

Domain Distribution Comparison of Transcription Factor and Chromatin

Arli Parikesit

Chair of Bioinformatics

University of Leipzig

25th TBI Seminar

Bled, Slovenia

19 February 2010



Introduction

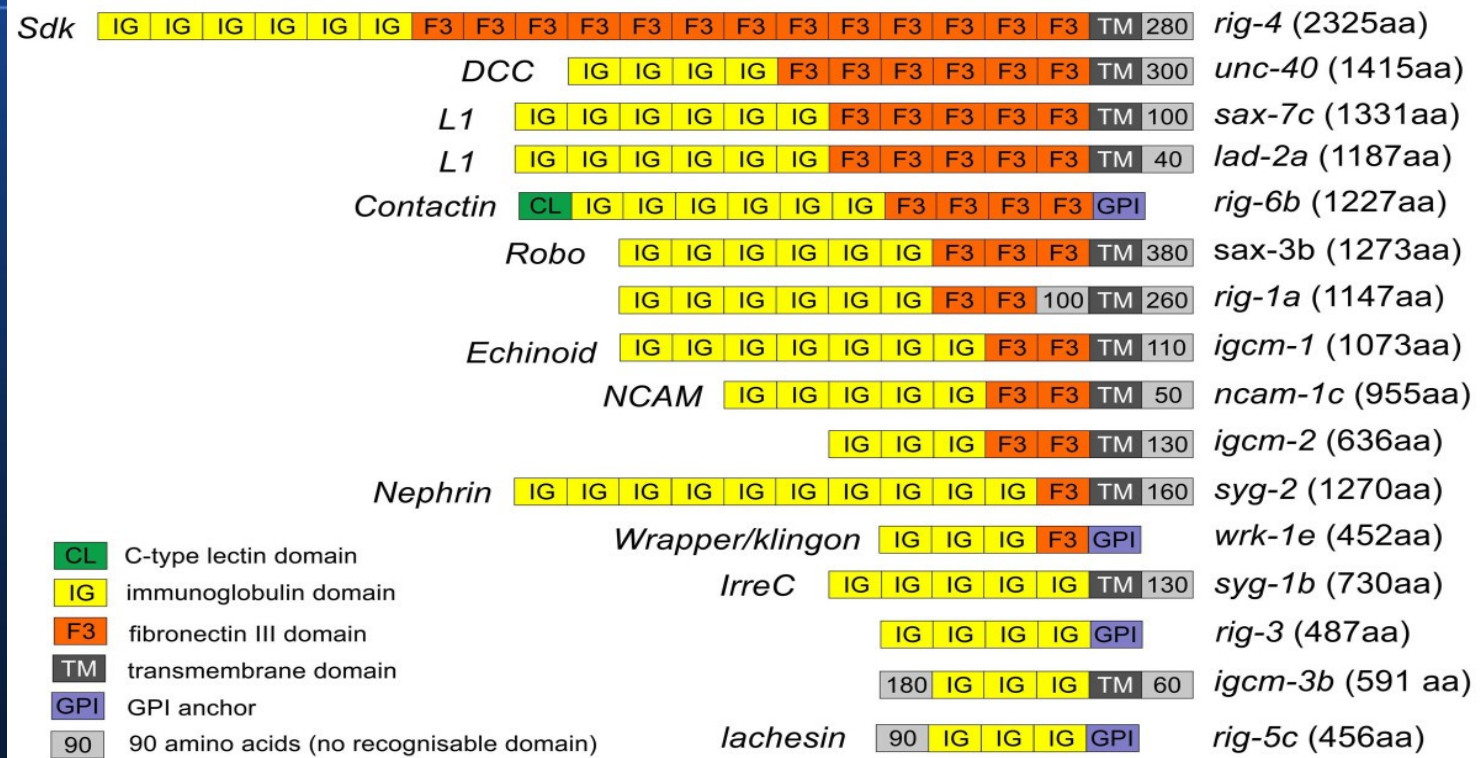
The complexity of gene regulatory mechanism

- Candidates for studying the regulatory mechanism have been selected. They are transcription factors, components of the miRNA machinery and proteins involved in chromatin regulation.
- Transcription factors are especially interesting because all domains of life have at least parts of them and they are involved in many regulatory processes in the cell. However, their interaction with the basal transcription machinery must be different in Archaea, Bacteria and Eukarya due to the differences in the machinery. Controls the transfer (transcription) of genetic information from DNA to RNA
- Chromatin modifications utilized as signals have been discovered in eukaryotes only, even though Archaea do have histones and enzymes carrying out chemical modifications are common among all living organisms (e.g. Phosphorylation). It facilitates DNA organization and prevent DNA aggregation and tangling which is important for replication, seggregation, and gene expression

The Necessity of Gene complexity

- More regulatory states.
- Higher stability (precision).
- More fine grained regulation

Fundamental and Functional unit of protein structure is the domain.



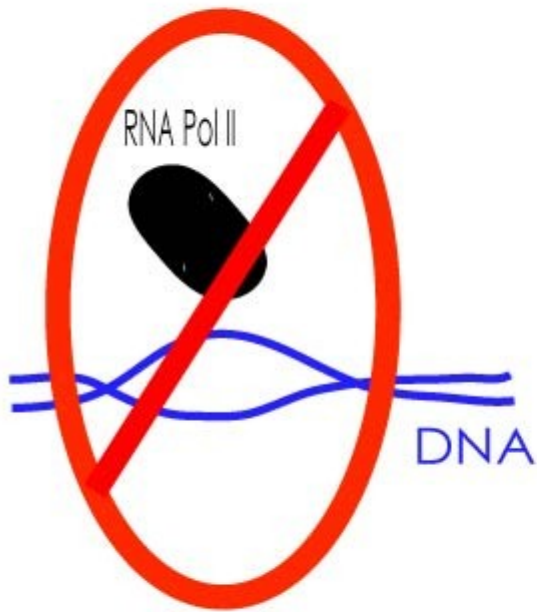
Independent of neighboring sequences, a domain folds into a distinct structure and mediates a biological functionality.

Research Objective

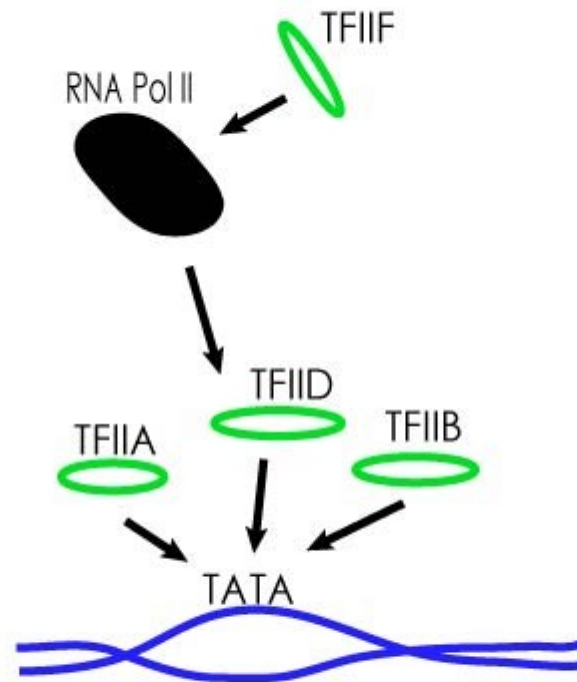
- Description of the functional evolution of the regulatory mechanisms. This will allow us to have insights into major innovations, adaptive changes, and the concepts of regulatory network evolution.
- Description of the occurrence of the gene regulation domain. Ongoing research: Micro RNA, Chromatin and Transcription factor activator domains.
- Evolutionary relationship between each type of regulation

Transcription Regulation

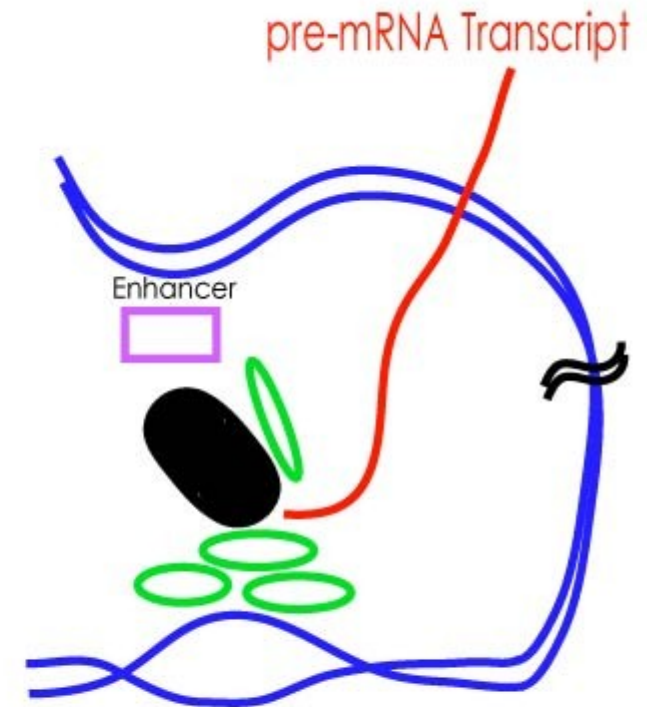
The Wonders of Transcription!



Eukaryotic RNA Polymerases cannot bind DNA alone!

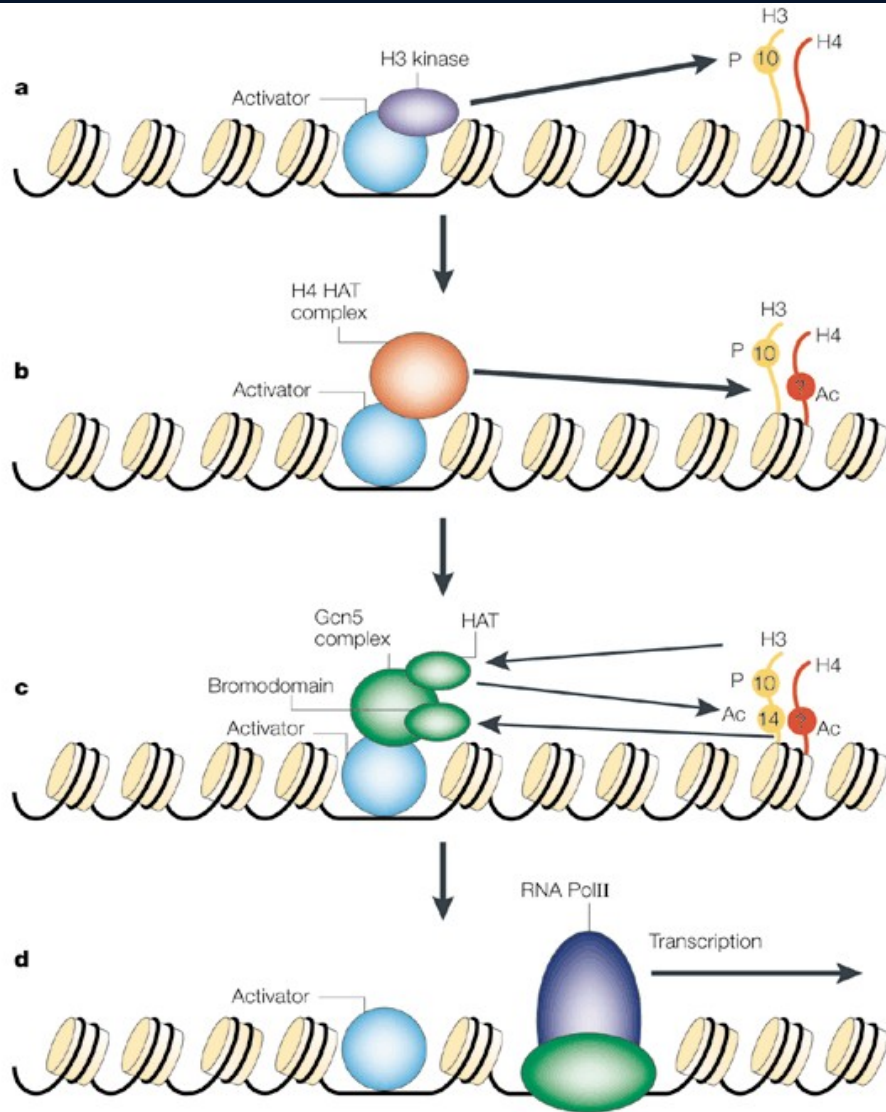


Basal transcription factors are **REQUIRED** for initiation!

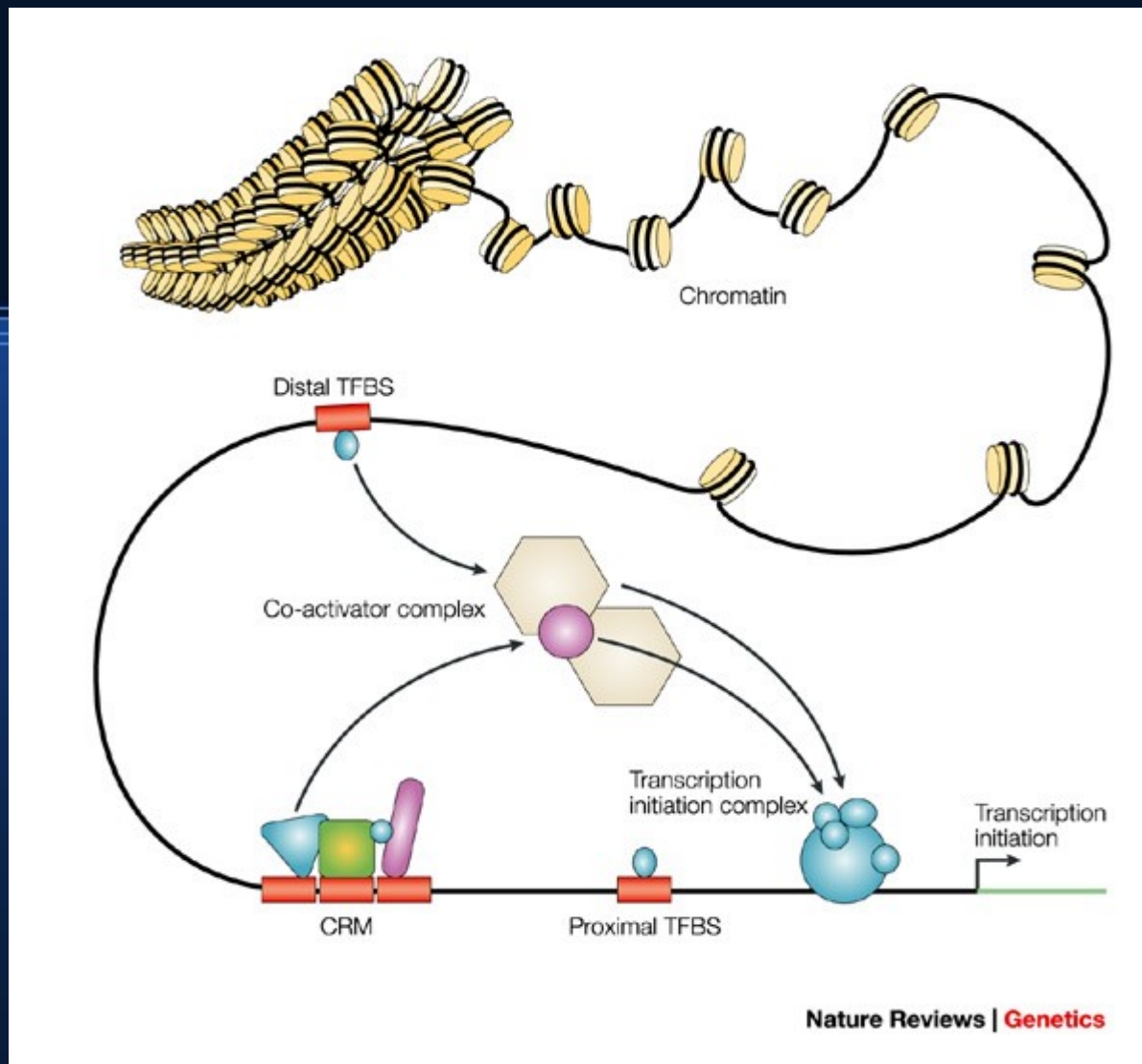


Distant enhancer sequences can also affect transcription!

Chromatin Regulation



- a)** In the first step, a histone H3 kinase is recruited to the promoter by the transcriptional activator to phosphorylate Ser10 on histone H3.
- b)** A histone H4 HAT (histone acetyltransferase) complex is recruited to the promoter to acetylate a lysine residue on histone H4.
- c)** The Gcn5 HAT complex is recruited to the promoter, possibly through multiple interactions including an interaction between the bromodomain of Gcn5 with acetyl-lysine modified histone H4 and an interaction between a phosphorylated histone H3 tail with the HAT domain. This recruitment results in the acetylation of Lys14 on histone H3.
- d)** The effect of these post-translational histone modifications is to recruit RNA polymerase II and the GTFs (general transcription factors) to DNA for transcriptional activation. Ac, acetylation; P, phosphorylation.



- Transcription factors (TFs) bind to specific sites (transcription-factor binding sites; TFBS) that are either proximal or distal to a transcription start site. Sets of TFs can operate in functional cis-regulatory modules (CRMs) to achieve specific regulatory properties. Interactions between bound TFs and cofactors stabilize the transcription-initiation machinery to enable gene expression. The regulation that is conferred by sequence-specific binding TFs is highly dependent on the three-dimensional structure of chromatin.

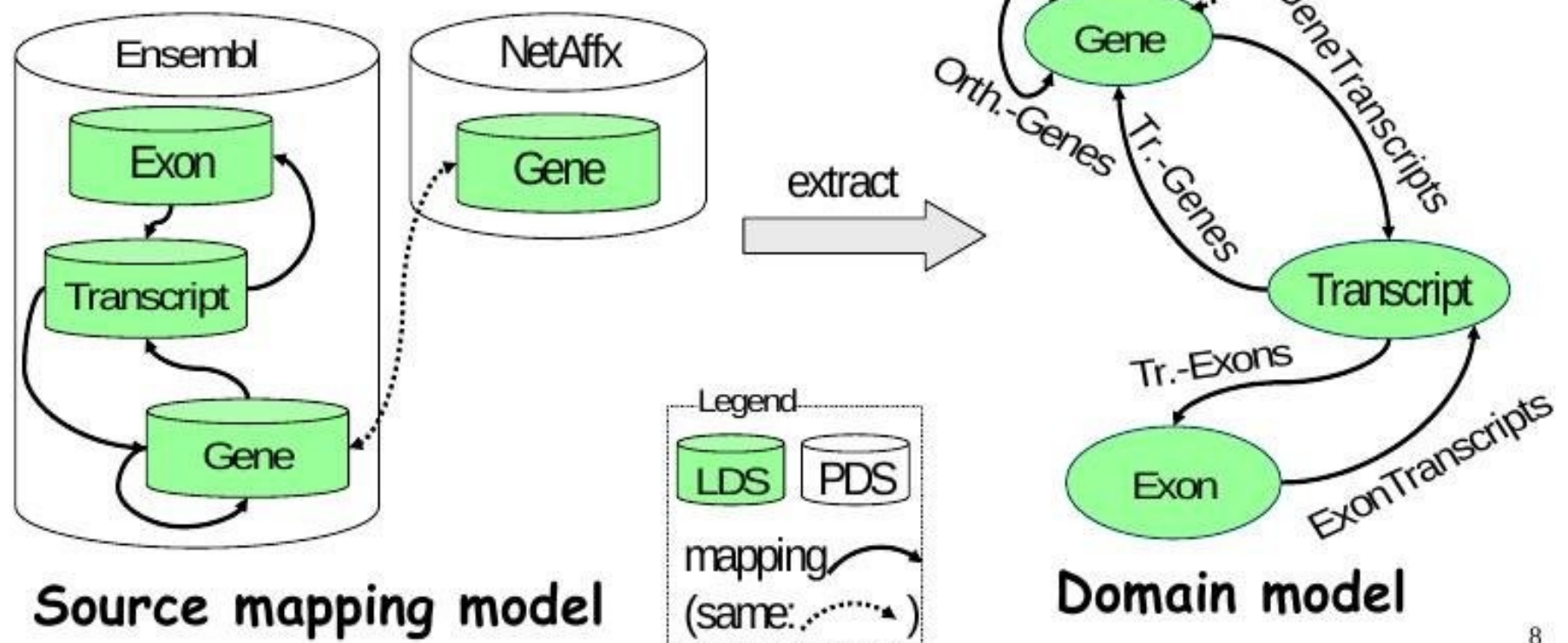
Method

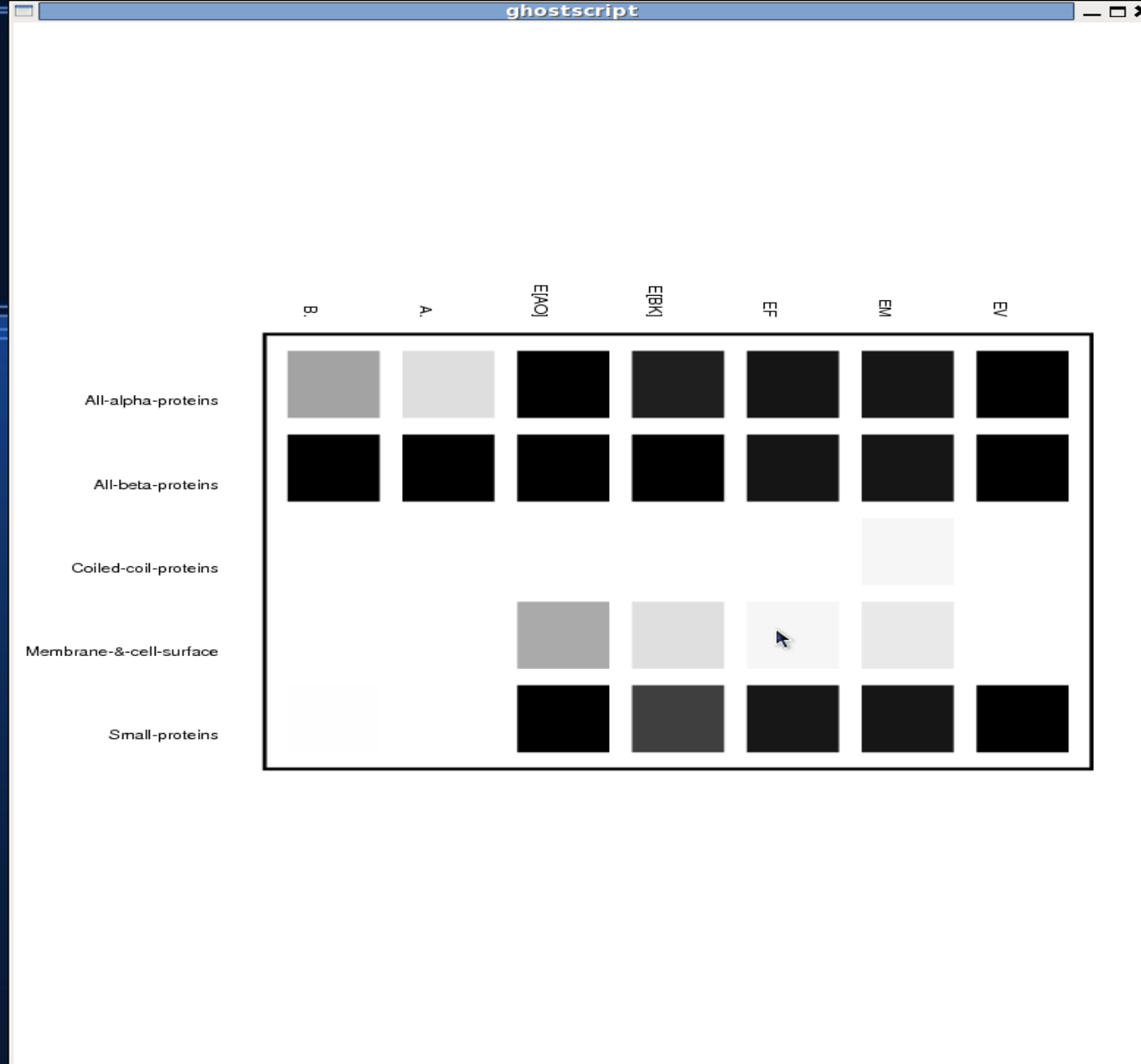
Domain Distribution on Phylogenetic trees:

- Extract the number of genes with domain X for every species.
- Compute the domain cooccurrence.
- Plot domain cooccurrence into the Graph.

Tools used: Toralf's Biofuice (Trans factor) and Sonja's SUPER script

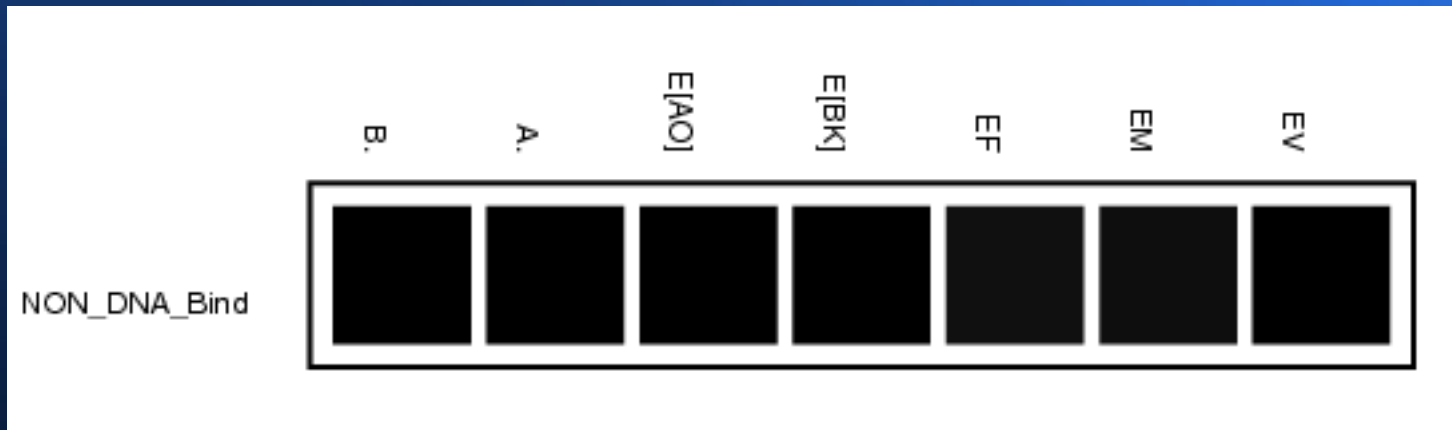
- Used by mediator for mapping/operator execution
- **Domain model** indicates available object types and relationships



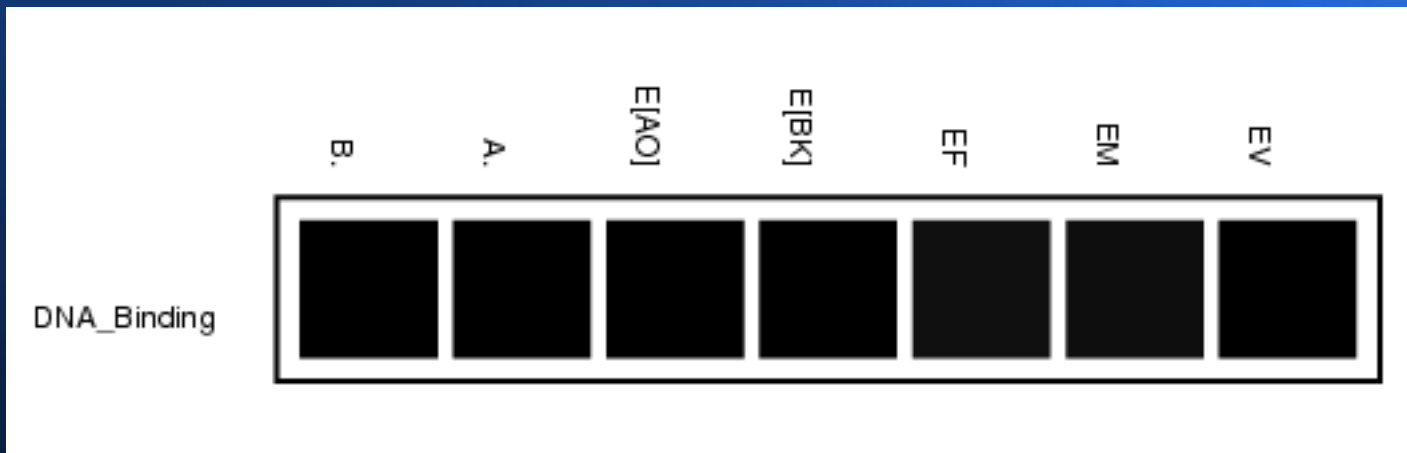


- The Transcription Factor Structural Domain Distribution. Totally no annotation for Coiled protein. No annotation for Bacteria and Arachea for Membrane and small protein

Non DNA Binding Domain annotation



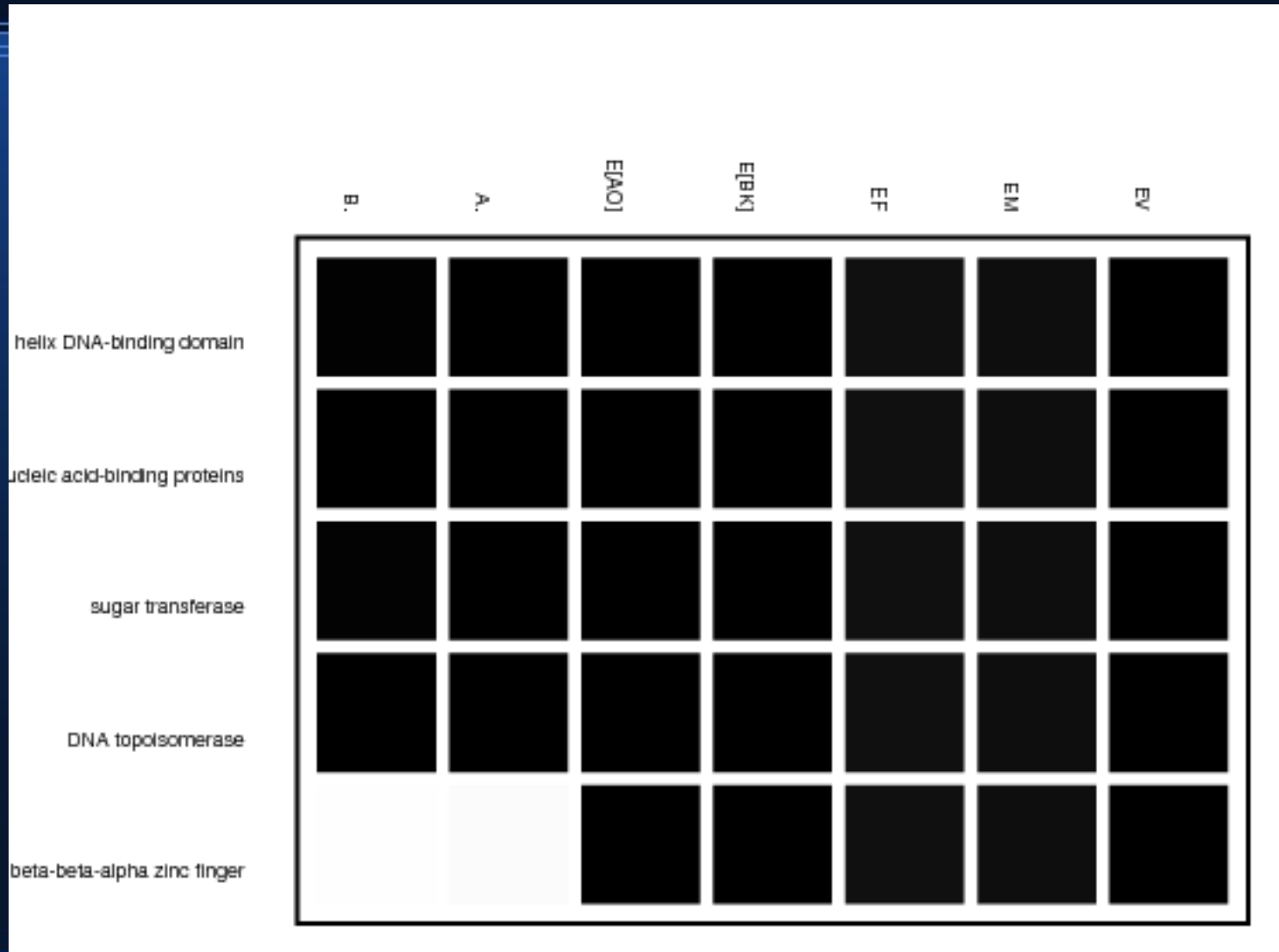
DNA Binding domain annotation



Non DNA Binding Domain Distribution

| | B. | A. | [EAO] | [EAK] | EF | EM | EV |
|-------------------------|----|----|-------|-------|----|----|----|
| MFS transporter | | | | | | | |
| Rossmann-fold domains | | | | | | | |
| triphosphate hydrolases | | | | | | | |
| methyltransferase | | | | | | | |
| Protein kinase-like | | | | | | | |

DNA Binding Domain Distribution

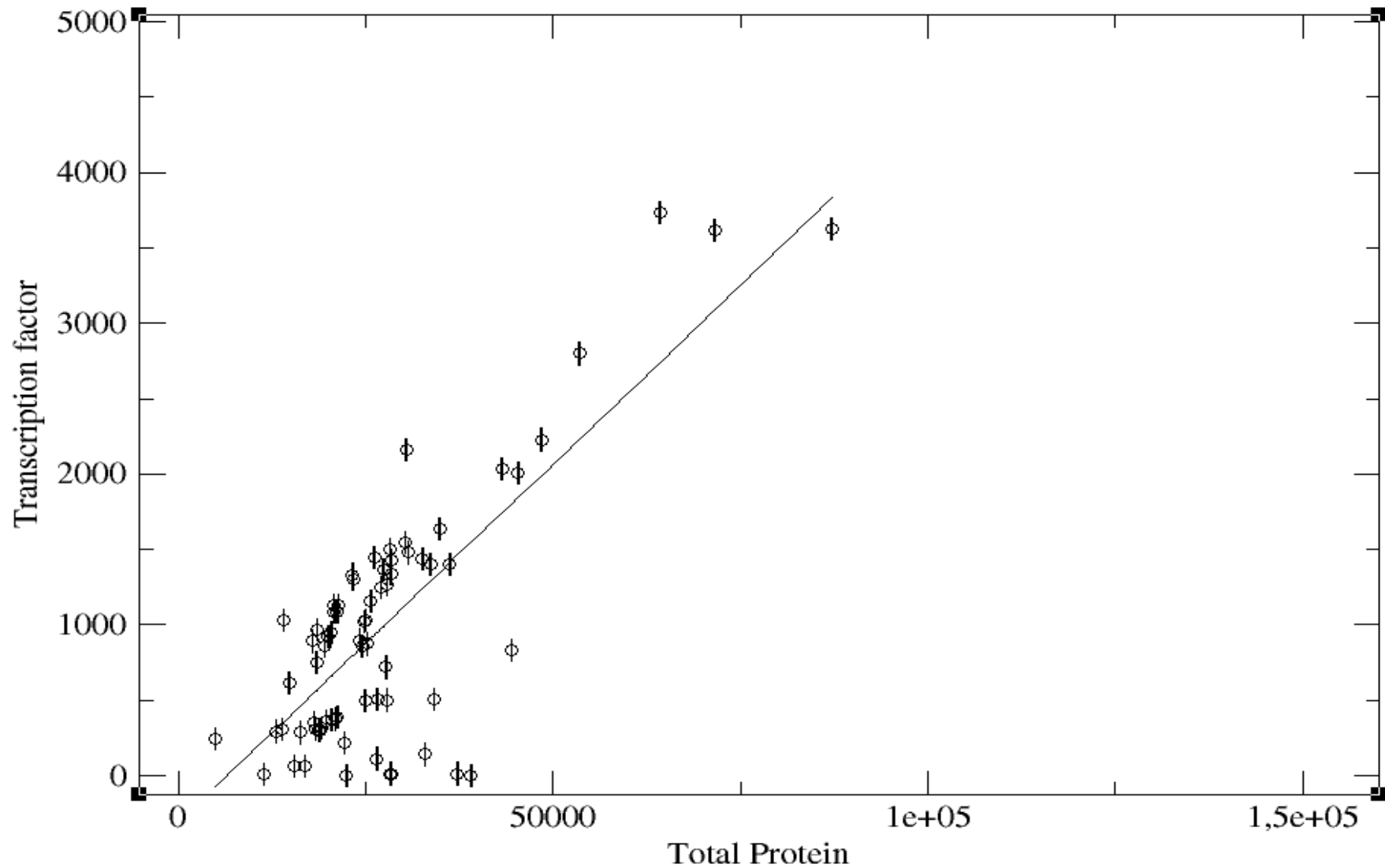


Domain Distribution

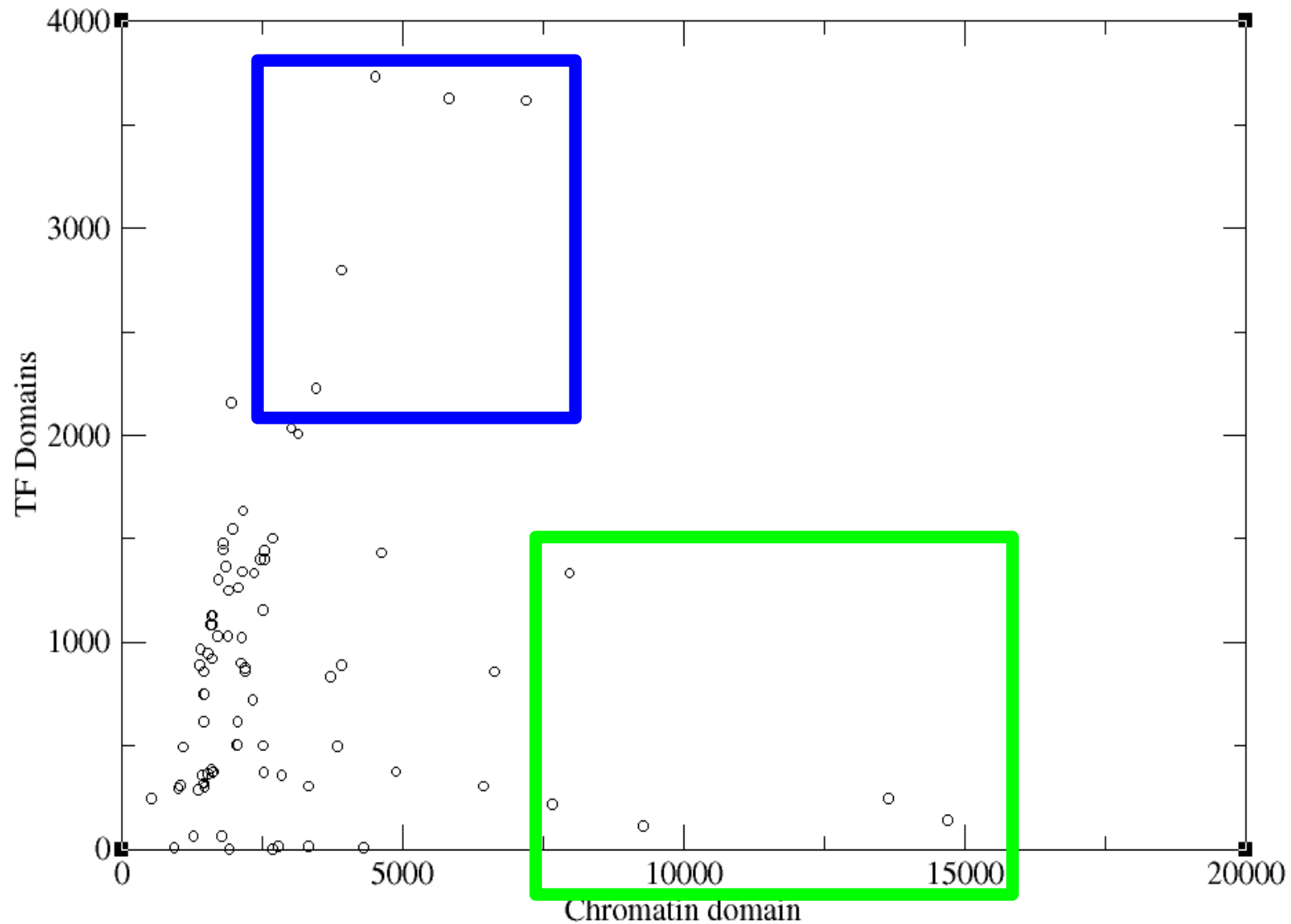
- The most widely available Transcription Factor (TF) Domain: 52540 (P-loop containing nucleoside triphosphate hydrolases)
- The most widely available DNA Binding TF: 46785 (Winged Helix DNA Binding Domain)
- The most widely available Non DNA Binding TF: 52540

Why P Loop Domain? Because of its assistance in the assembly, operation, or disassembly of protein complexes

Transcription Factor vs Total Domain Plot



Transcription Factor vs Chromatin Domain



Species with large amount of annotated TF domain

hs --> human

in --> Kangaroo rat

mm --> mouse

op --> opossum

to --> Fugu

Species with large amount of annotated Chromatin domain

is --> sea squirt (ciona)

o1 --> Medaka (japan kill f sh)

o1 --> Medaka [different data set]

sc --> yeast

u1 --> Dolphin

The result is phylogenetically ambiguous, because Yeast and Dolphin can't be on the same clade

Ambiguous Annotation

- Chromo Domain, which is a Chromatin regulation Domain, is annotated as Transcription Factor in SCOP Database → Wrong annotation

Conclusion

- Structural Domain Distribution: No annotation for Coiled protein. No annotation for Bacteria and Arachea for Membrane and small protein.
- Functional Domain Distribution: It's annotated in every clade, for DNA binding and non DNA binding domain

Conclusion

- The most widely available Transcription Factor (TF) Domain: 52540 (P-loop containing nucleoside triphosphate hydrolases)
- The more total domain in a species will give more Transcription Factor domain
- The Transcription Factor versus Chromatin Domain graph is giving a phylogenetically ambiguous result

Outlook

- Different strategy to annotate transcripts/proteins is necessary
- Pool all transcript and rerun the HMM engine to the whole genome

Reference

Chen, Kevin,. Rajewsky, Nikolaus: The Evolution of gene regulation by Transcription Factors and MicroRNAs. Nature Vol 8. pp 93-103.

Kleene, Kenneth: Patterns of Translational Regulation in the Mammalian Testis. Molecular Reproduction and Development 1996. 43:268-281.

Mann, Mathias., Jensen, Ole N: Proteomic Analysis of Post Translational Modifications. Nature review 2003 Vol 21. pp 255-261

Reference

Reik, Wolf: Stability and Flexibility of epigenetic gene regulation in Mammalian Development. Nature Vol 447. pp 425 -432.

S Wuchty, E. Almaas: Evolutionary cores of domain co-occurrence networks. BMC Evolutionary Biology 2005, 5:24

THANKS TO...

Sonja Prohaska

Axel Wintsche

Toralf Kirsten

Peter Stadler

DAAD

THANK YOU FOR YOUR ATTENTION