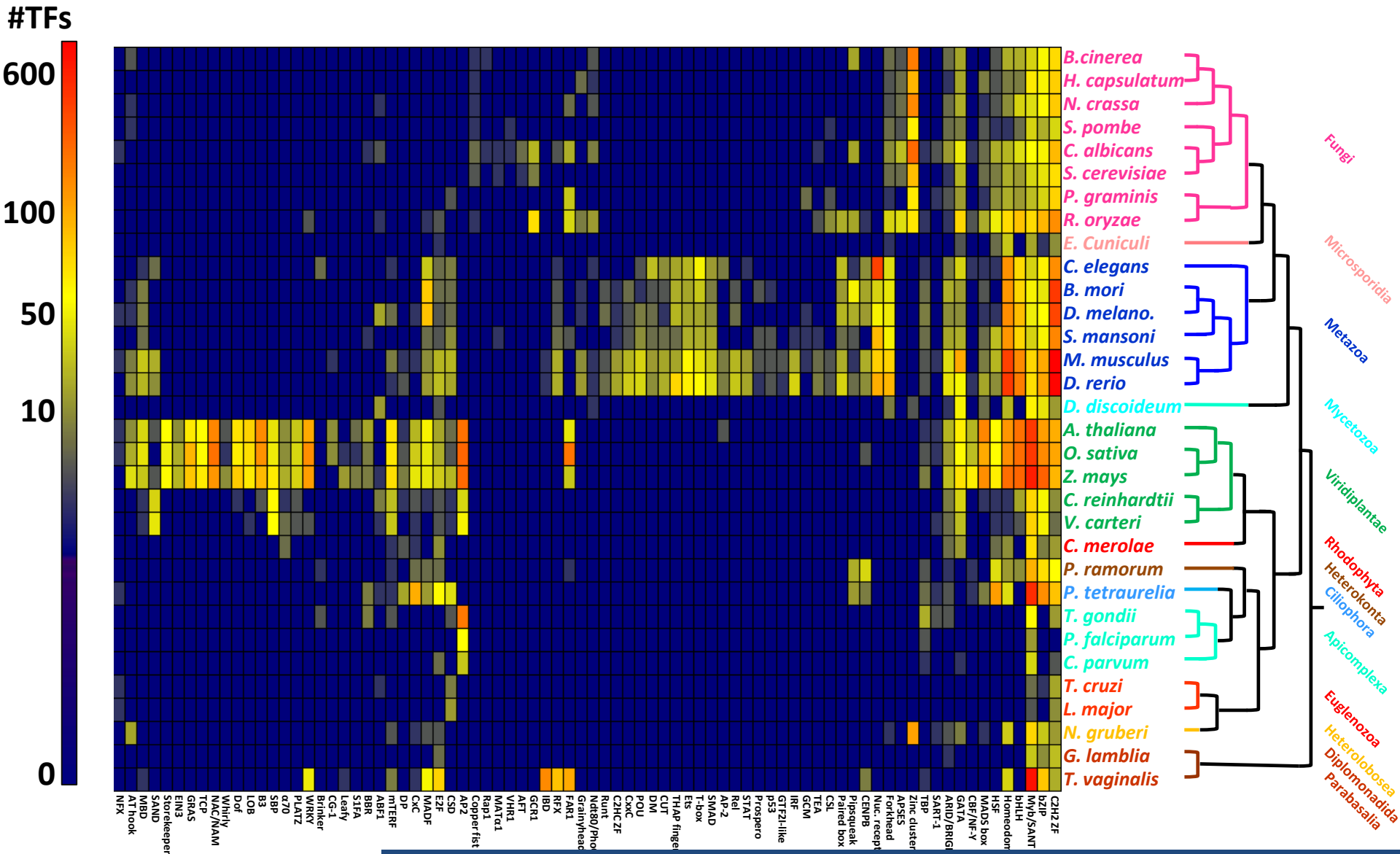
Weirauch and Hughes, 2011
In: A Handbook of Transcription Factors

TF families across Eukaryota



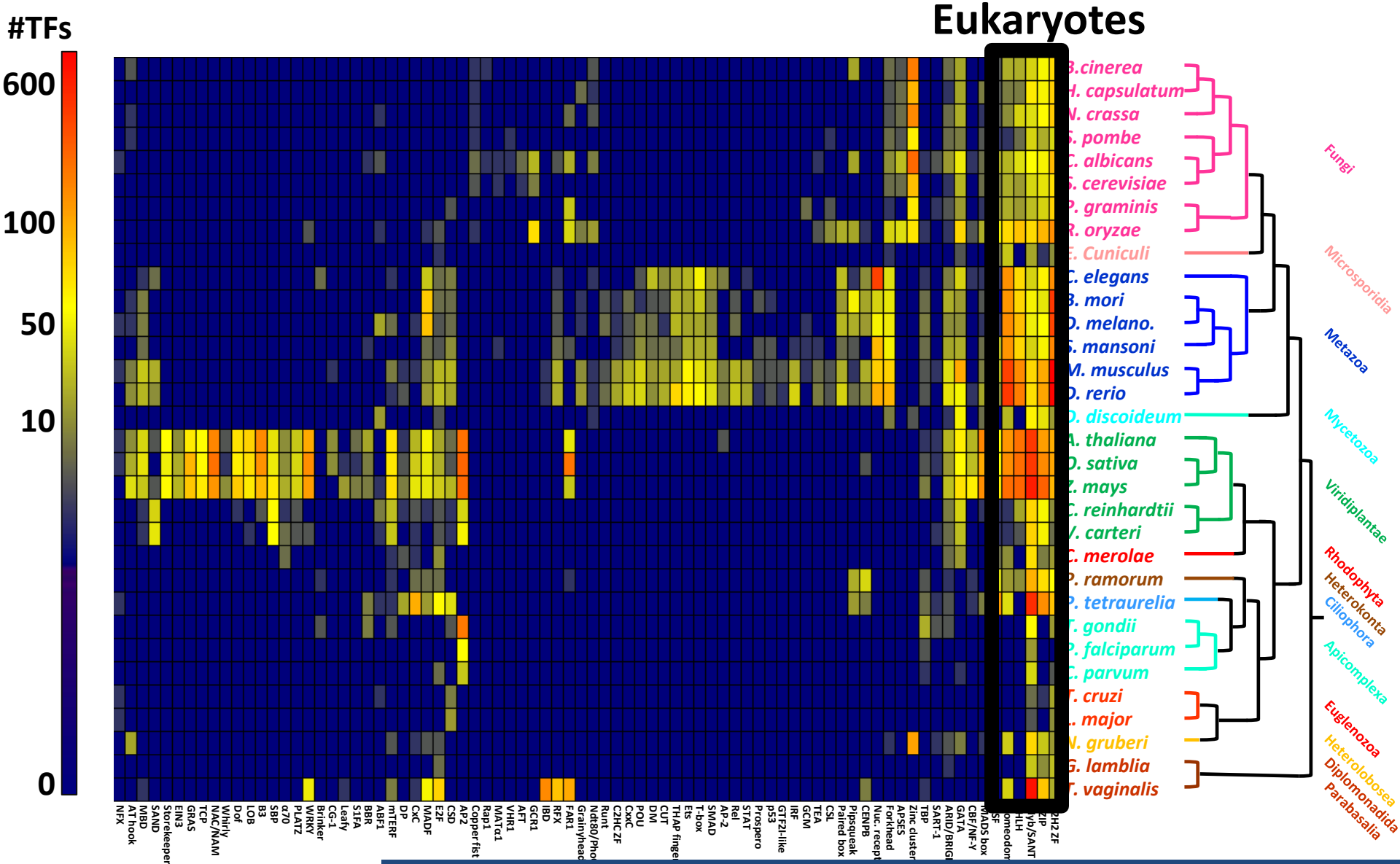
Weirauch and Hughes, 2011. In: A Handbook of Transcription Factors

TF families across Eukaryota



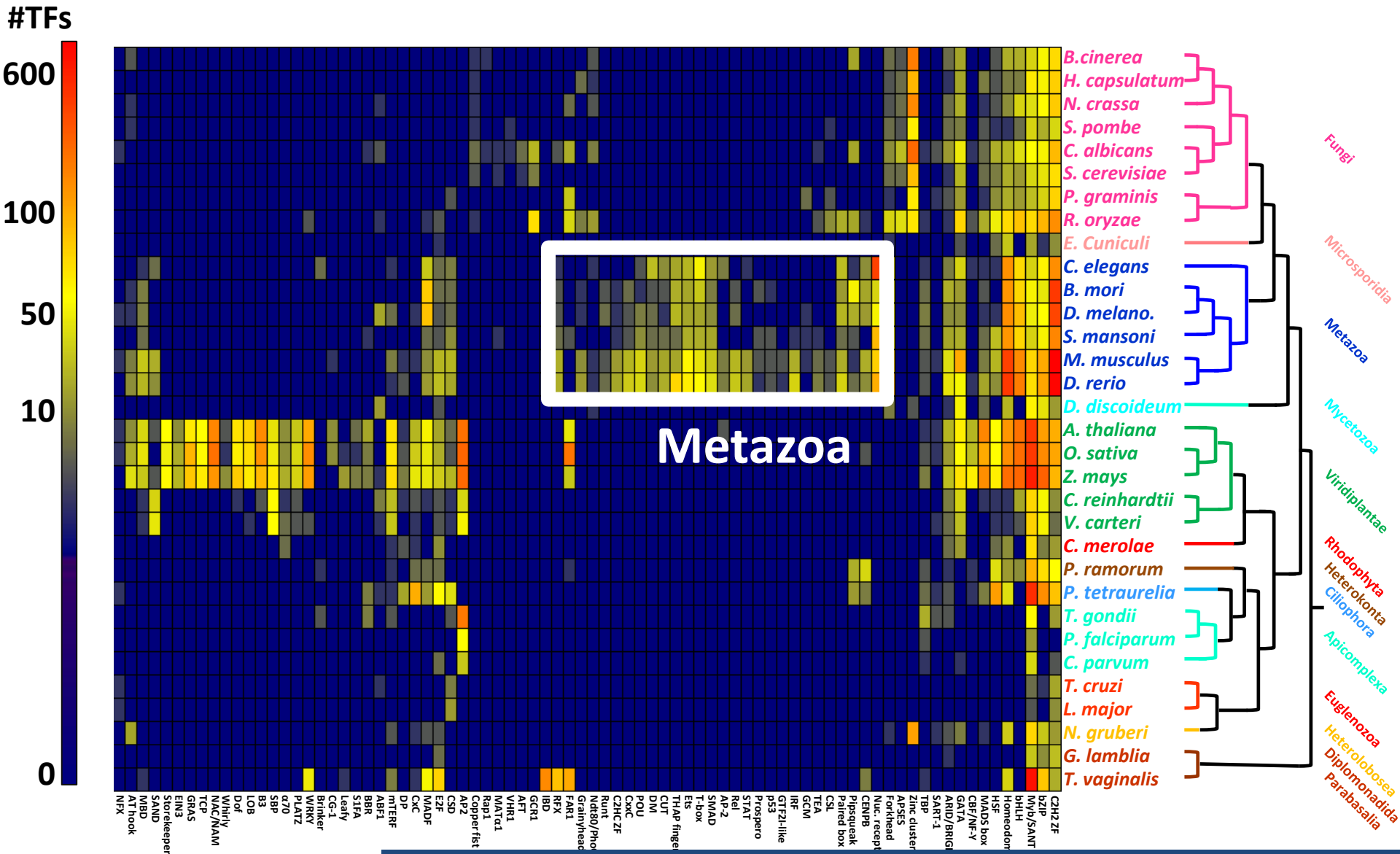
Weirauch and Hughes, 2011. In: A Handbook of Transcription Factors

TF families across Eukaryota



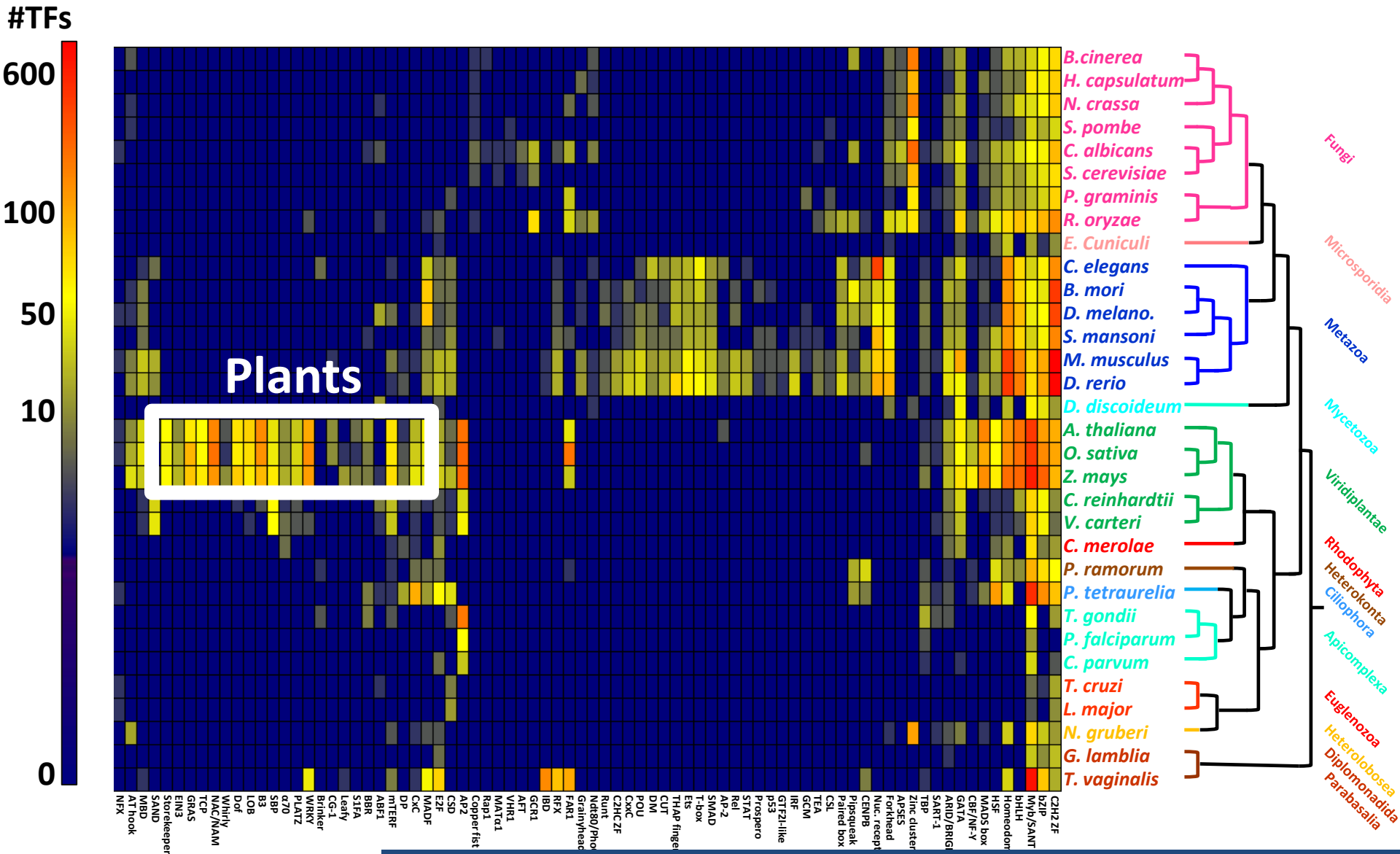
Weirauch and Hughes, 2011. In: A Handbook of Transcription Factors

TF families across Eukaryota



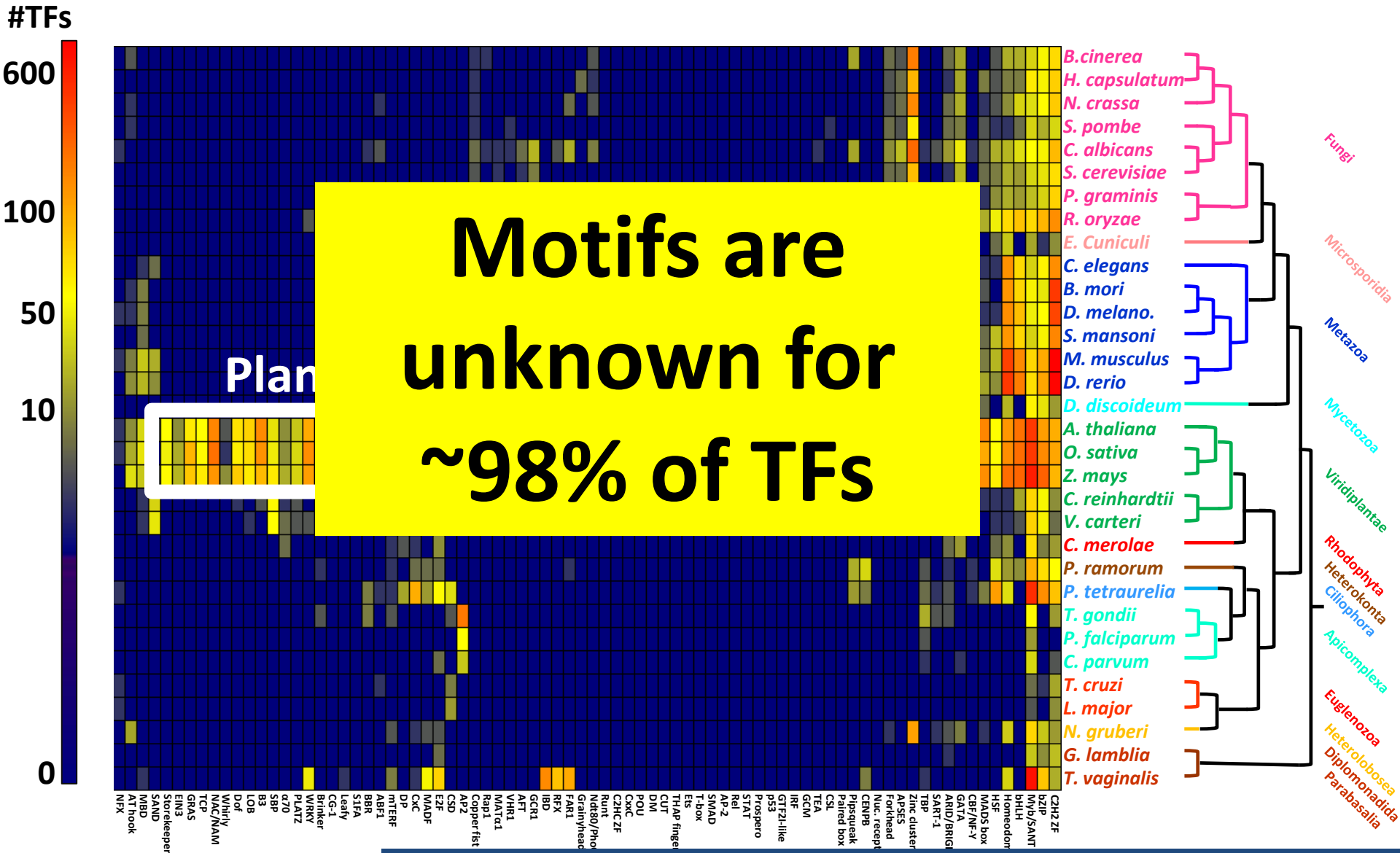
Weirauch and Hughes, 2011. In: A Handbook of Transcription Factors

TF families across Eukaryota



Weirauch and Hughes, 2011. In: A Handbook of Transcription Factors

TF families across Eukaryota



Weirauch and Hughes, 2011. In: A Handbook of Transcription Factors

Protein-binding microarrays (PBMs)

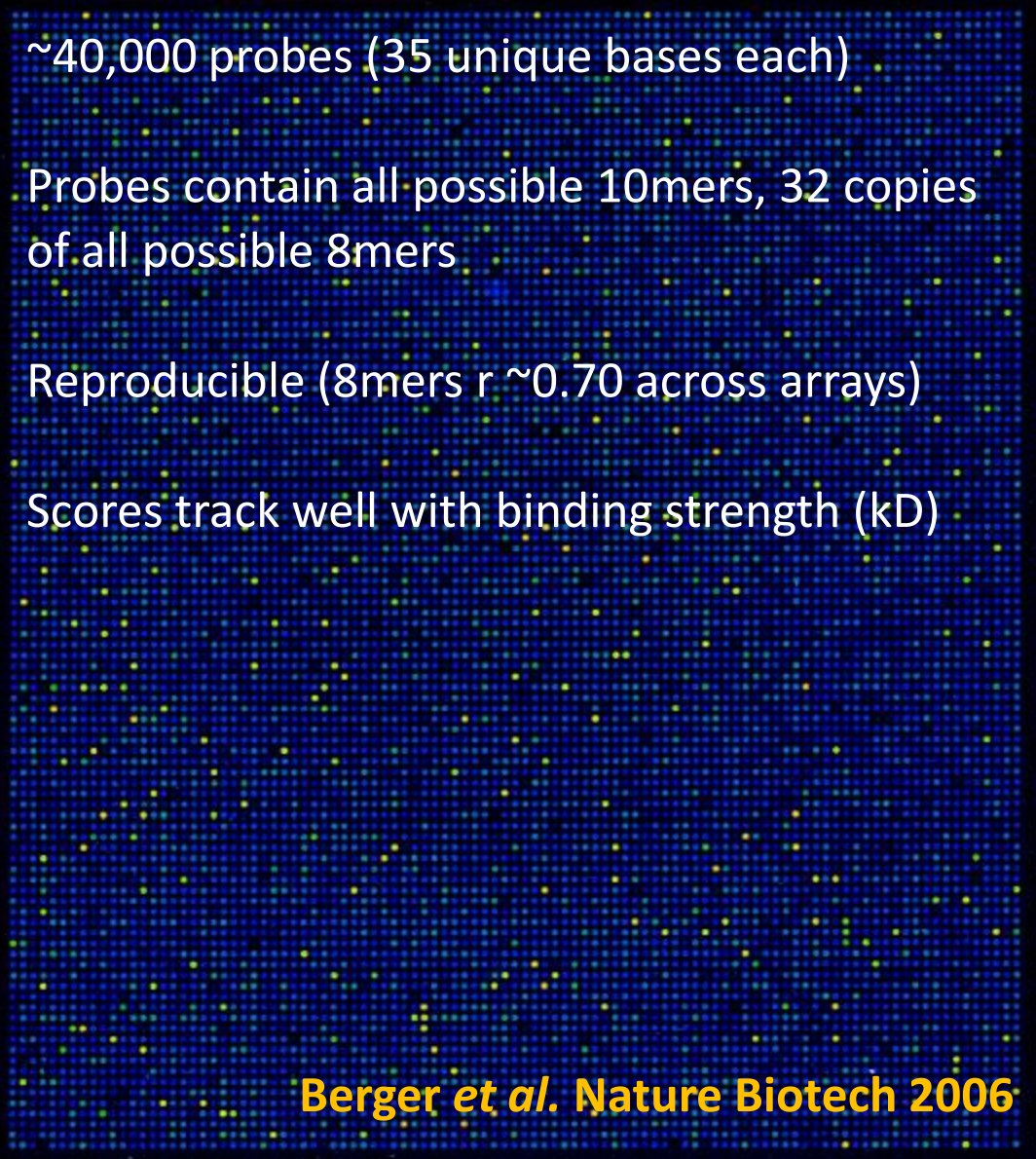
Make plasmid (DNA-binding domain)

↓

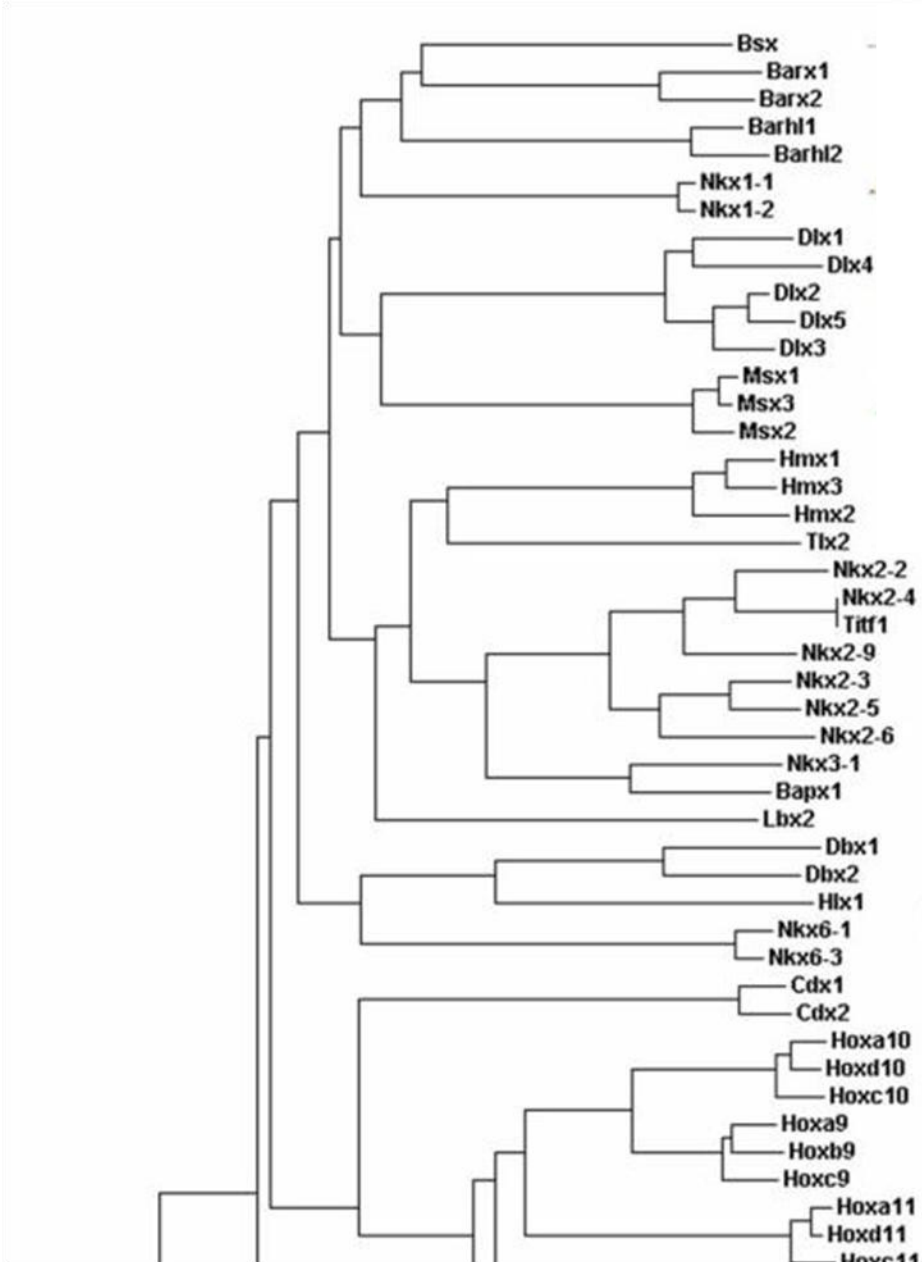
Bind to DNA microarray

↓

Determine binding preferences

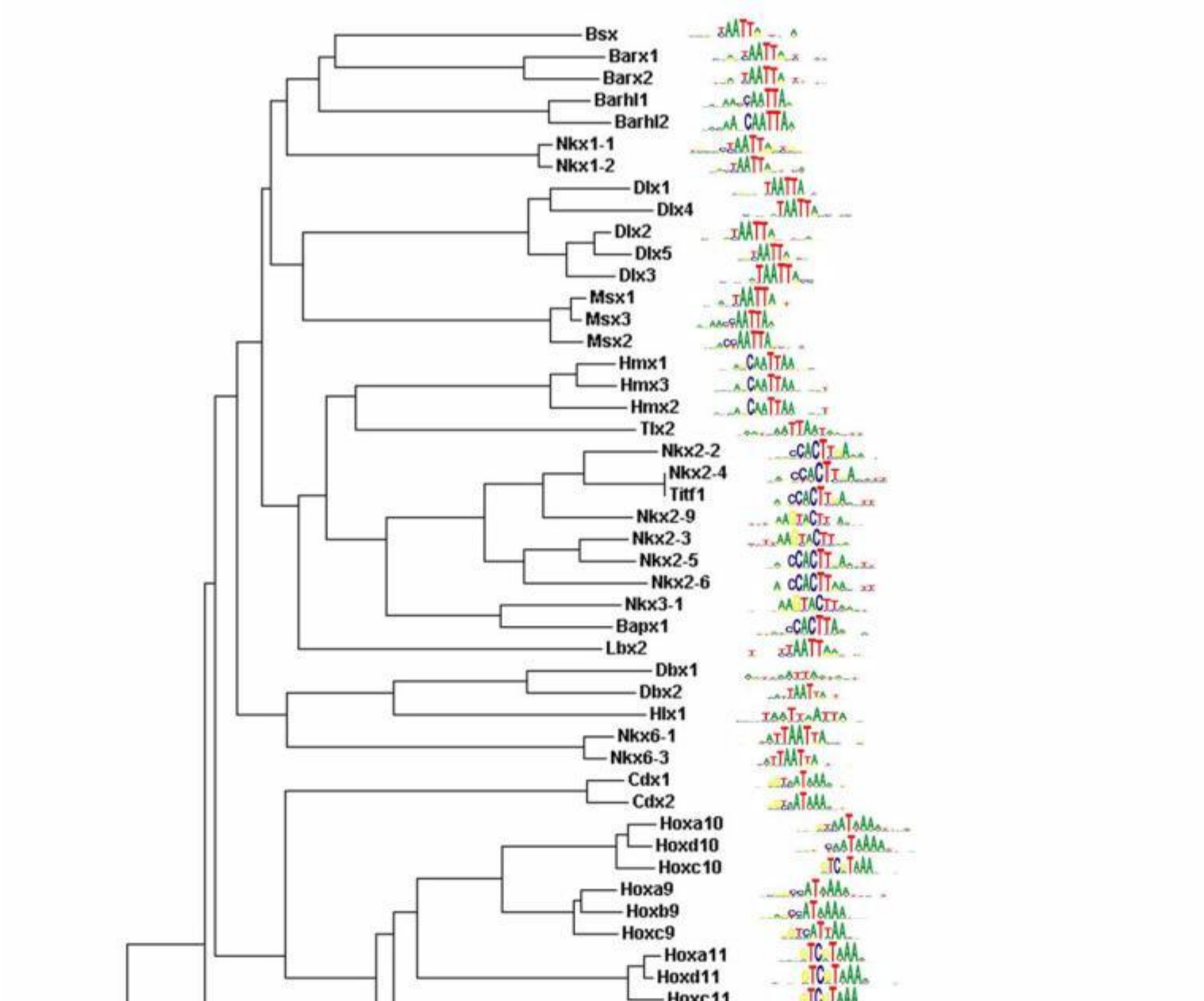


Similar DBDs recognize similar motifs



160 Mouse Homeodomains (Berger *et al.* 2008, Cell)

Similar DBDs recognize similar motifs



The “65% rule”

- 65% or greater AA DBD sequence identity -> nearly identical motif
  - Independent of evolutionary distance
  - Can use a mouse TF to predict yeast homolog’s binding preference
- Simple rule performs comparably to advanced methods
- Works for RNA binding proteins too!
  - Ray, Kazan, Cook, Weirauch, *et al.* , Nature 2013

Berger *et al. Cell* 2008; Alleyne *et al. Bioinformatics* 2009

TF “picking” strategy

Choose ~2000 TFs for characterization, in order to:

TF “picking” strategy

Choose ~2000 TFs for characterization, in order to:

1. Determine motif inference thresholds for the >80 TF families

TF “picking” strategy

Choose ~2000 TFs for characterization, in order to:

1. Determine motif inference thresholds for the >80 TF families

2. Maximize the total number of motif inferences across all organisms

TF “picking” strategy

Choose ~2000 TFs for characterization, in order to:

1. Determine motif inference thresholds for the >80 TF families
2. Maximize the total number of motif inferences across all organisms
3. Ensure all major clades are sampled

TF “picking” strategy

Choose ~2000 TFs for characterization, in order to:

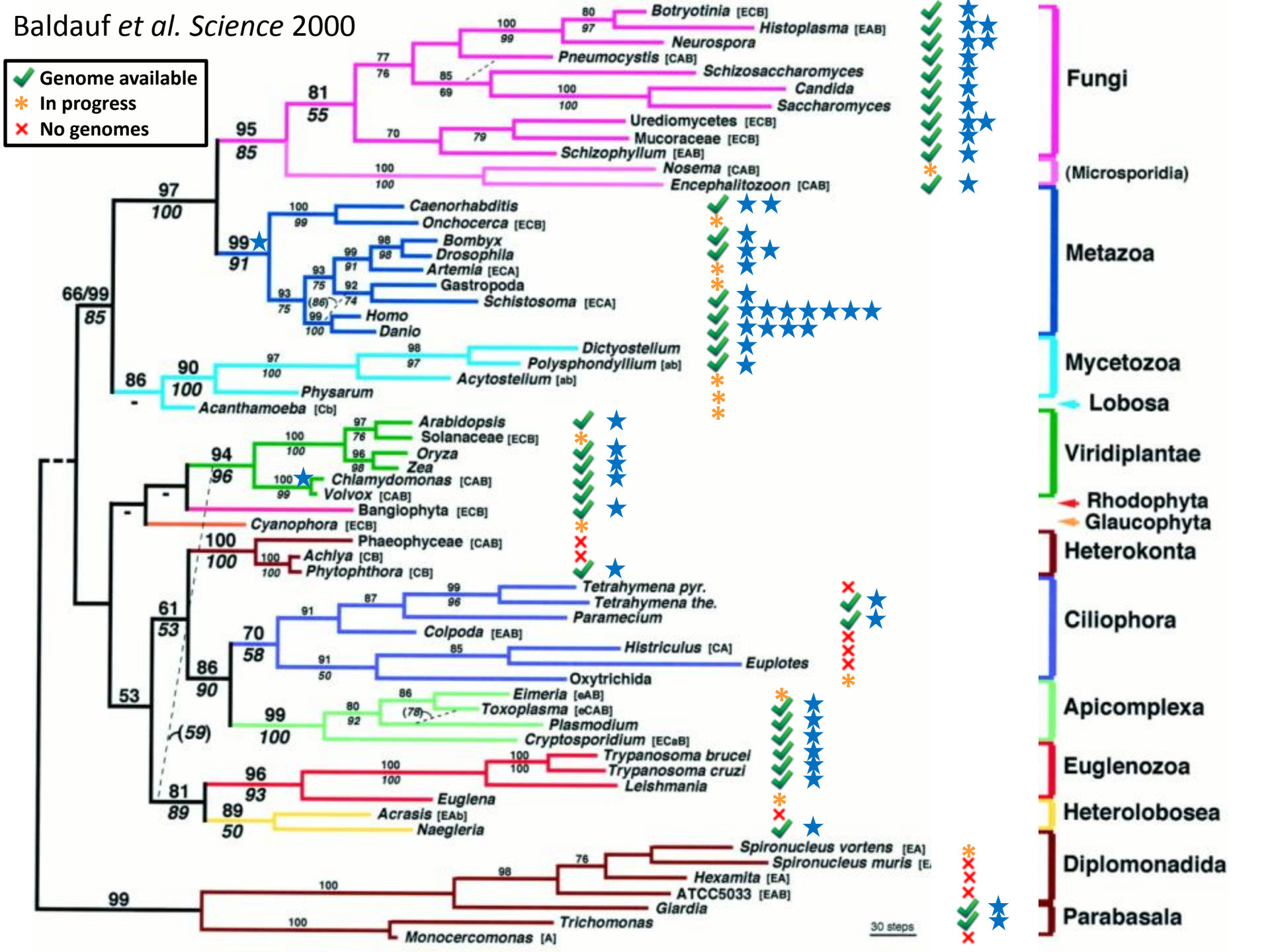
1. Determine motif inference thresholds for the >80 TF families

2. Maximize the total number of motif inferences across all organisms

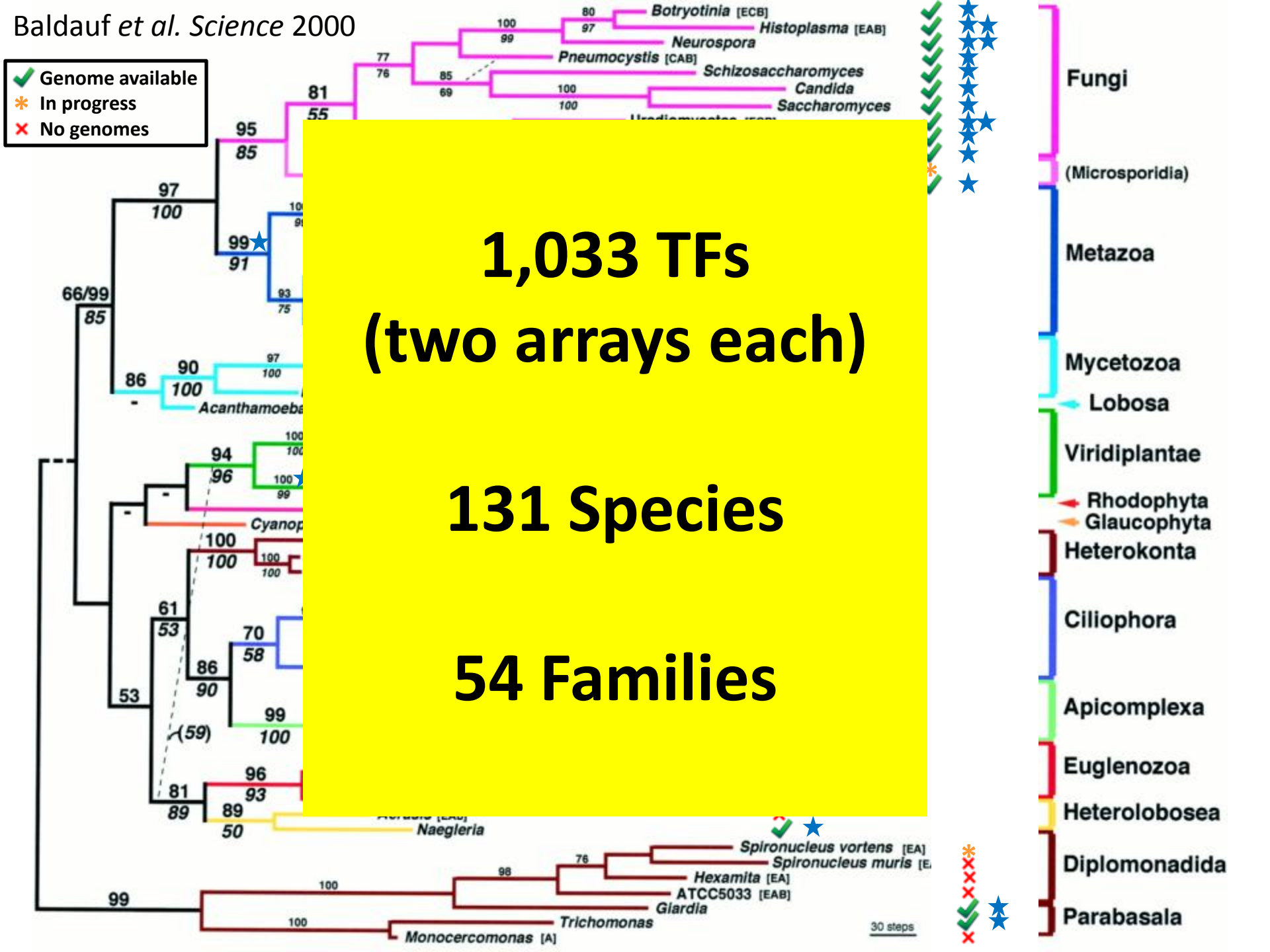
3. Ensure all major clades are sampled

4. Characterize motifs for select, diverse model organisms
   - *Arabidopsis thaliana* (plant)
   - *Dictyostelium discoideum* (slime mold)
   - *Ostreococcus tauri* (algae)
   - *Neurospora crassa* (fungus)
   - *Mus musculus* (animal)

Baldauf *et al. Science* 2000



Baldauf *et al. Science* 2000



1,033 TFs
(two arrays each)

131 Species

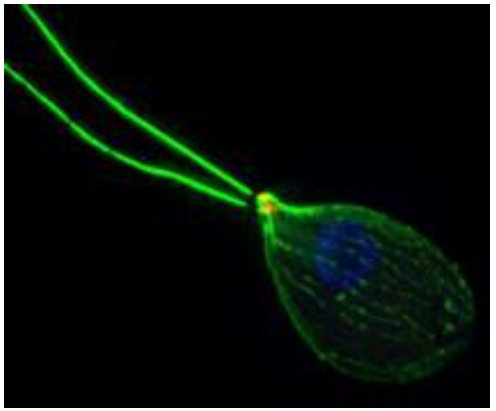
54 Families

T. nigroviridis

T. pseudonana

N. crassa

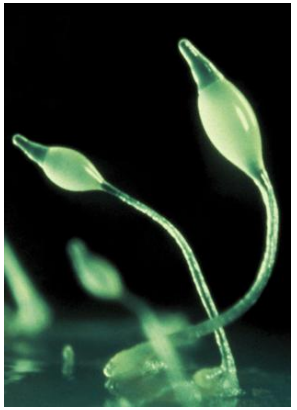
S. commune



N. gruberi

G. lamblia

B. mori



D. discoideum

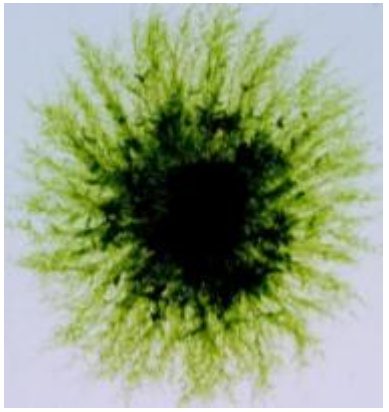
M. gallopavo

P. tetraurelia

O. sativa

N. vectensis

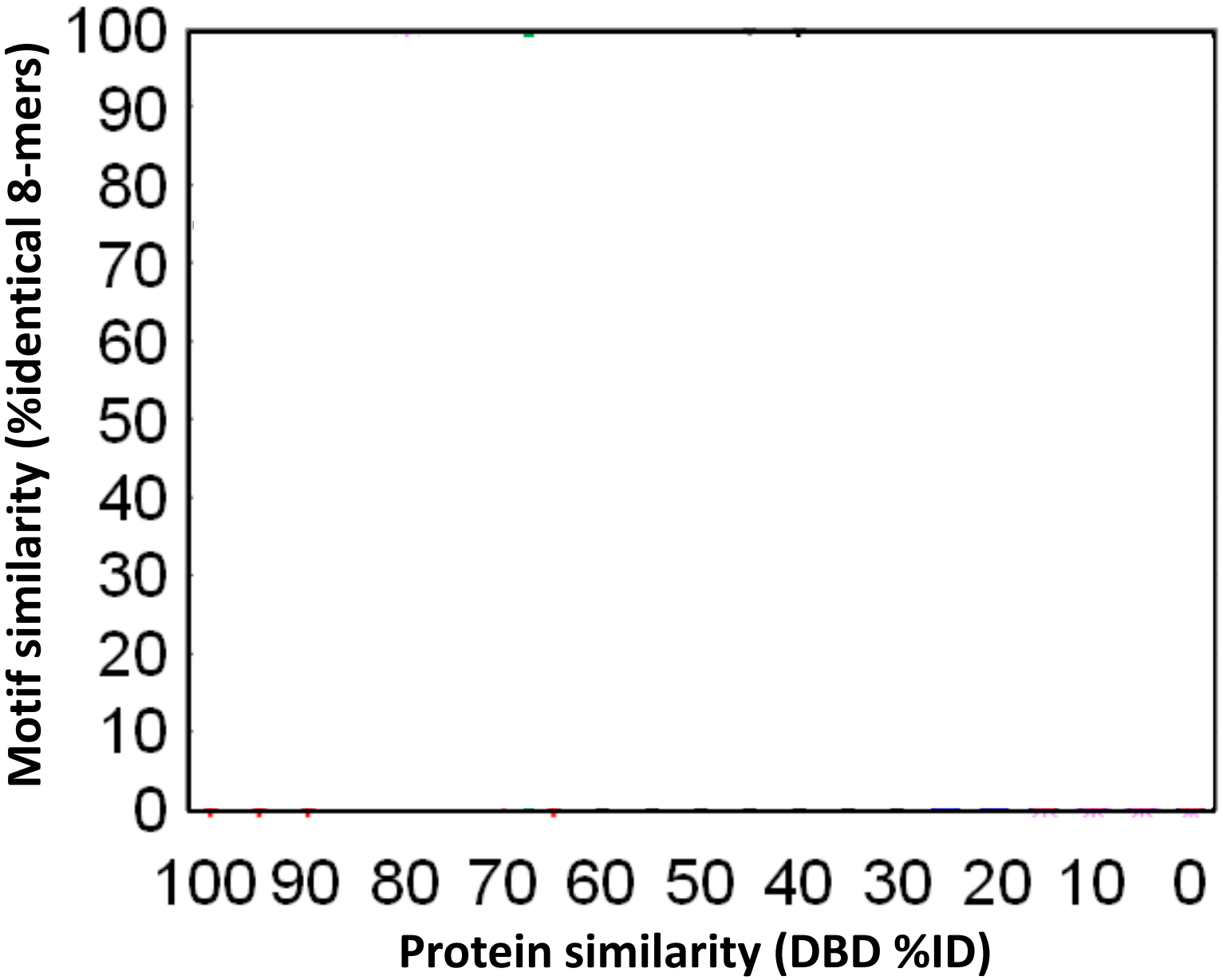
C. intestinalis



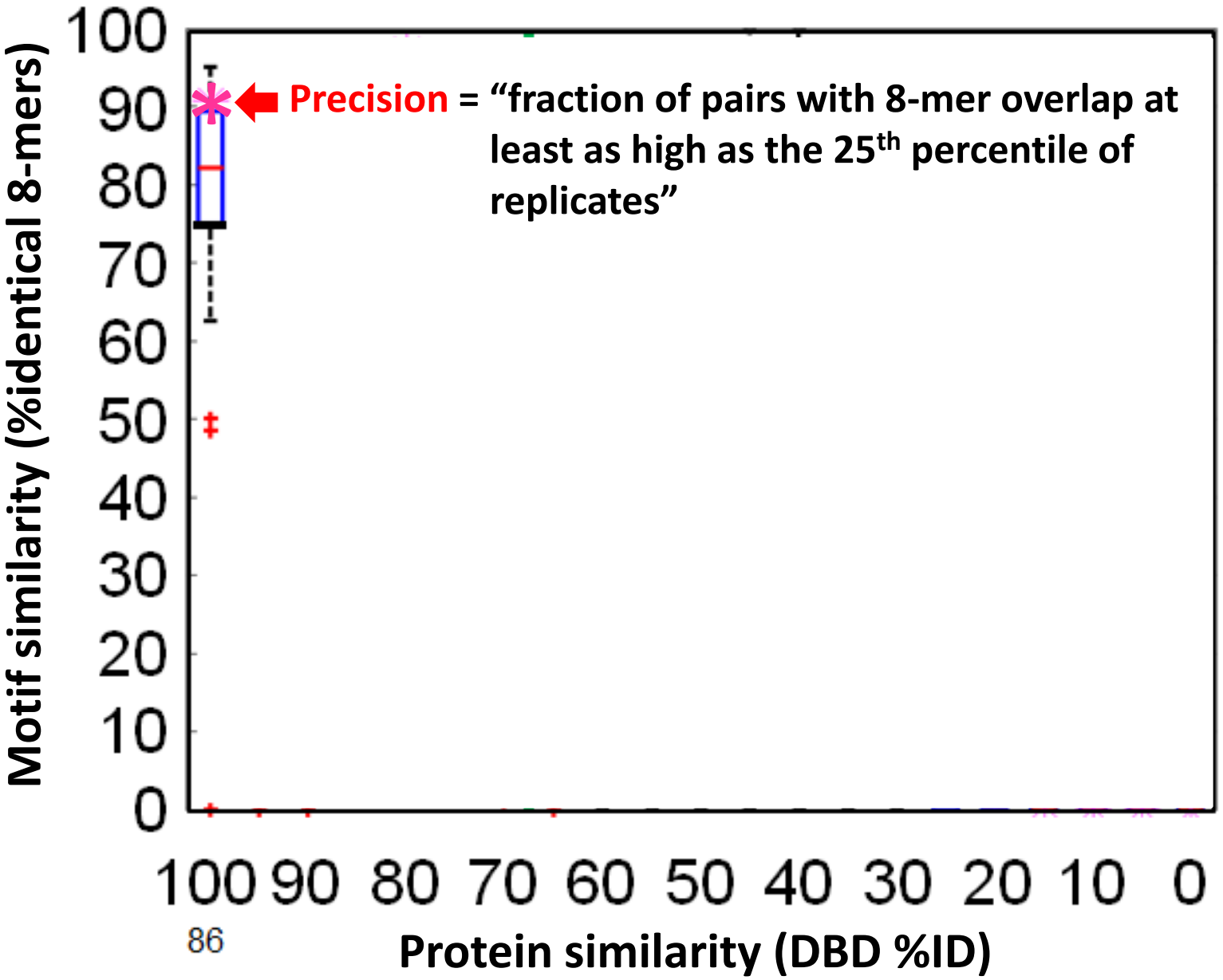
P. patens

A. carolinensis

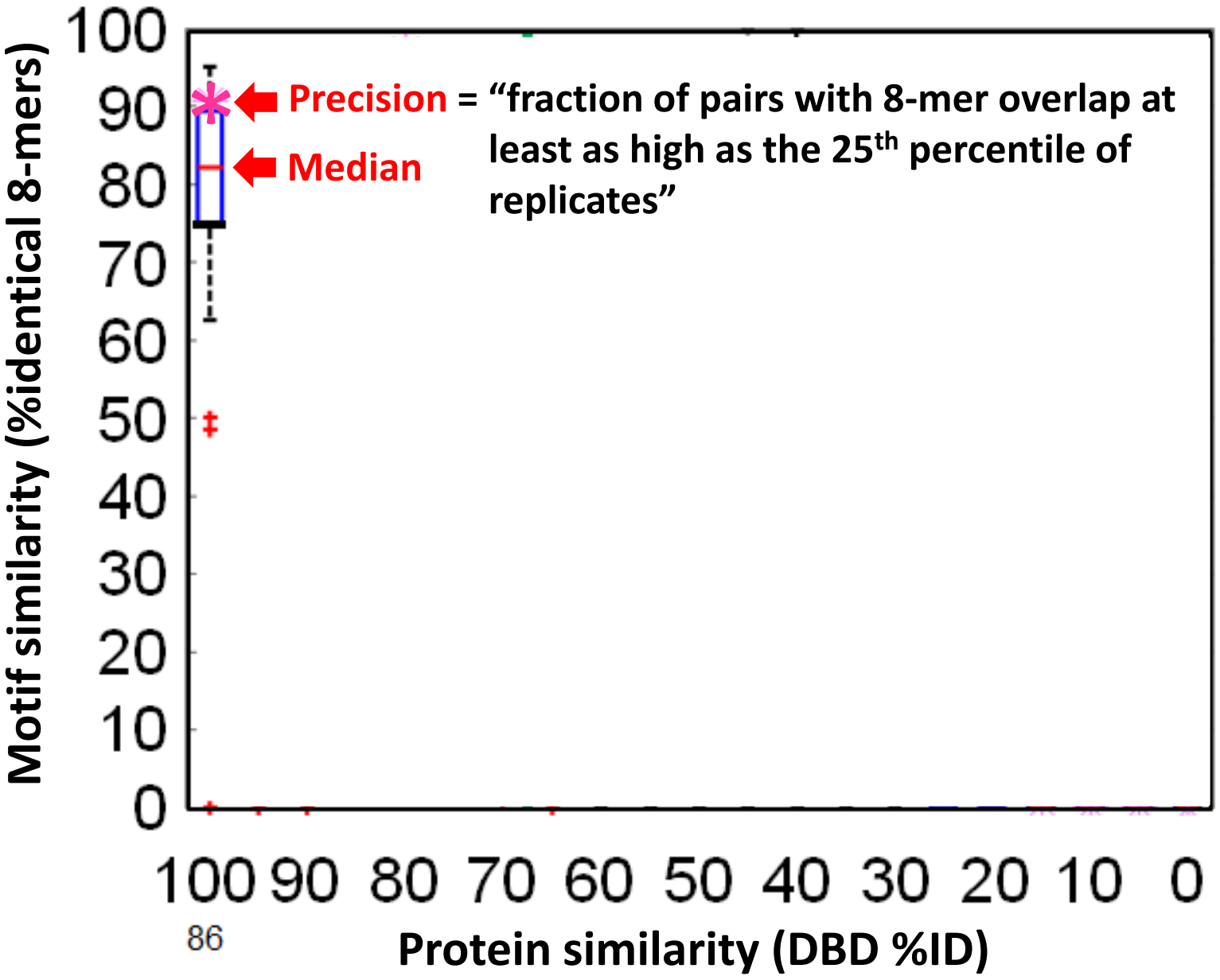
65% rule, revisited



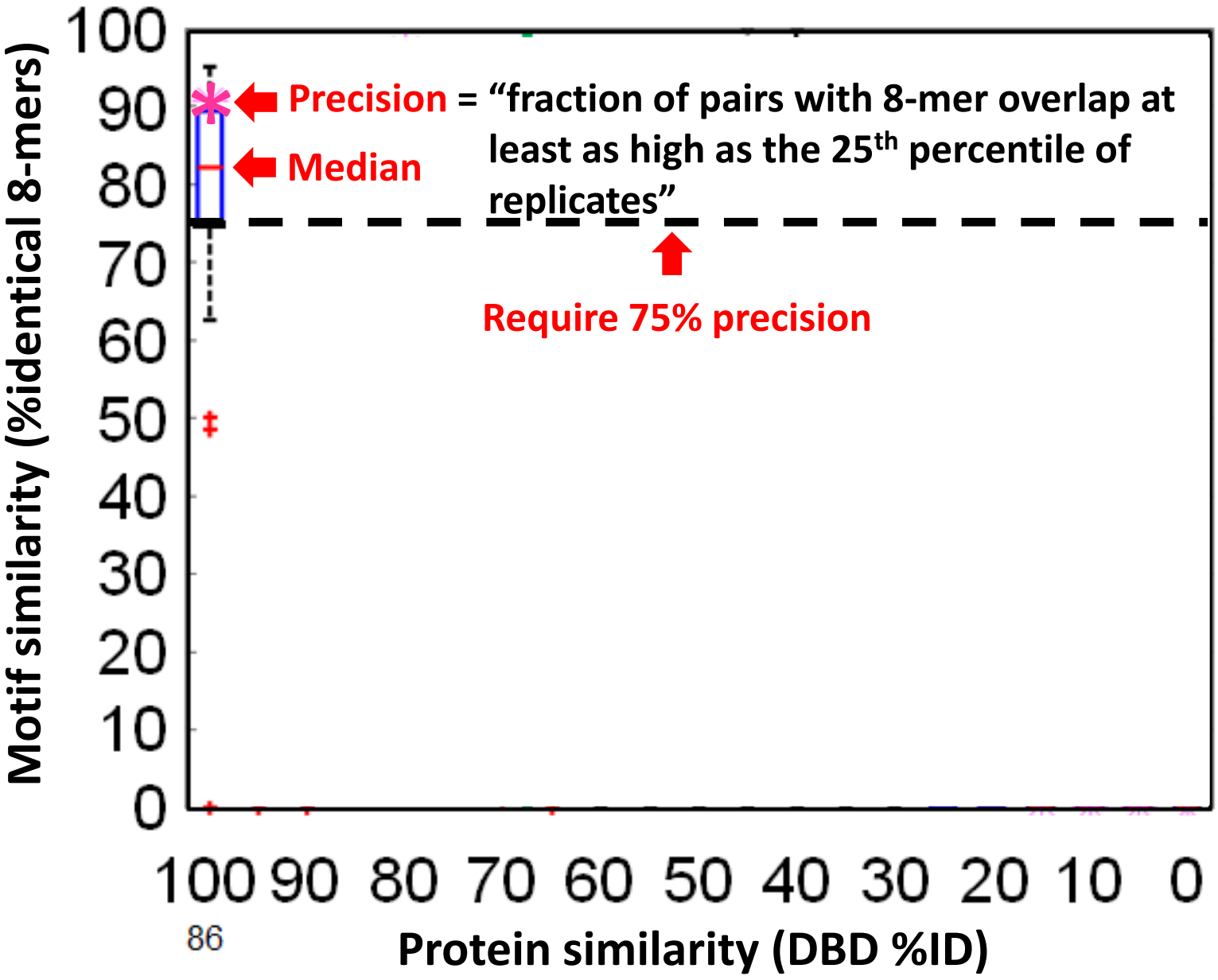
65% rule, revisited



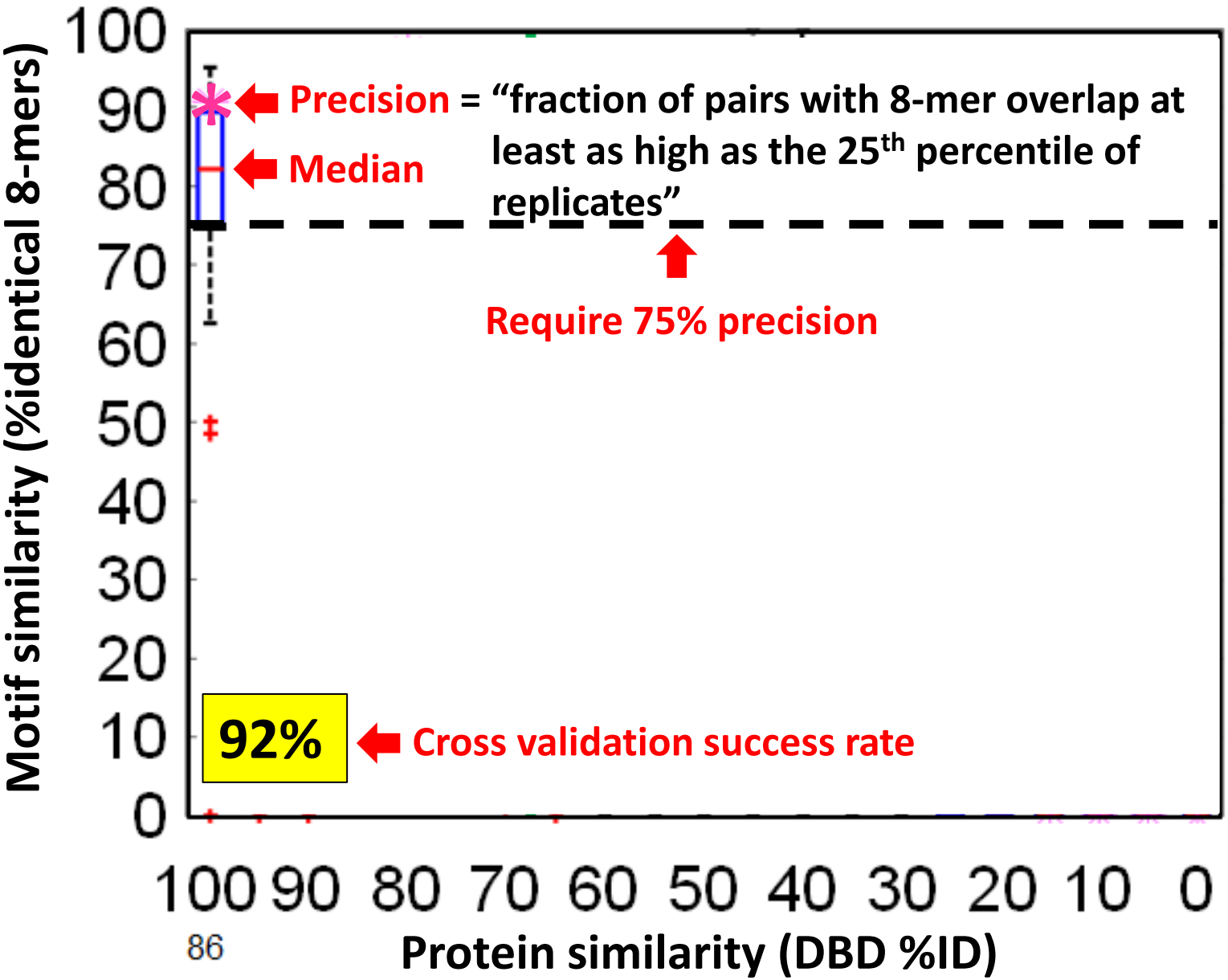
65% rule, revisited



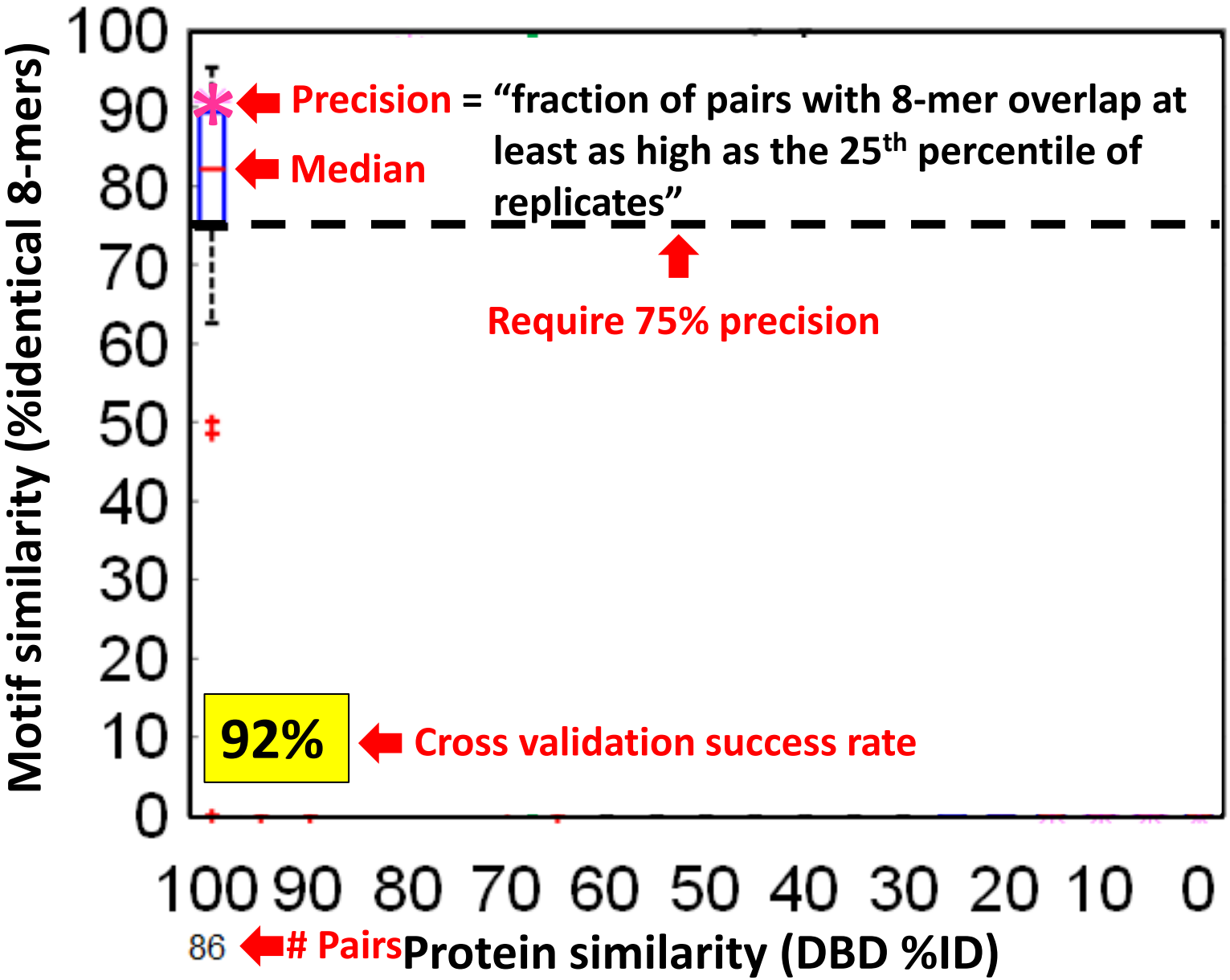
65% rule, revisited



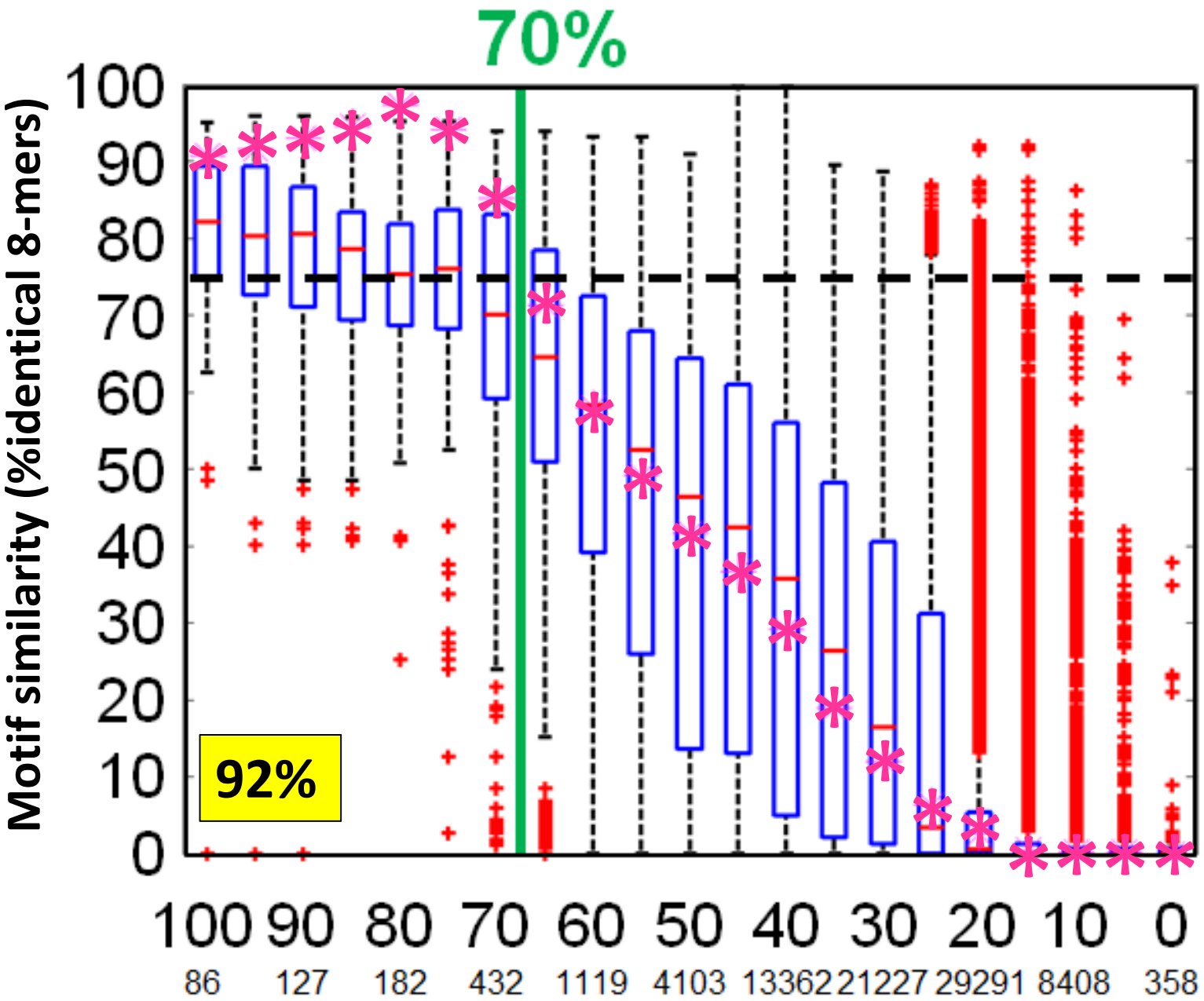
65% rule, revisited



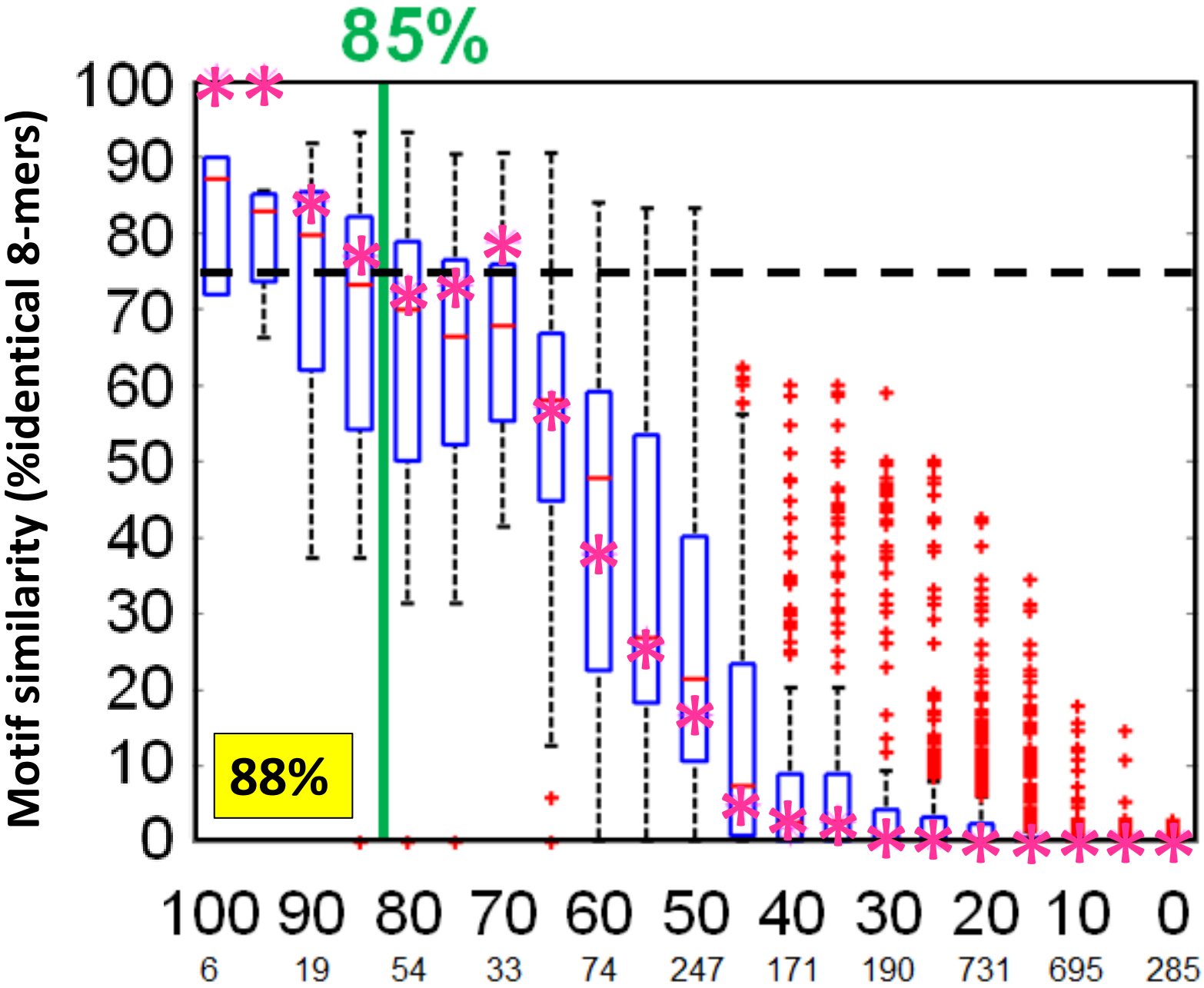
65% rule, revisited



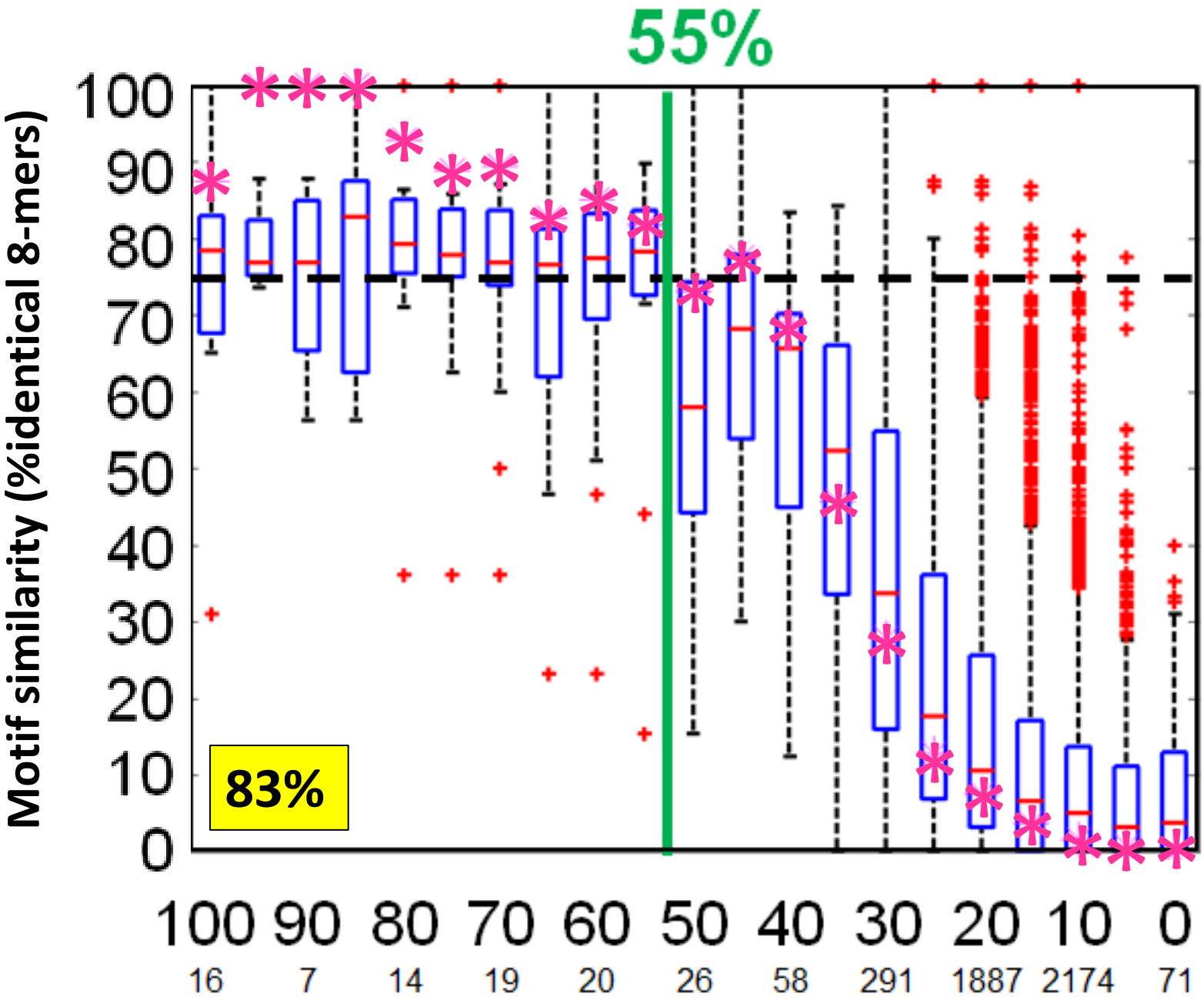
Homeodomains



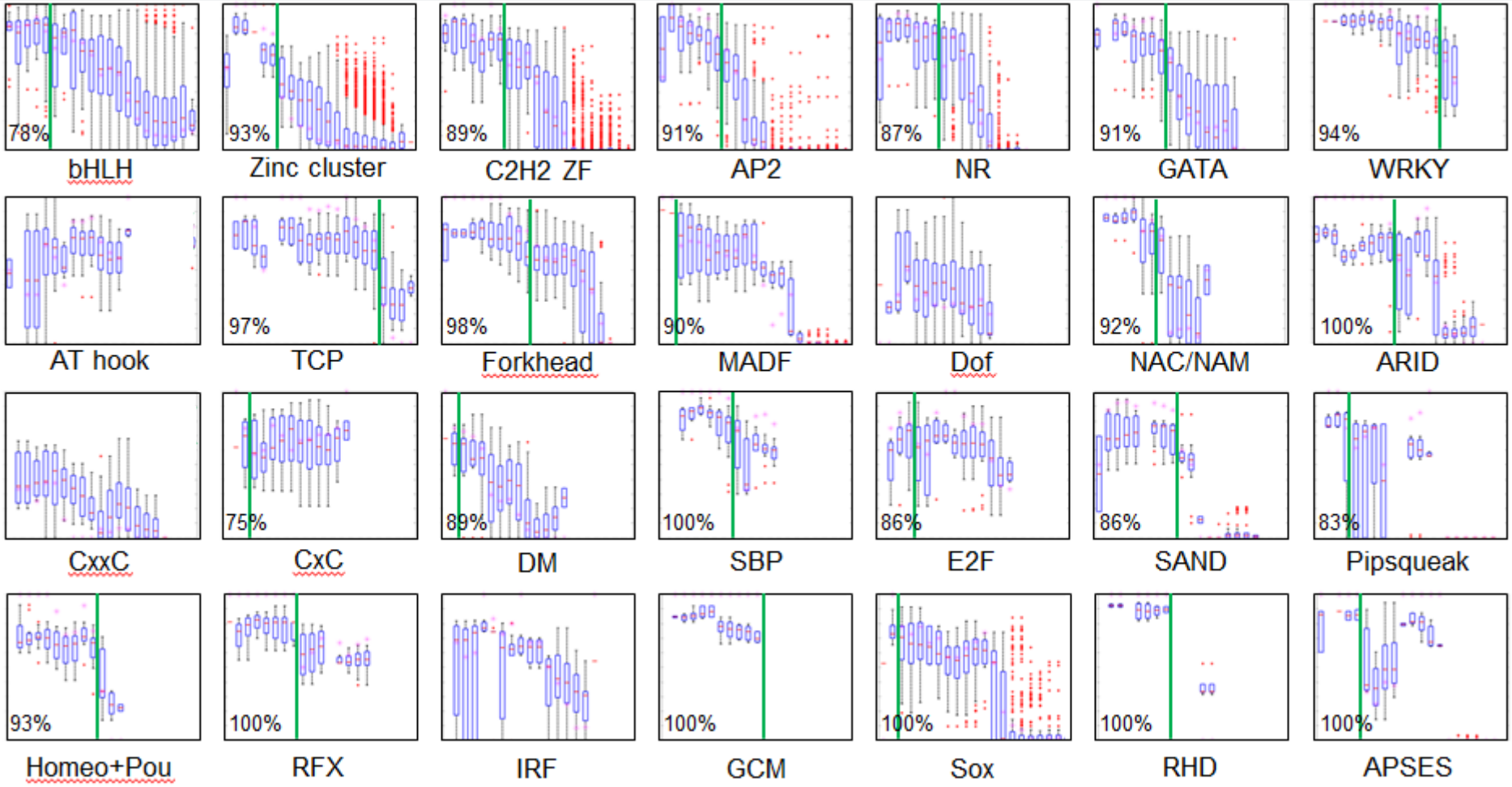
Myb/SANT


85%
88%
Motif similarity (%identical 8-mers)
100
90
80
70
60
50
40
30
20
10
0
100 90 80 70 60 50 40 30 20 10 0
6 19 54 33 74 247 171 190 731 695 285

bZIPs

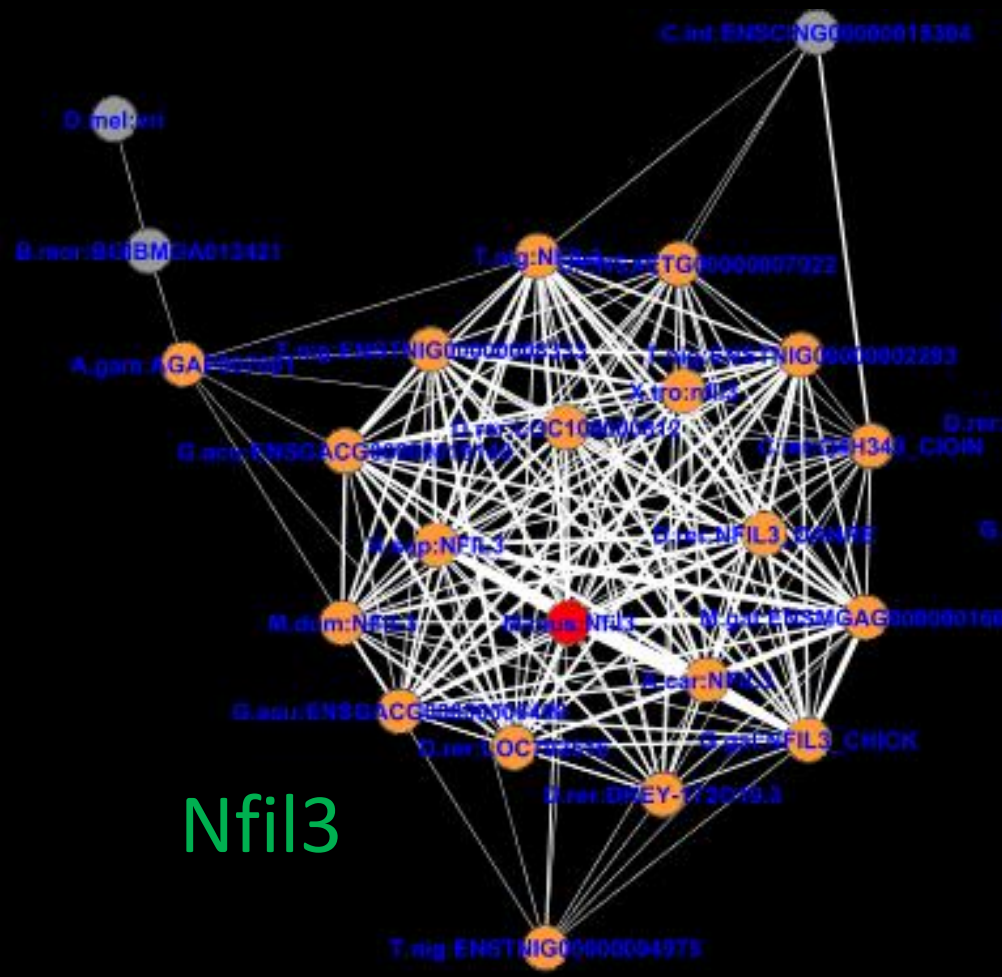

55%
Motif similarity (%identical 8-mers)
100
90
80
70
60
50
40
30
20
10
0
83%
100 90 80 70 60 50 40 30 20 10 0
16 7 14 19 20 26 58 291 1887 2174 71

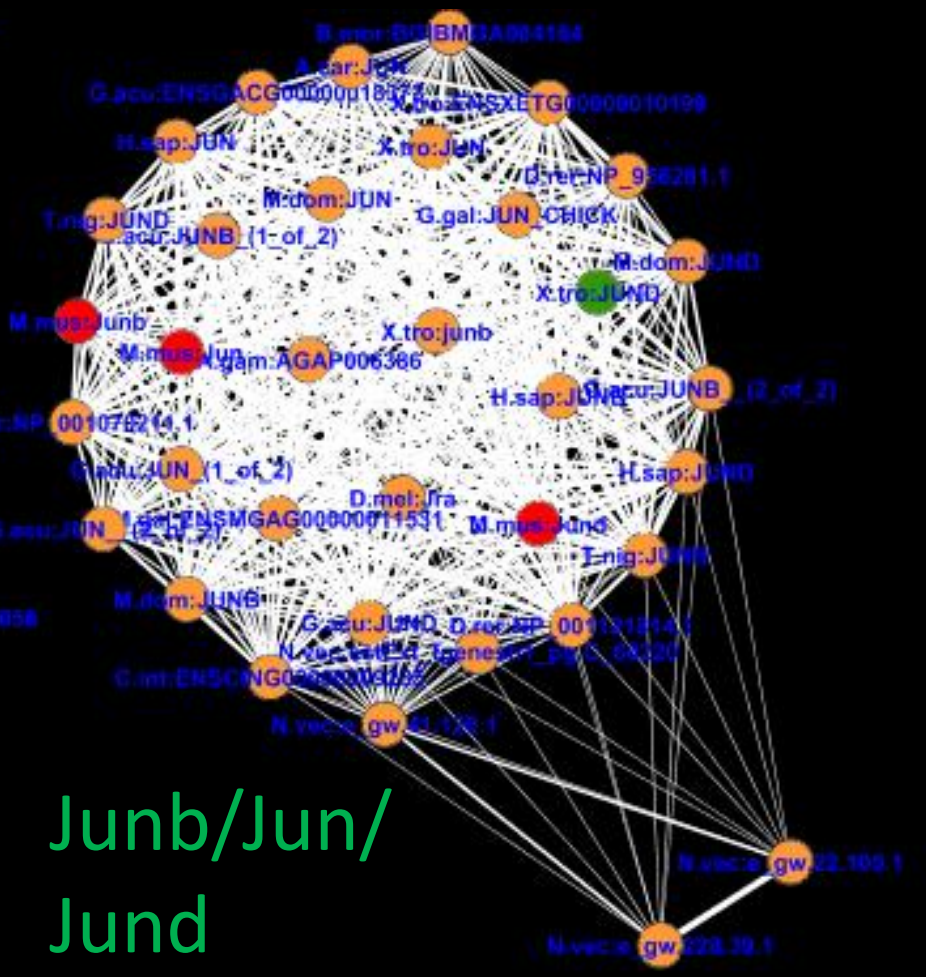
Thresholds can be drawn for most TF families

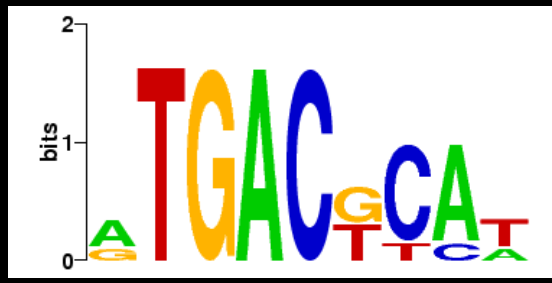


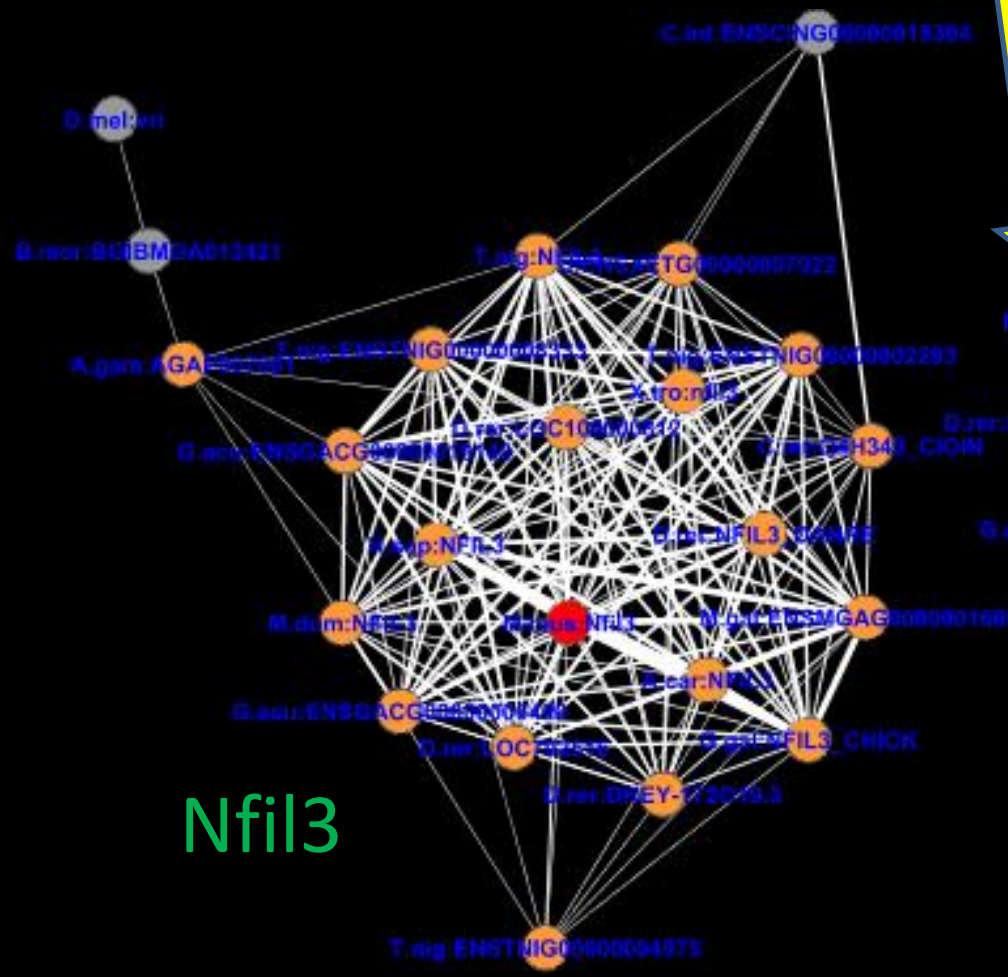
Overall Cross Validation success rate: 89%

bZIP subfamilies



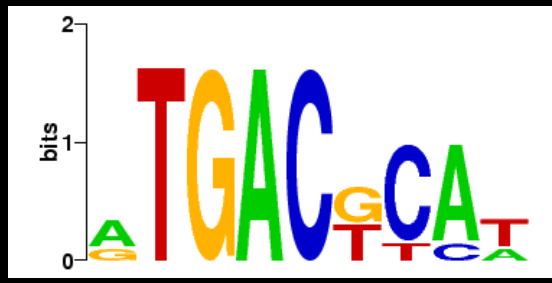




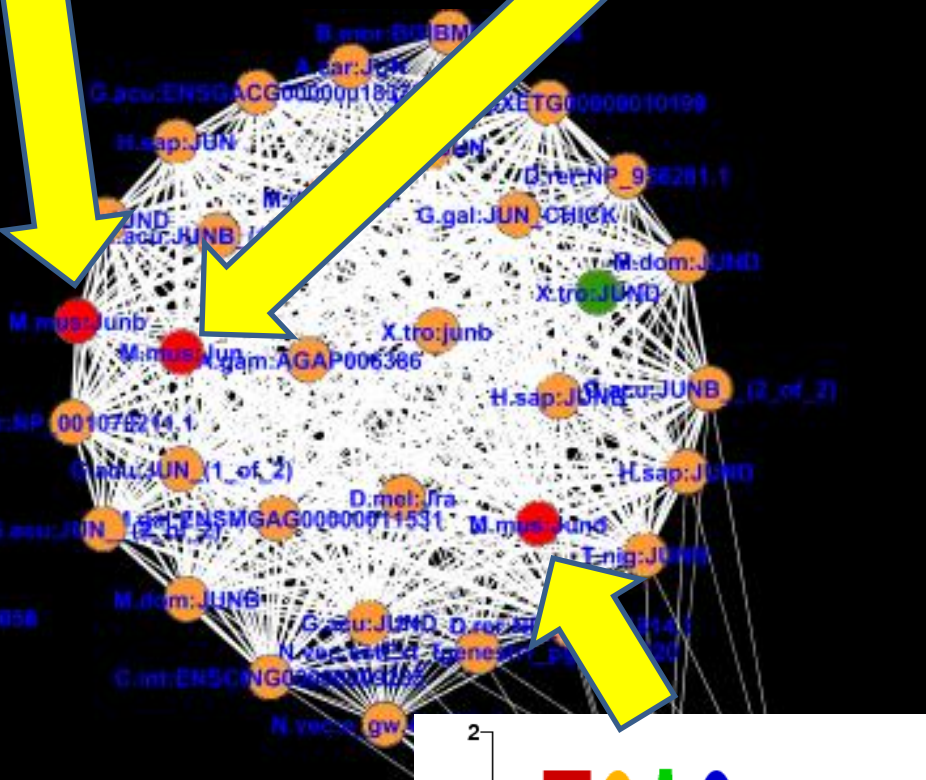


Nfil3

Junb/Jun/
Jund

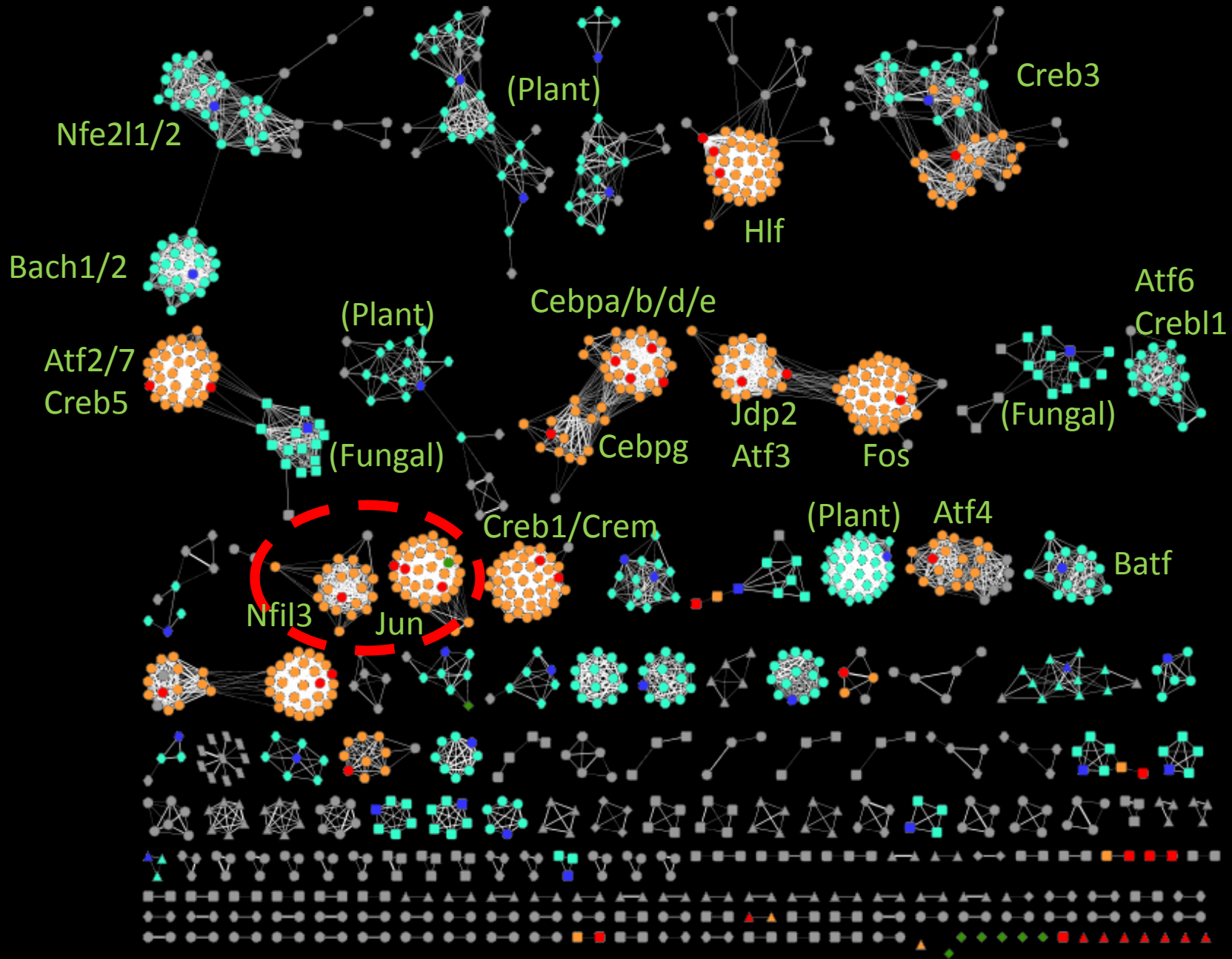


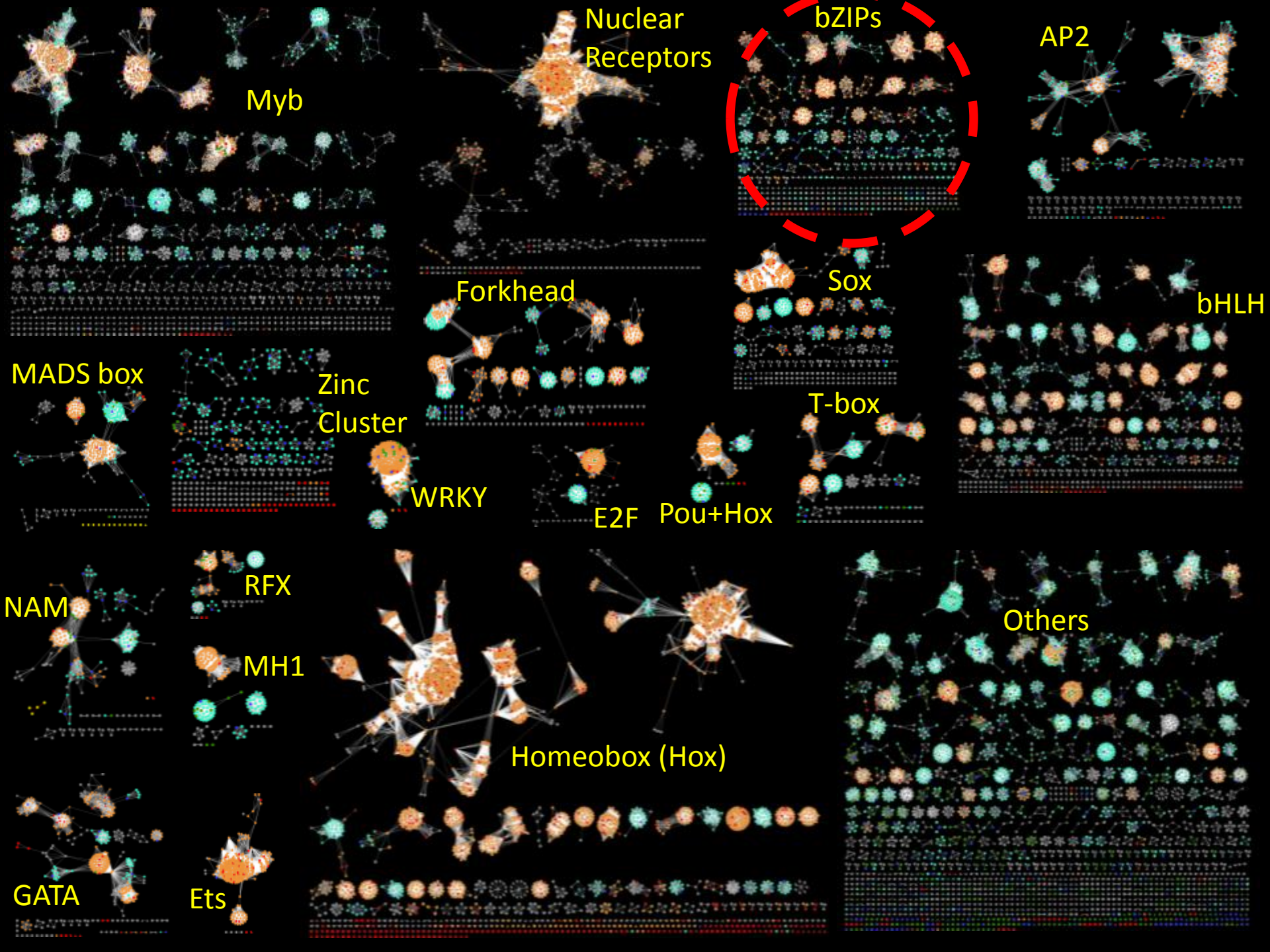
Nfil3



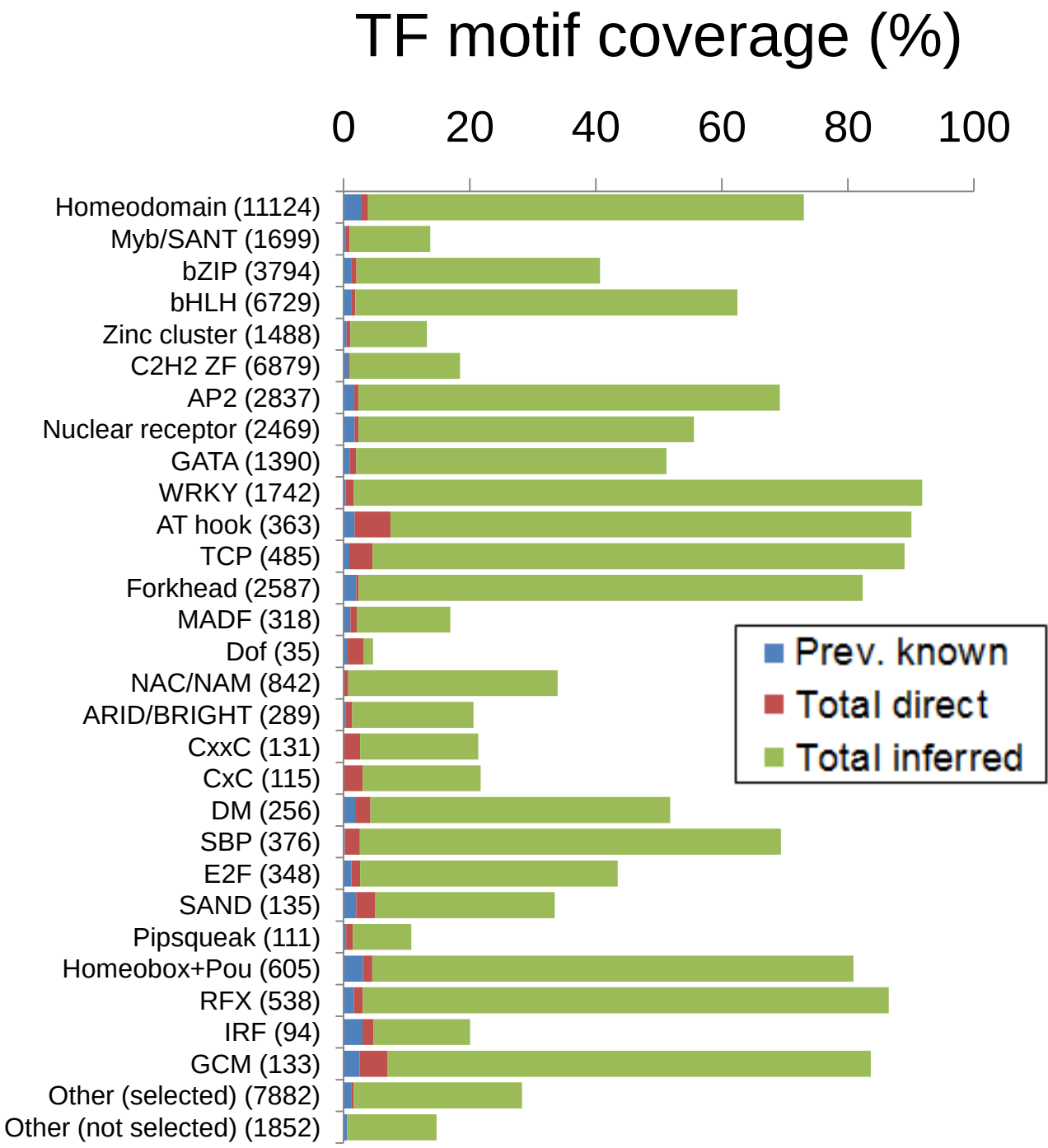
Junb/Jun/
Jund

Basic Leucine Zippers (bZIPs)

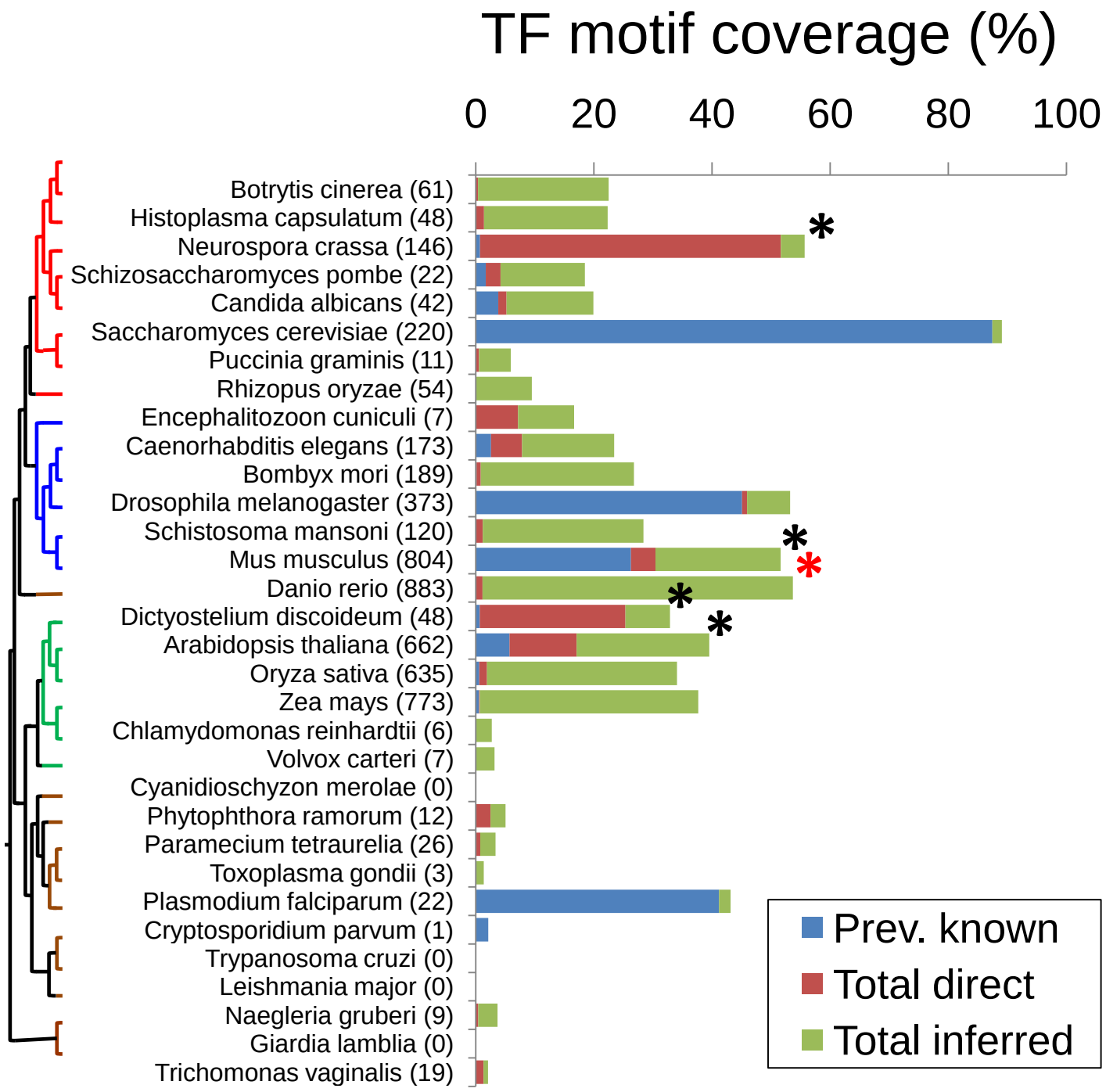



Myb
Nuclear Receptors
bZIPs
AP2
Forkhead
Sox
bHLH
MADS box
Zinc Cluster
T-box
WRKY
E2F
Pou+Hox
NAM
RFX
MH1
Homeobox (Hox)
Others
GATA
Ets

Coverage by Family



Coverage by Species



Coverage summary

~162,000	**Total estimated number of sequenced eukaryotic TFs**

Coverage summary

~162,000	Total estimated number of sequenced eukaryotic TFs
~2,000 (<2%)	Number of TF motifs from other studies

Coverage summary

~162,000	Total estimated number of sequenced eukaryotic TFs
~2,000 (<2%)	Number of TF motifs from other studies
~63,000 (39%)	Number of TF motifs we currently have determined or can infer

Coverage summary

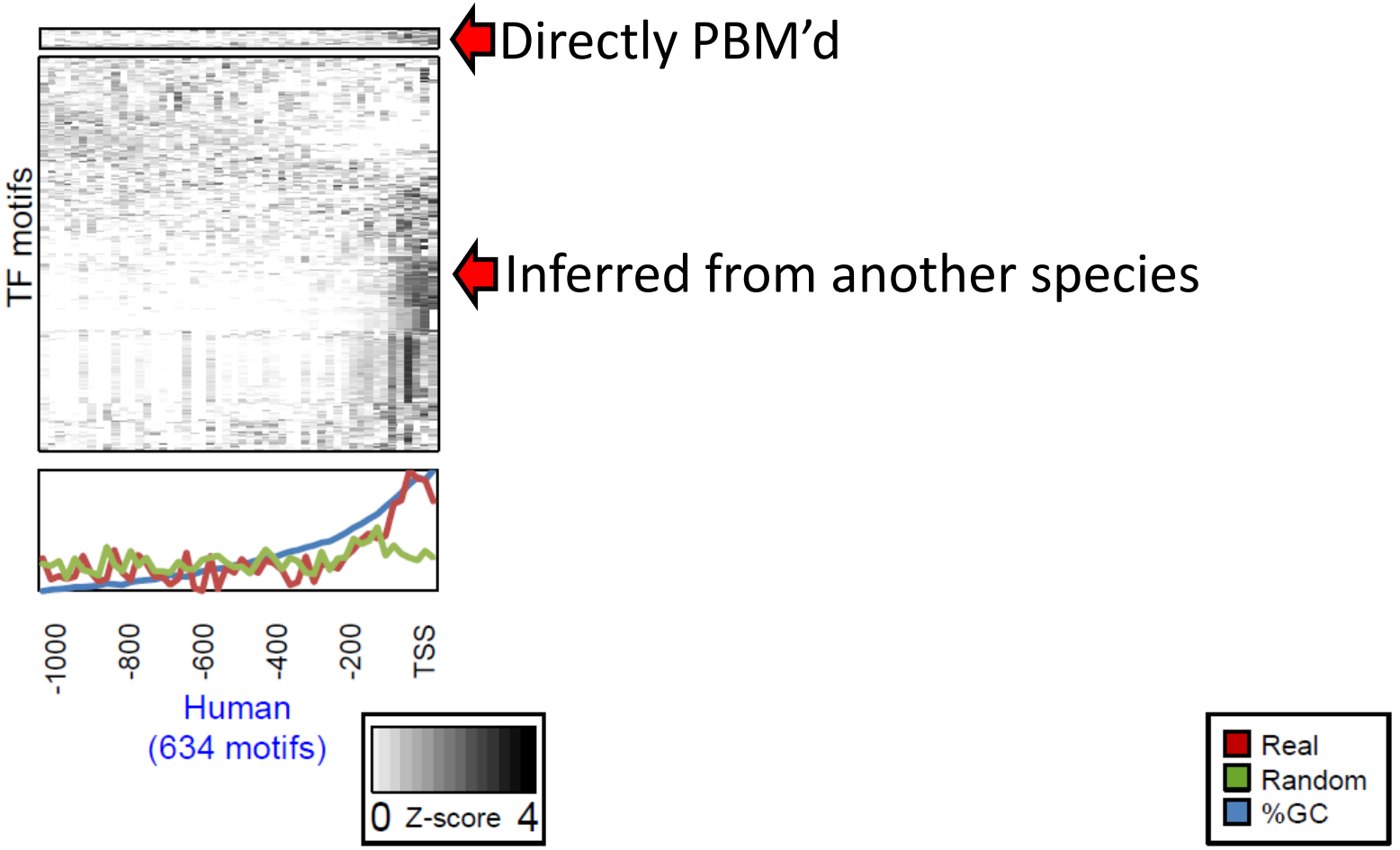
~162,000	Total estimated number of sequenced eukaryotic TFs
~2,000 (<2%)	Number of TF motifs from other studies
~63,000 (39%)	Number of TF motifs we currently have determined or can infer
~80,000 (49%)	Number of TF motifs (lenient, 50% precision threshold)

Coverage summary

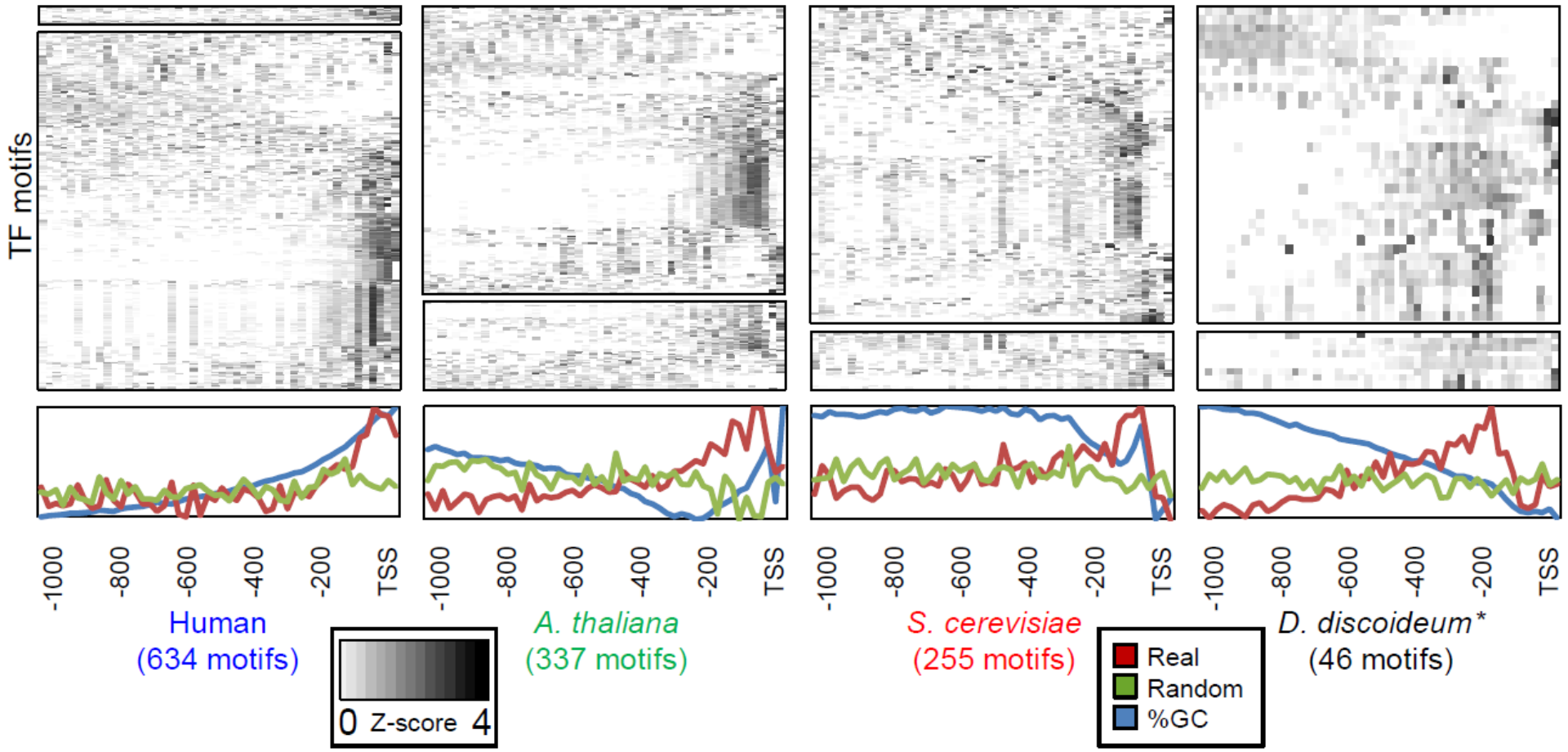
~162,000	Total estimated number of sequenced eukaryotic TFs
~2,000 (<2%)	Number of TF motifs from other studies
~63,000 (39%)	Number of TF motifs we currently have determined or can infer
~80,000 (49%)	Number of TF motifs (lenient, 50% precision threshold)

We recently ordered primers
for ~3000 more proteins

Positional bias of motifs in eukaryotic promoters

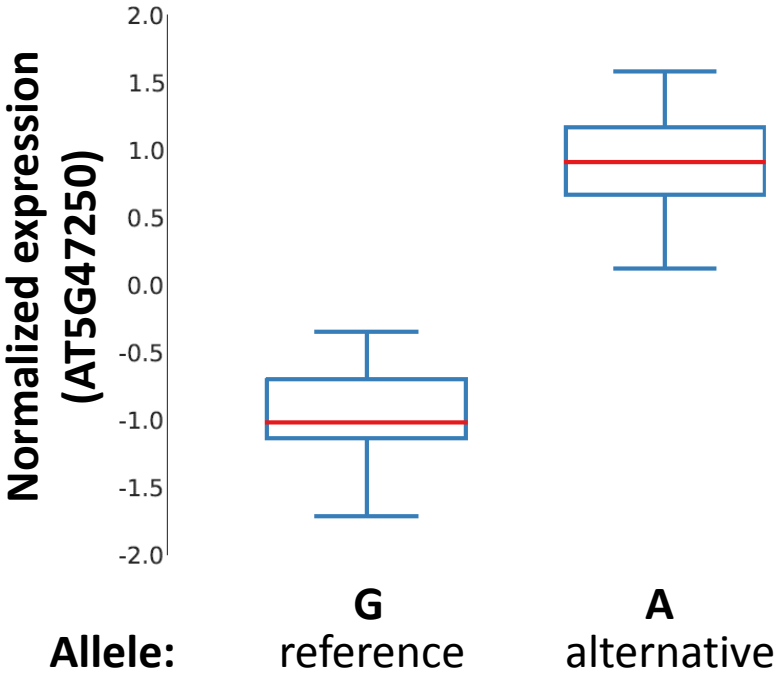


Positional bias of motifs in eukaryotic promoters

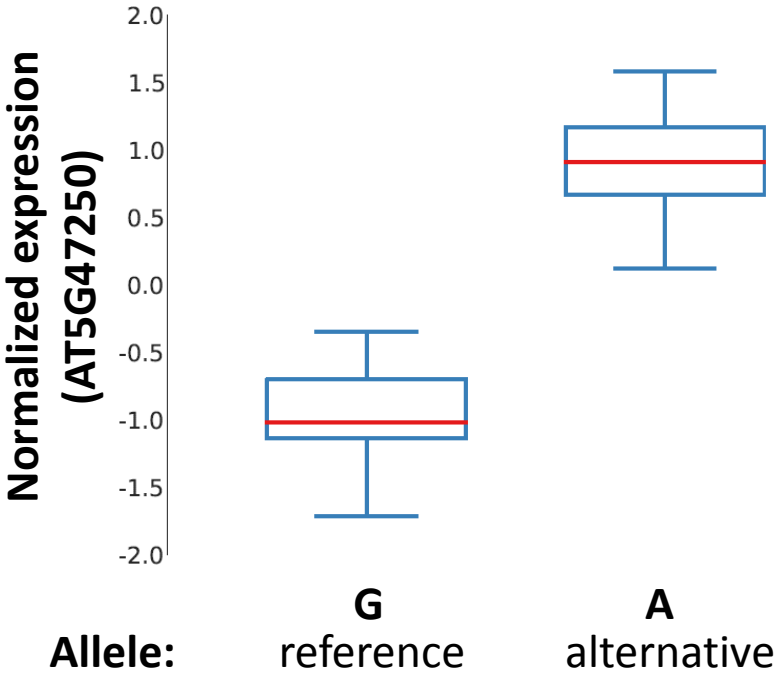


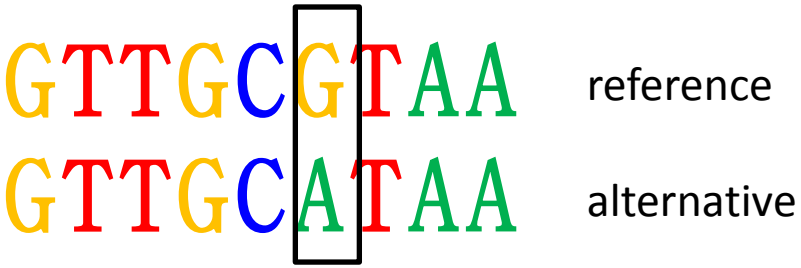
Both direct + inferred motifs “peak” near TSS

Motifs tie genotype to expression

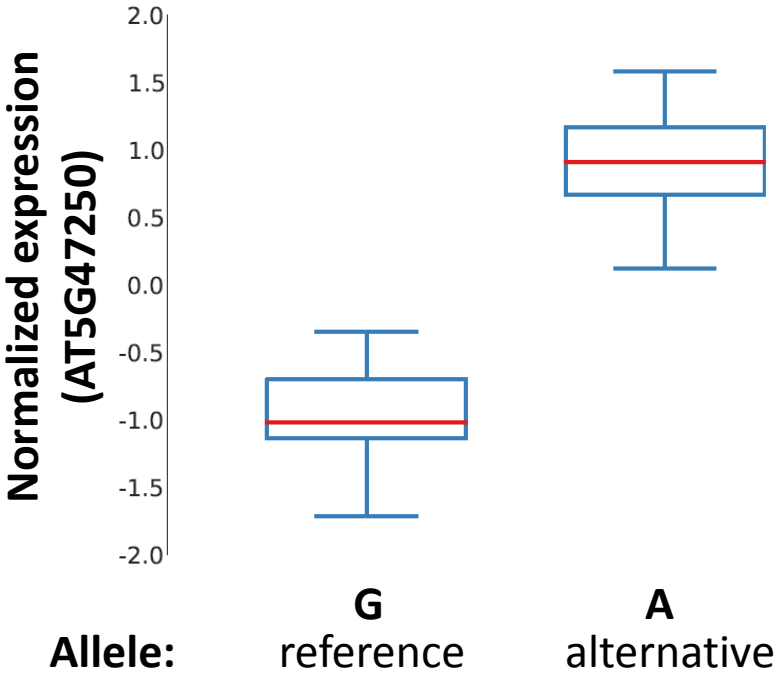


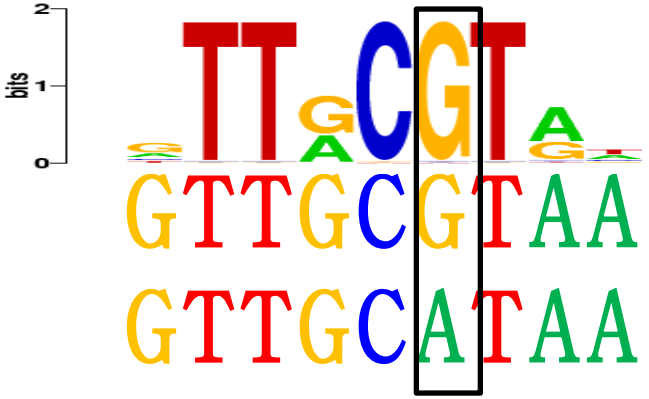
Motifs tie genotype to expression



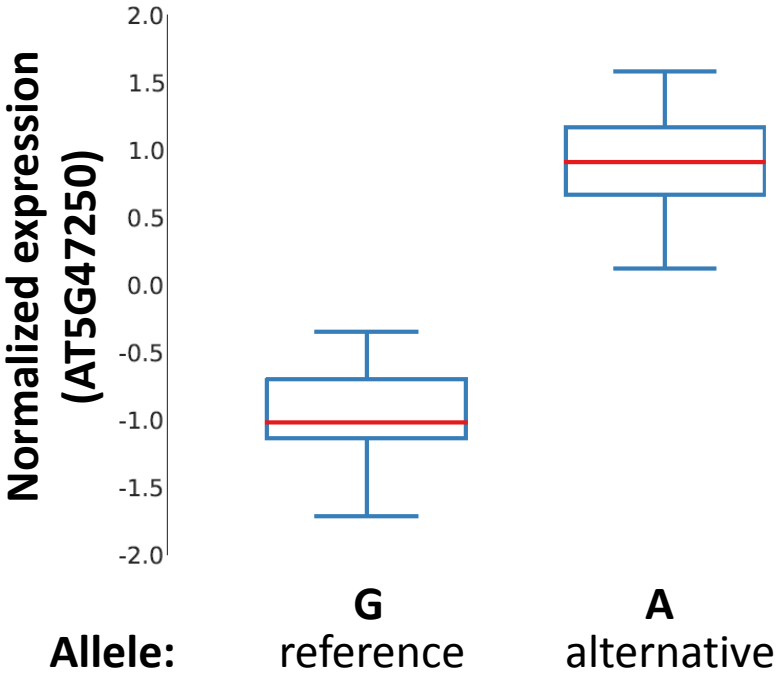


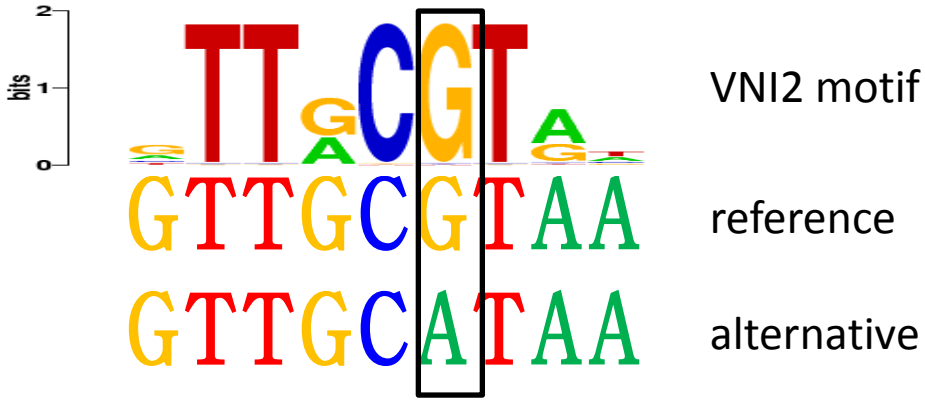
Motifs tie genotype to expression





Motifs tie genotype to expression





- VNI2 is a repressor (Yamaguchi *et al.* Plant Cell, 2010).
  - The alternative (A) allele "breaks" a binding site for this repressor, causing the de-repression of AT5G47250

- Both genes involved in defense response against the same pathogen

Identifying TFs affected by human disease SNPs

- **Procedure**:
  - Use all PBM data to “score” risk and non-risk alleles of known disease-associated SNPs
  - Rank all TFs based on the likelihood that their binding would be disrupted by the SNP

Identifying TFs affected by human disease SNPs

- **Procedure**:
  - Use all PBM data to “score” risk and non-risk alleles of known disease-associated SNPs
  - Rank all TFs based on the likelihood that their binding would be disrupted by the SNP

- **Results**:
  - Applied to set of 16 experimentally confirmed TF/disease SNP pairs

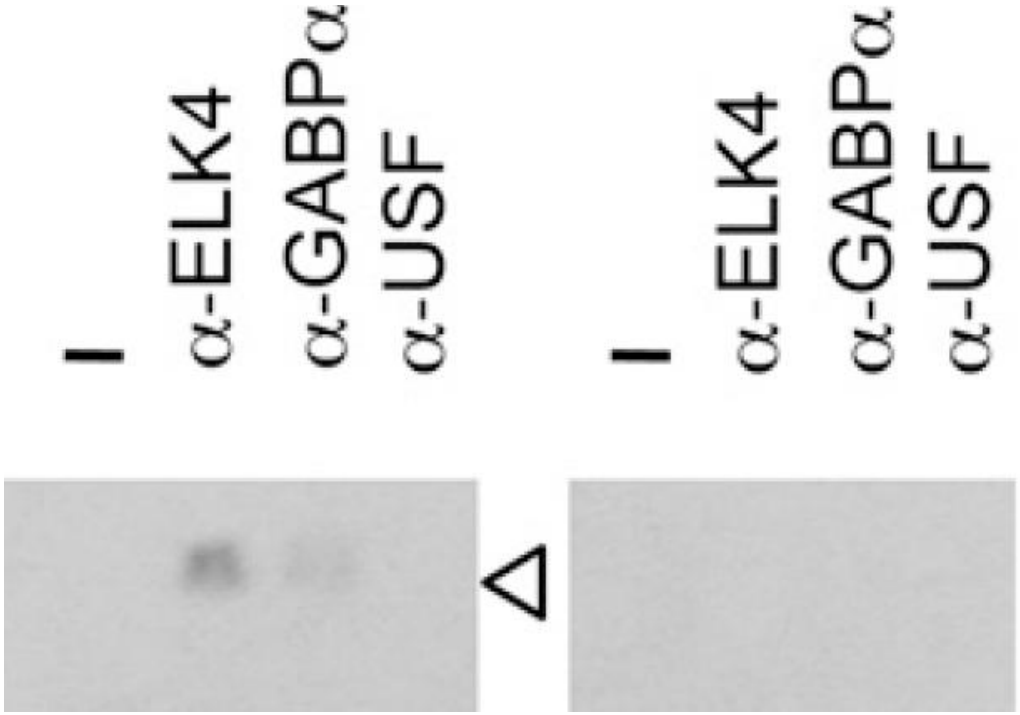
Identifying TFs affected by human disease SNPs

- **Procedure**:
  - Use all PBM data to "score" risk and non-risk alleles of known disease-associated SNPs
  - Rank all TFs based on the likelihood that their binding would be disrupted by the SNP

- **Results**:
  - Applied to set of 16 experimentally confirmed TF/disease SNP pairs
  - Could rank correct TF (or highly related) in the top five for 10 out of 16

Example: differential TF binding to rs554219 in breast cancer tumors

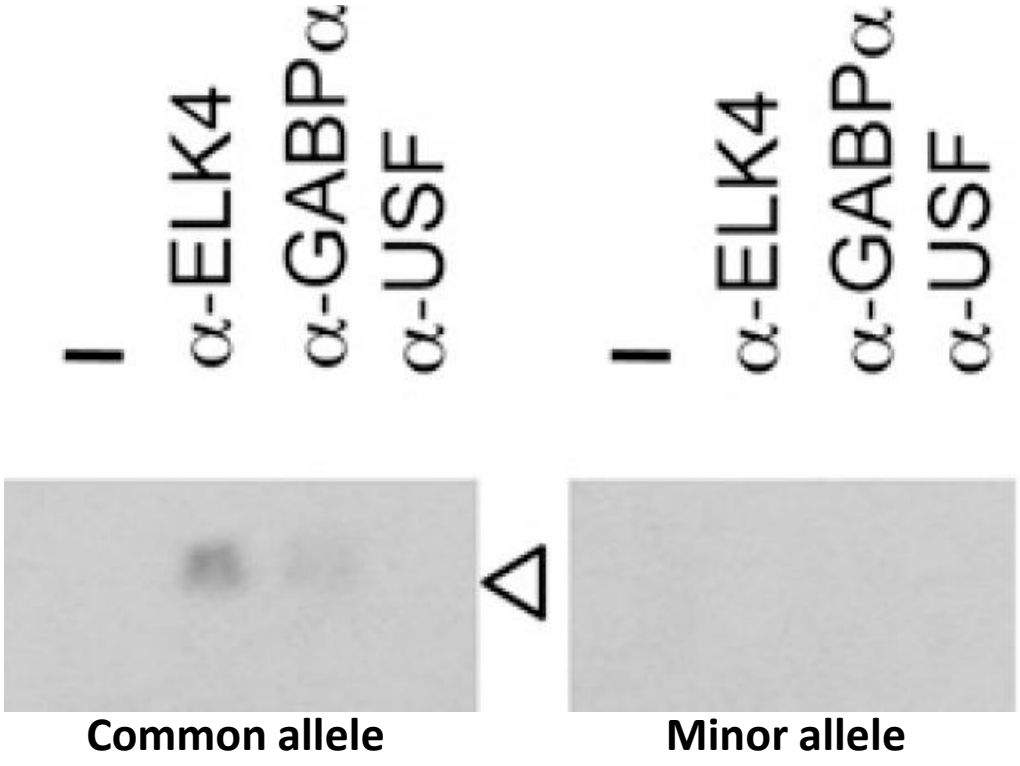
Supershift EMSA



French et al. *AJHG* 2013

Example: differential TF binding to rs554219 in breast cancer tumors

Supershift EMSA



Computational

TF	Common	Minor
1. ELK4	0.498	0.131
2. GABPα	0.496	0.329

(ranked out of >800 possible TFs)

French et al. *AJHG* 2013

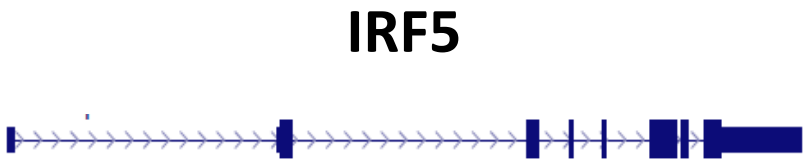
Application to lupus: IRF5 and rs4728142

- **IRF5 (Interferon Regulatory Factor 5)**
  - Controls important inflammatory pathways involved in lupus, autoimmunity, and infectious disease
  - Strongly associated with lupus
  - Over-expressed in lupus patients

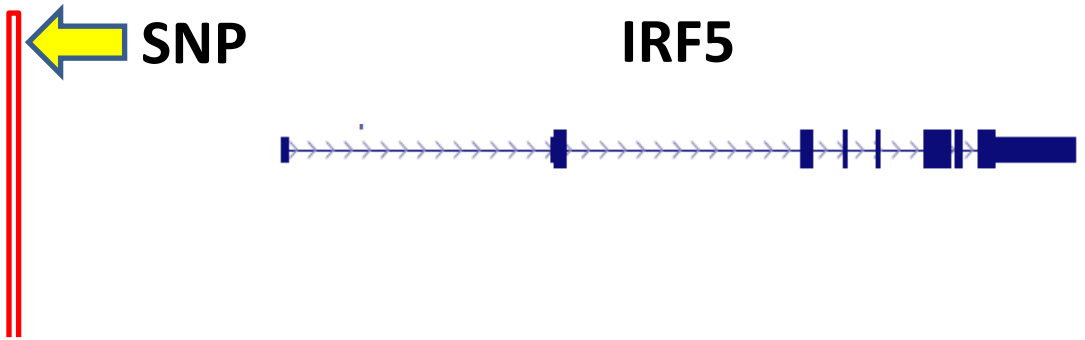
- **rs4728142**
  - Most likely of six possible causal variants for IRF5's association with lupus (Harley/Kottyan labs, targeted deep sequencing of 1,000s of cases and controls)
  - Risk allele results in increased binding of *some protein* in blood cell nuclear extracts (Kristjansdottir *et al.,* 2008), and in B cell lines (Harley lab, unpublished data)

Kottyan et al., Hum Mol Genet. 2014

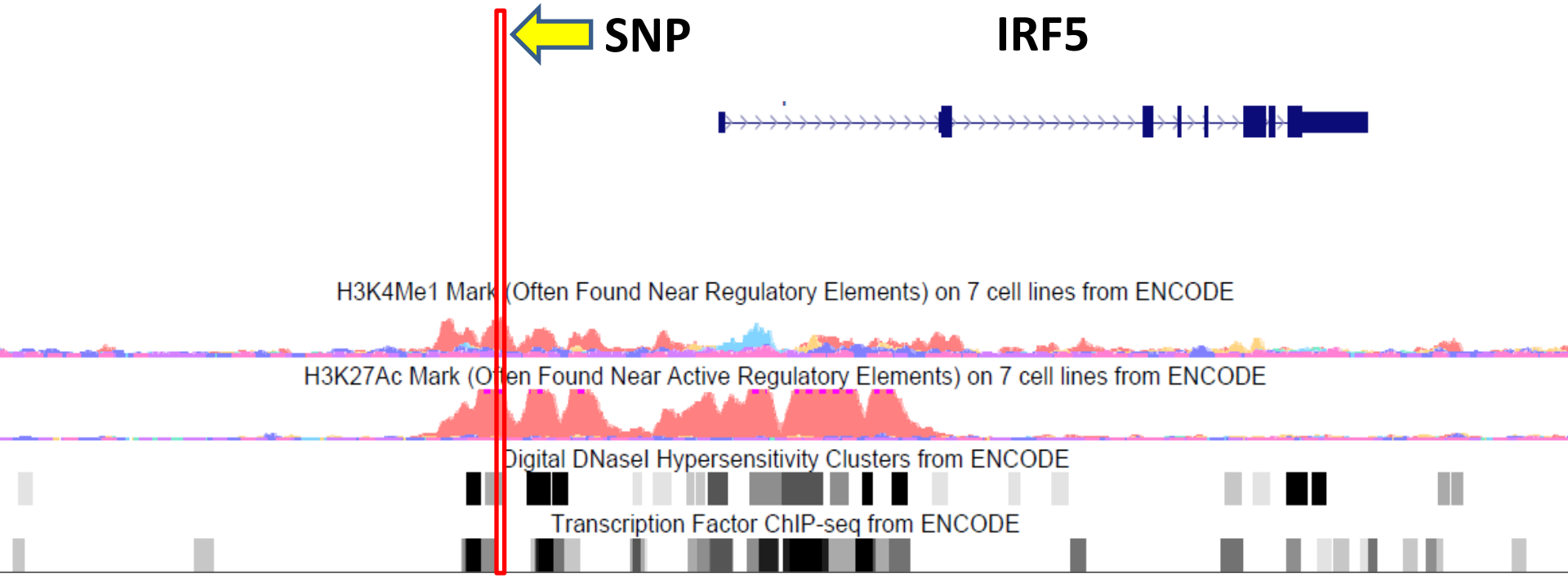
Genomic context of rs4728142



Genomic context of rs4728142



Genomic context of rs4728142



Prediction: stronger binding of Zbtb3

Zbtb3 binding motif

CGGTGTT (non-risk)

CAGTGTT (risk)

Quantitative score from PBM:
0.03 (won't bind) to 0.45 (strong binding)

Zbtb3 is highly expressed in B cells

Kottyan et al., Hum Mol Genet. 2014

Hypothesis:
the rs4728142 risk allele "creates" a binding site for ZBTB3 in B cells, resulting in higher expression of the IRF5 gene in lupus patients

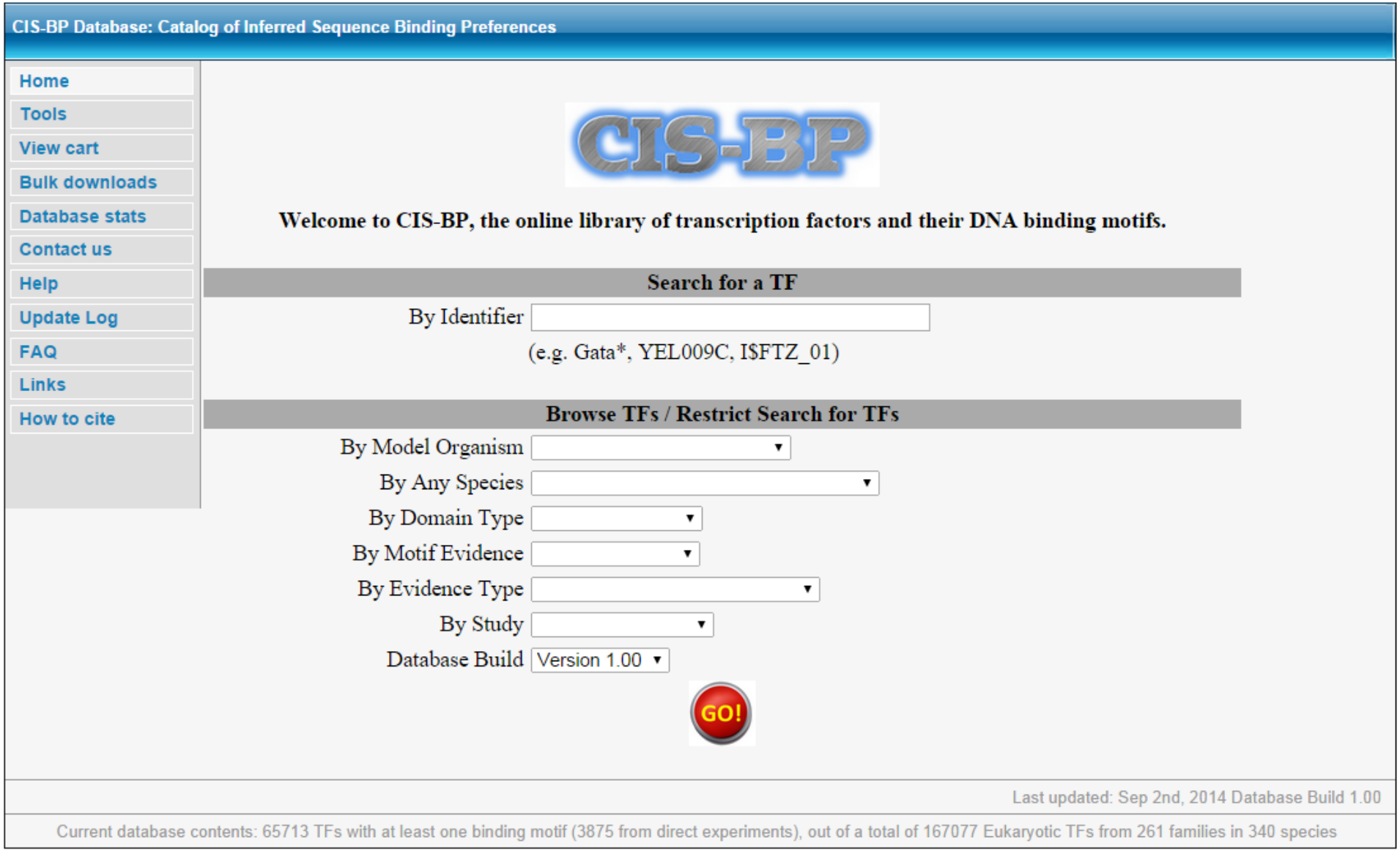
Kottyan et al., Hum Mol Genet. 2014

Hypothesis:
the rs4728142 risk allele "creates" a binding site for ZBTB3 in B cells, resulting in higher expression of the IRF5 gene in lupus patients

CONFIRMED VIA SUPERSHIFT EMSA

Kottyan et al., Hum Mol Genet. 2014

Website frontpage



http://cisbp.ccbr.utoronto.ca/

TF pages

CIS-BP Database: Catalog of Inferred Sequence Binding Preferences

- Home
- Tools
- Download cart
- Bulk downloads
- Database stats
- Contact us
- Help
- How to cite

Add to cart

CREB1 (*Homo sapiens*)
bZIP

TF Information

Pfam ID	Interpro ID	Gene ID	CIS-BP ID	Sequence source
PF00170 (bZIP_1)	IPR011616	ENSG00000118260	T086497	Ensembl (2011-Oct-26)

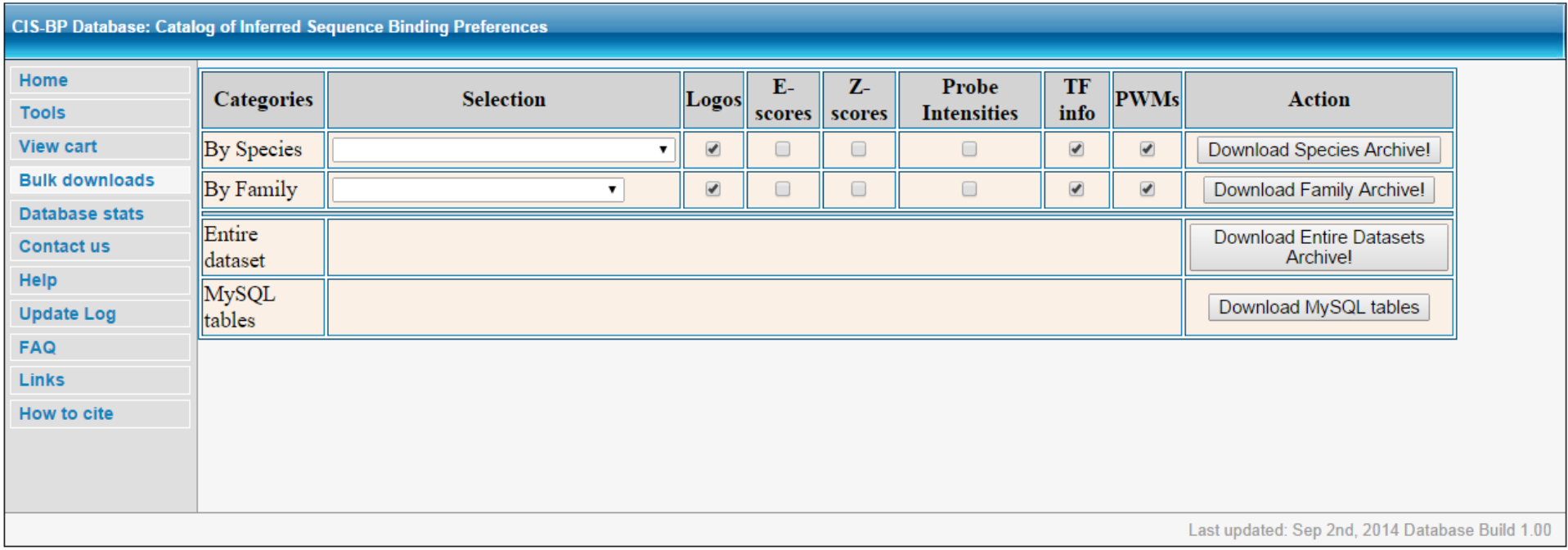
Directly determined binding motifs

Name/Motif ID	Species	Forward	Reverse	Type/Study/Study ID	DBD Identity
CREB1 M1305	*Homo sapiens*			SELEX Portales-Casamar et al.(2010) MA0018.1	(Direct)

Motifs from related TFs

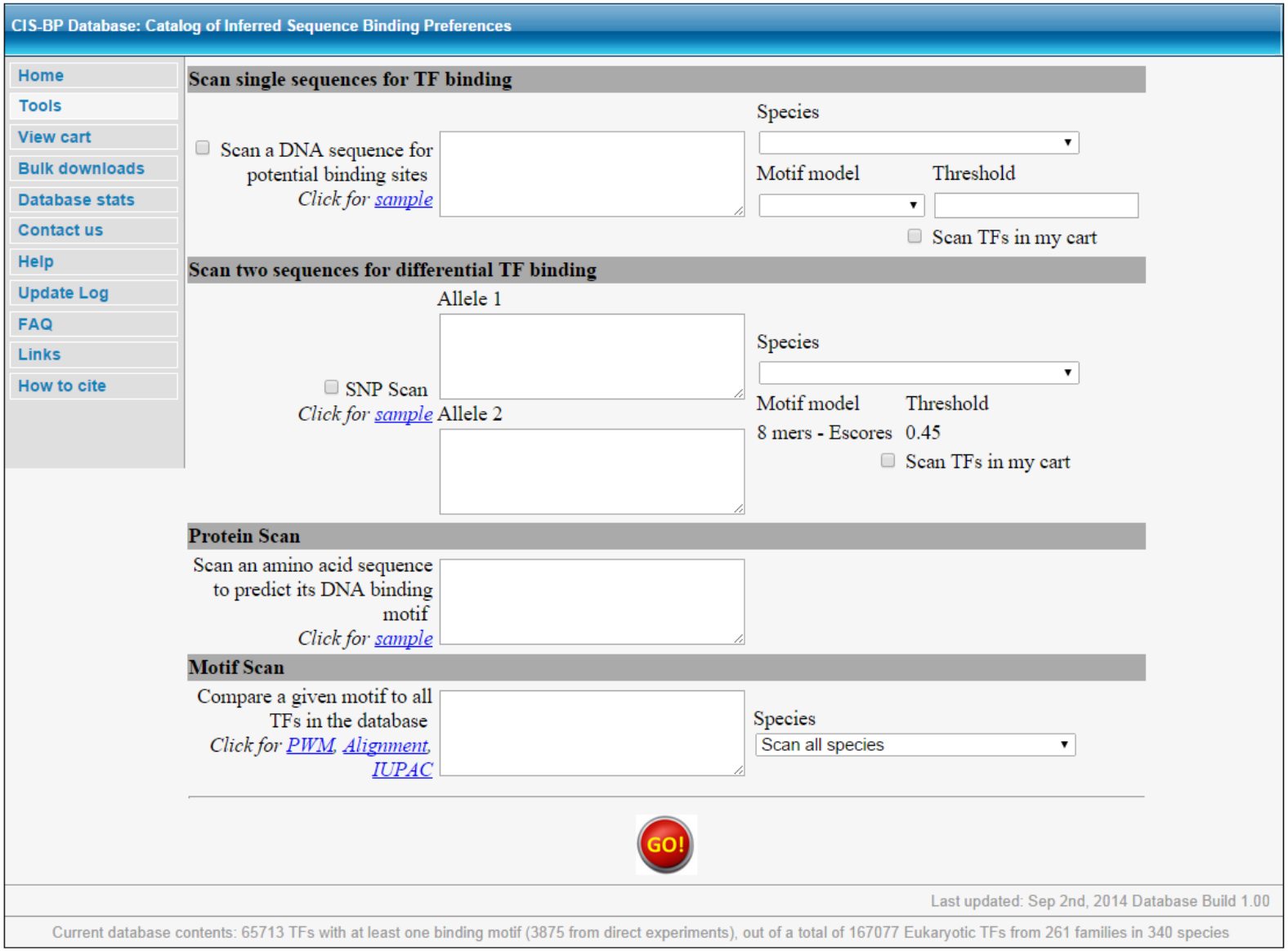
Name/Motif ID	Species	Forward	Reverse	Type/Study/Study ID	DBD Identity
Creb1 M0352	*Mus musculus*			PBM Zoo et al.(0) pTH5080	1.000
Crem M0375	*Mus musculus*			PBM Zoo et al.(0) pTH5002	0.934
Atf1	*Mus musculus*			PBM Zoo et al.(0)	0.852

Bulk Downloads



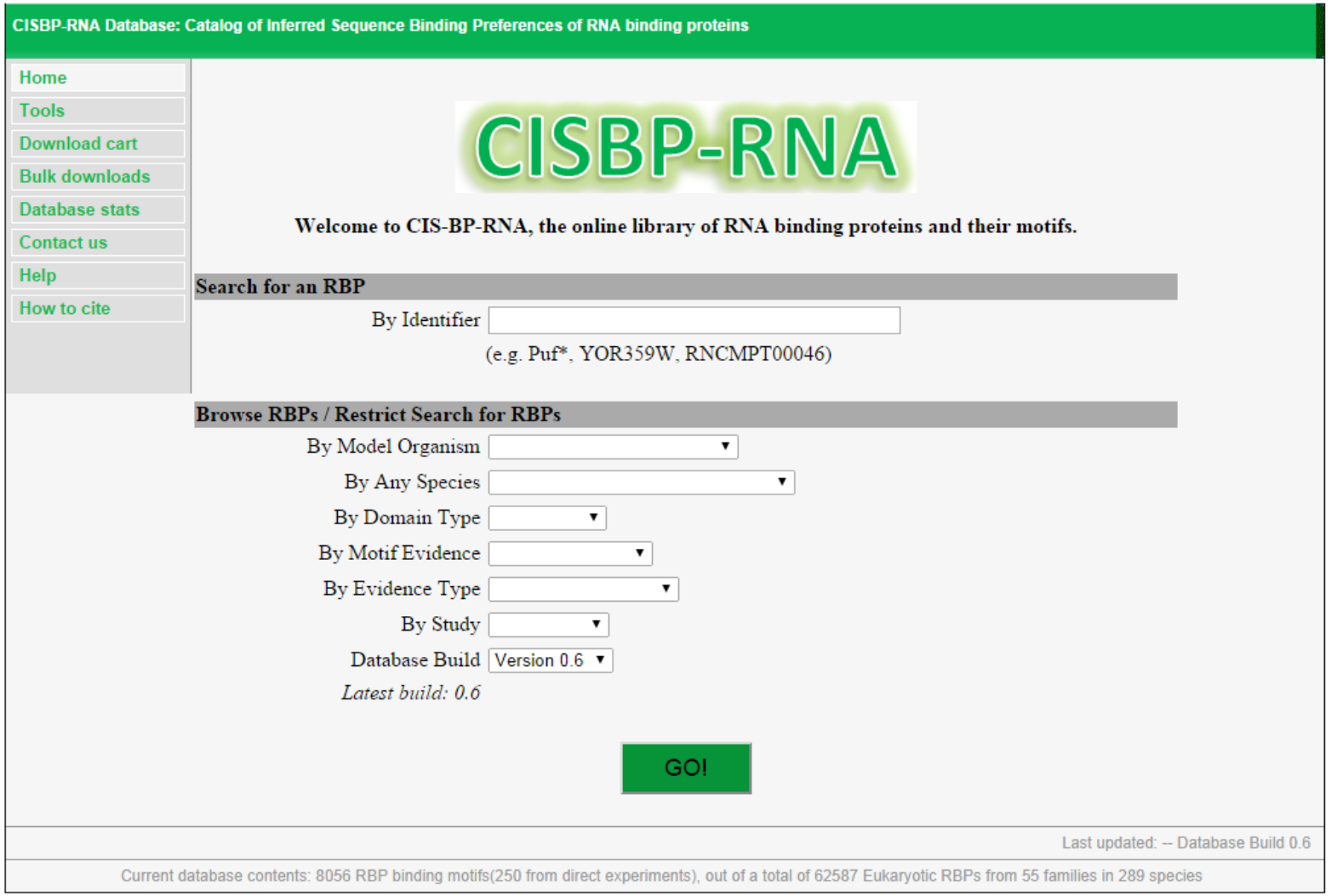
http://cisbp.ccbr.utoronto.ca/

Tools page



http://cisbp.ccbr.utoronto.ca/

There's one for RNA binding proteins too!



Ray, Weirauch *et al.*, Nature 2013

http://cisbp-rna.ccbr.utoronto.ca/

U. Toronto

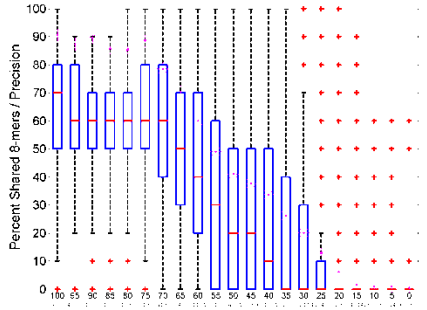
- Tim Hughes
- Ally Yang
- Mihai Albu
- Atina Cote
- Hong Zheng
- Hamed Shateri Najafabadi
- Sam Lambert
- Ishminder Mann
- Kate Cook
- Harm van Bakel
- Shaheynoor Talukder
- Andrew Vorobyov
- Anton van der Ven
- Wilfred de Vega
- Nicole Park
- Geanany Rasanathan
- Yogesh Hooda
- Sanie Mnaimneh
- Kenneth Chu
- Oliver Boright
- Jerry Li
- Agnieszka Janska

Elsewhere

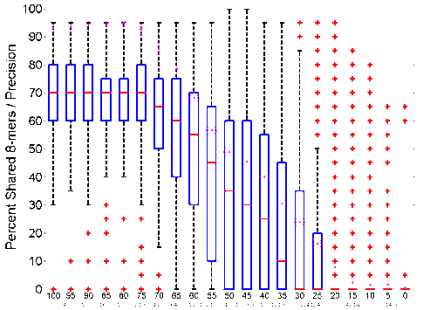
- Joseph Ecker
- Luis Larrondo
- Gunnar Ratsch
- Marian Walhout
- Alejandro Montenegro-Montero
- Philipp Drewe
- François-Yves Bouget
- Gadi Shaulsky
- Jean-Claude Lozano
- Mary Galli
- Mathew Lewsey
- Eryong Huang
- Tuhin Mukherjee
- Xiaoting Chen
- John Reece-Hoyes
- Sridhar Govindarajan
- Oliver Stegle

Ref:
Weirauch *et al.* Cell, 2014.
Contact:

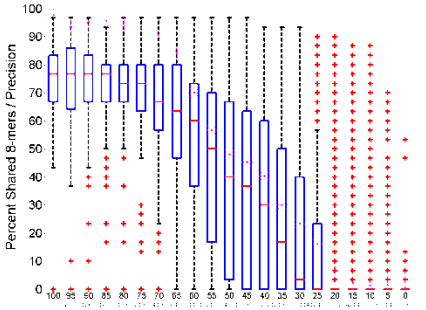
Homeodomains
(8 different measures of 8-mer simiarity)



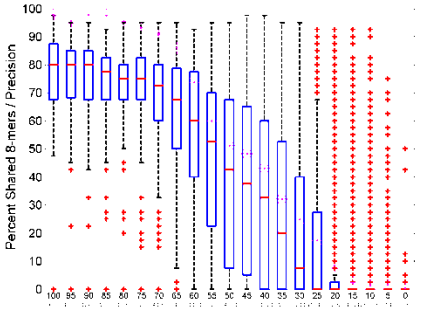
Zscore10



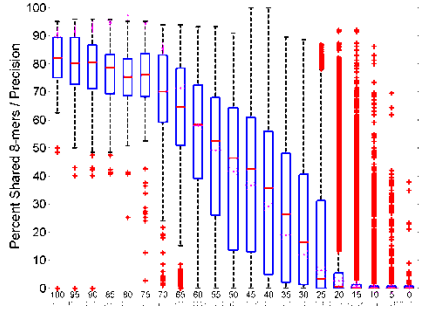
Zscore20



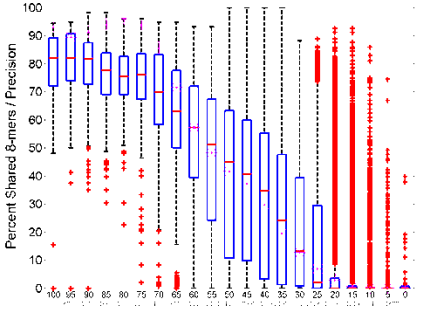
Zscore30



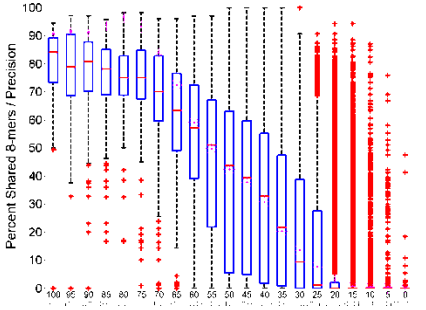
Zscore40



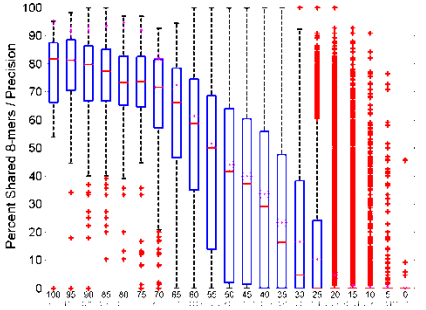
Escore45



Escore46

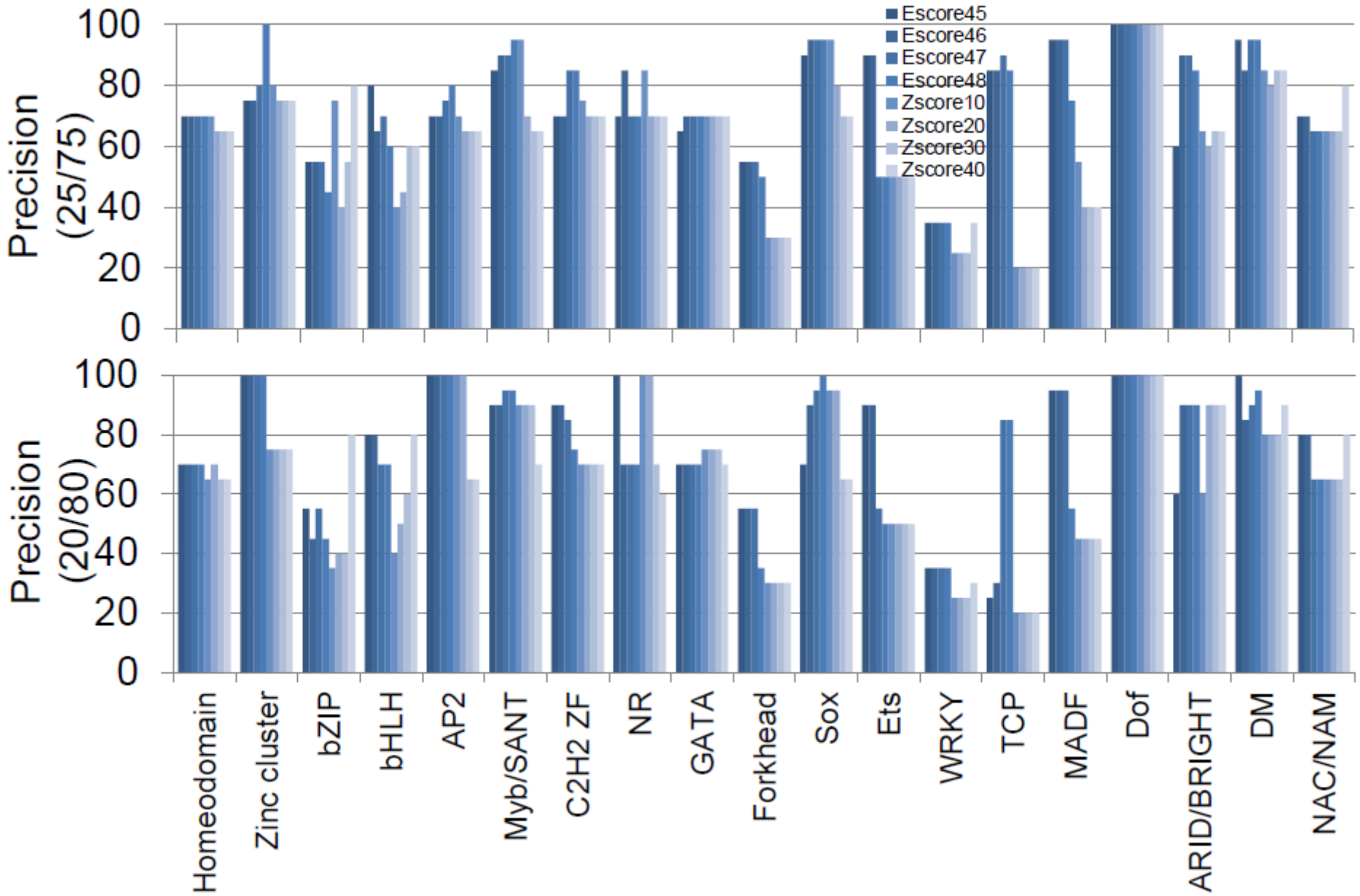


Escore47

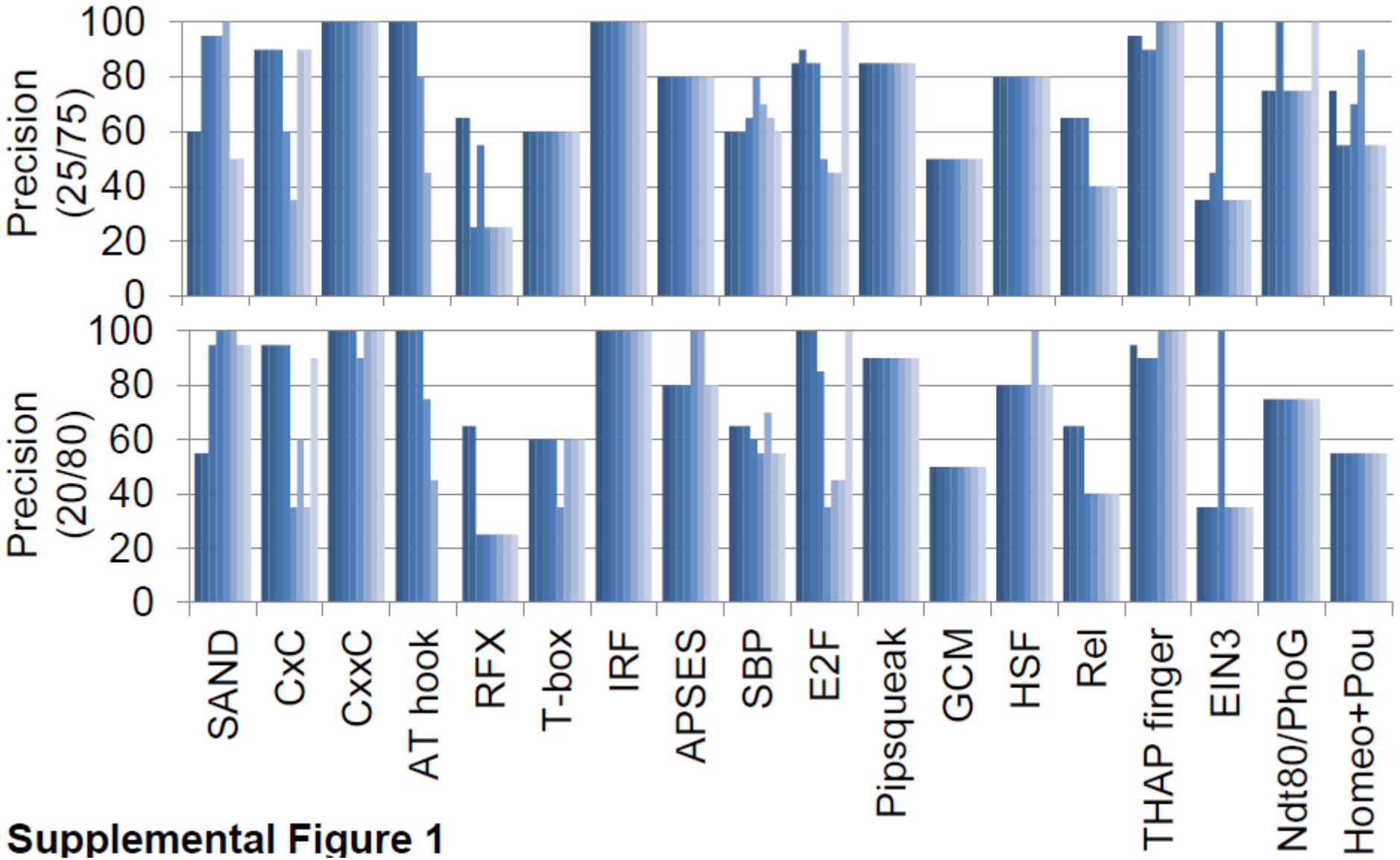


Escore48

Cutoffs across families and parameters

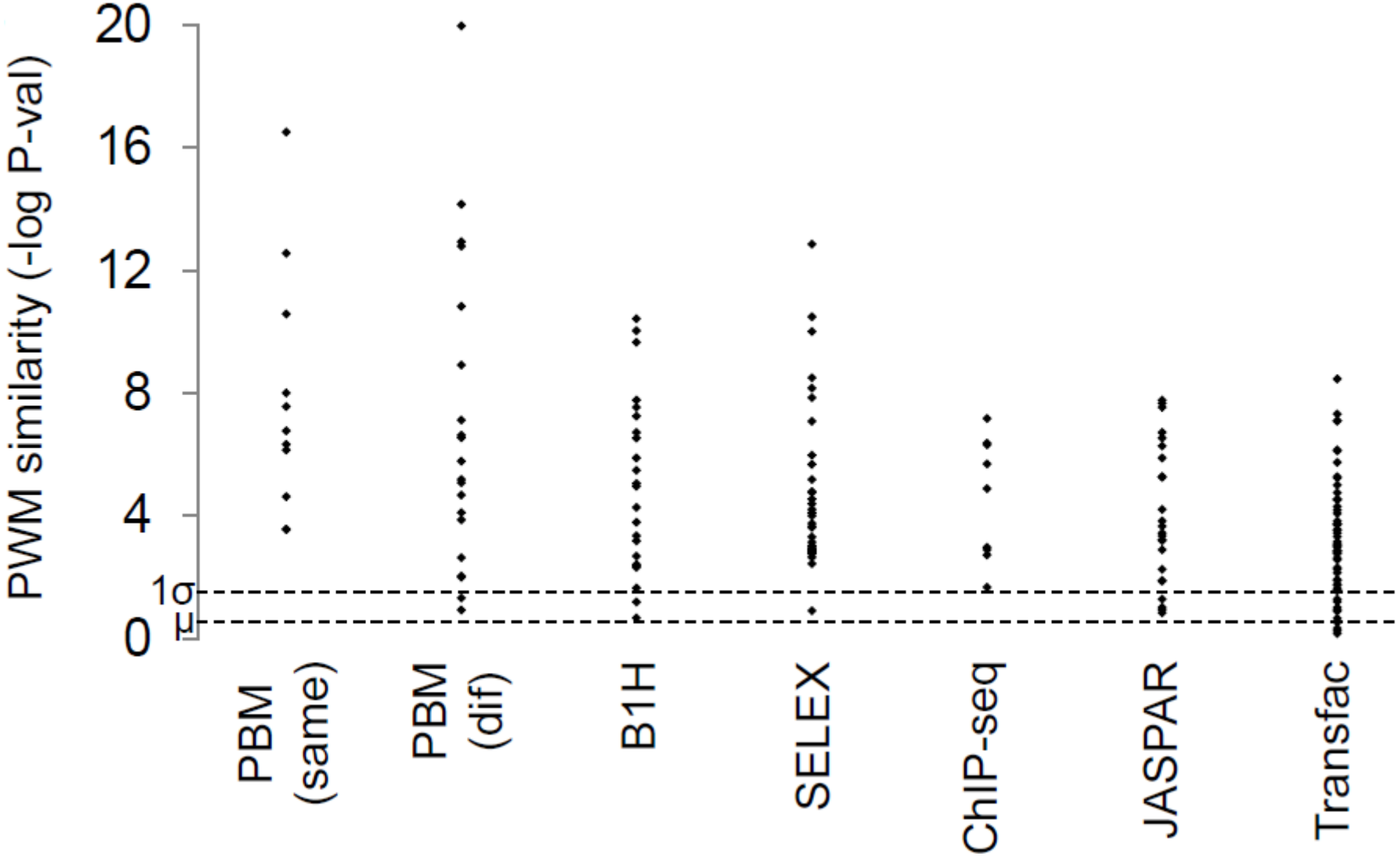


Cutoffs across families and parameters

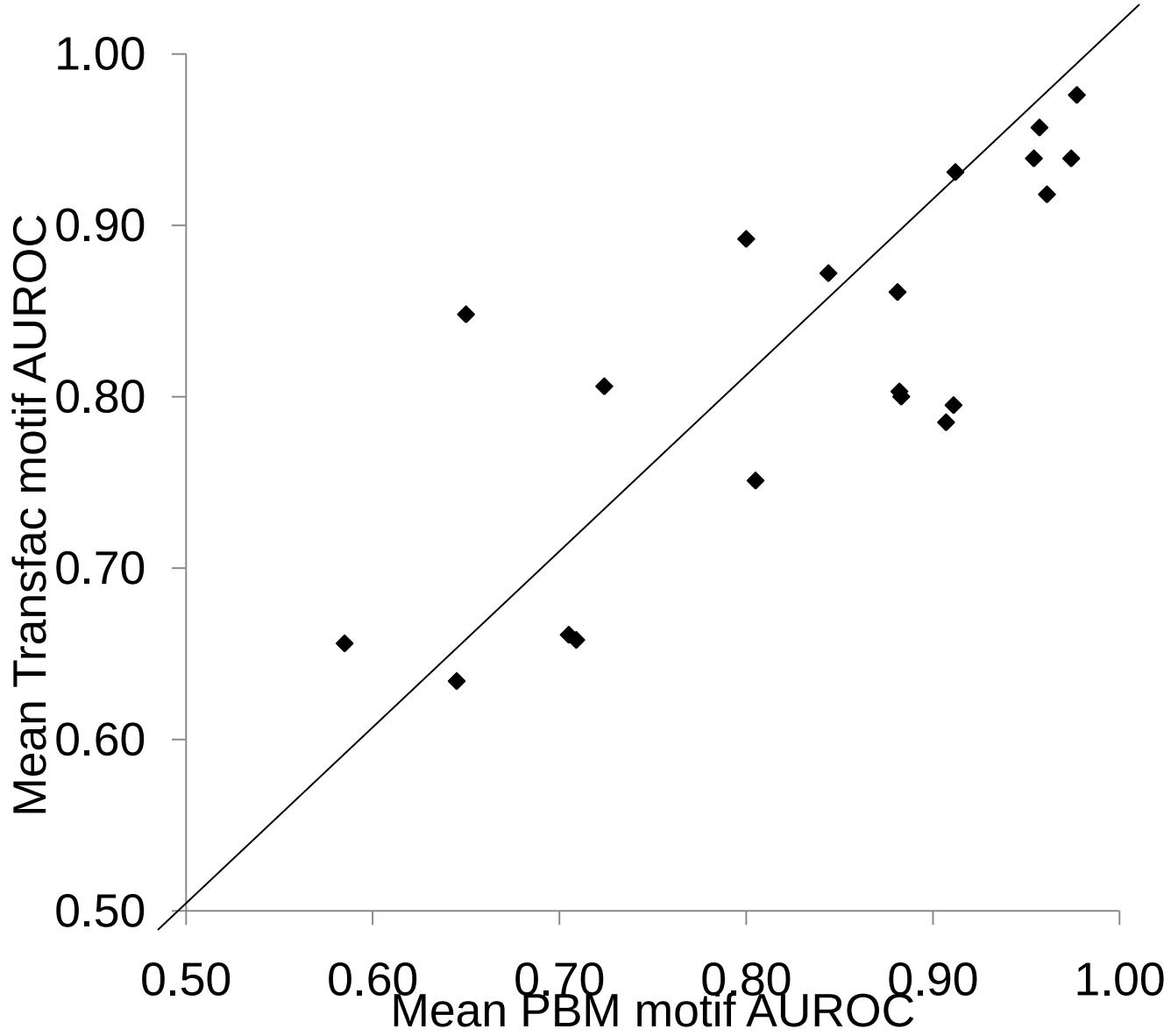


Supplemental Figure 1

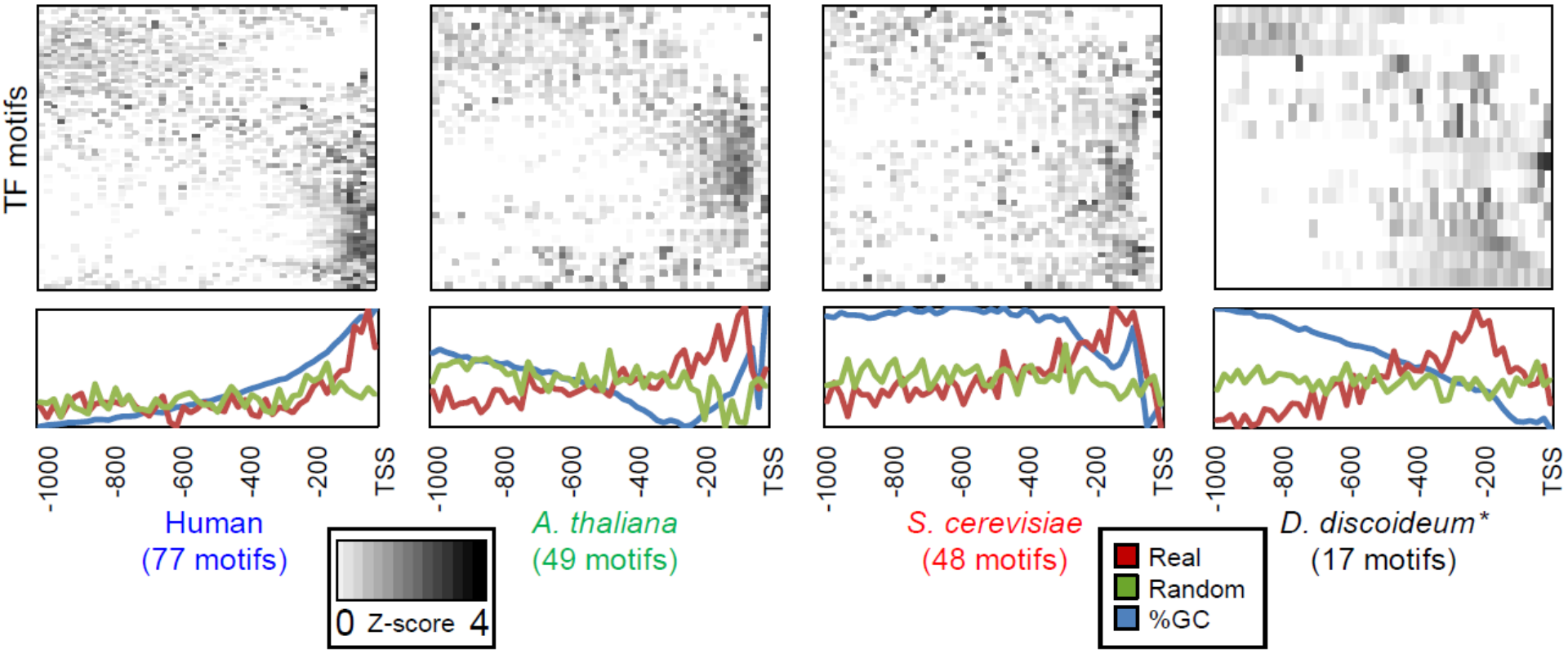
Our motifs vs. other studies



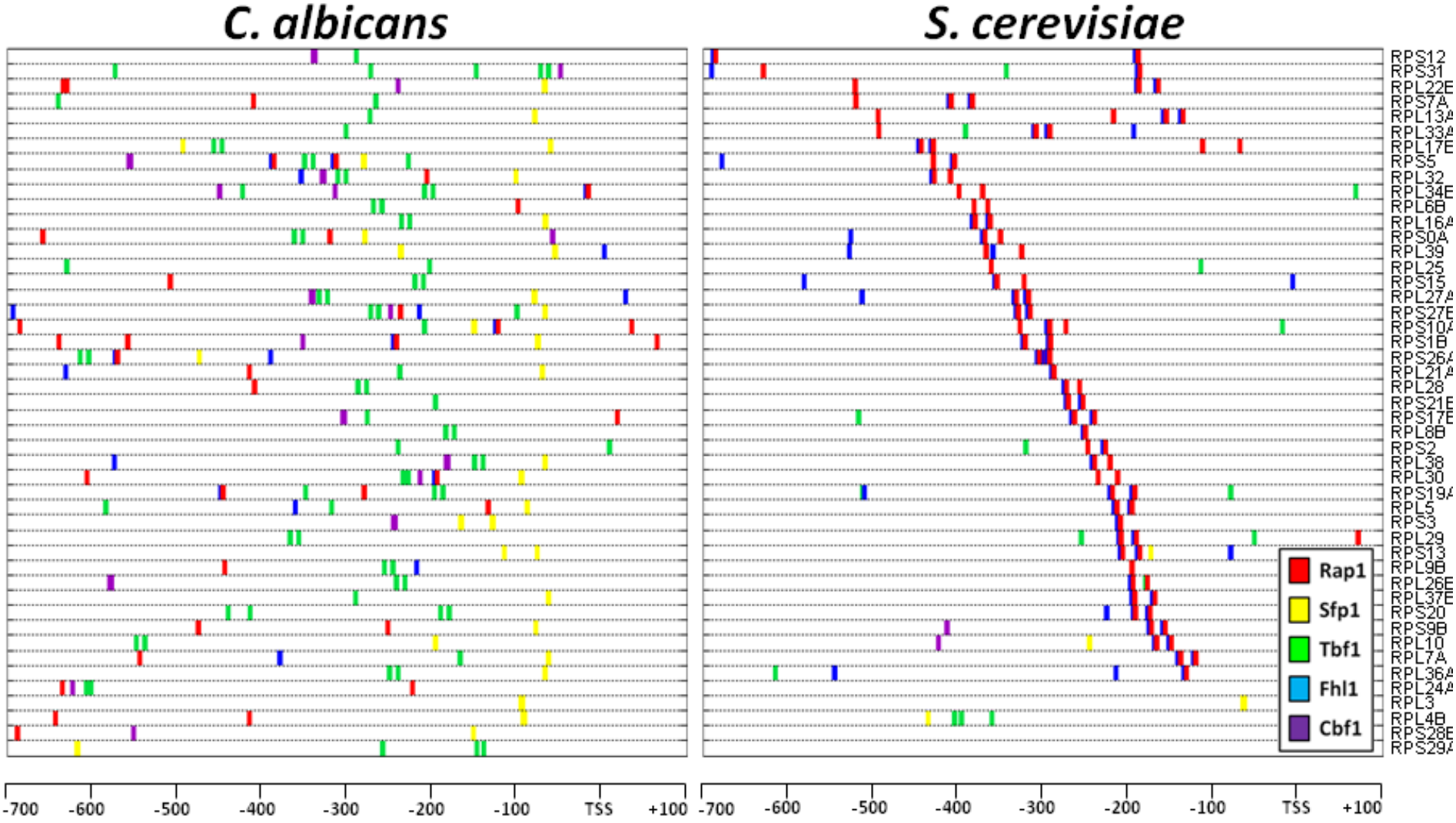
PBM-derived motifs predict ChIP peaks as accurately as literature-derived motifs



Positional bias of (non-redundant) motifs in eukaryotic promoters



Completely different in another yeast



Major TF controller switched from Tbf1 to Rap1