Enhancers

Day 4

info@genexplain.com

Enhancer

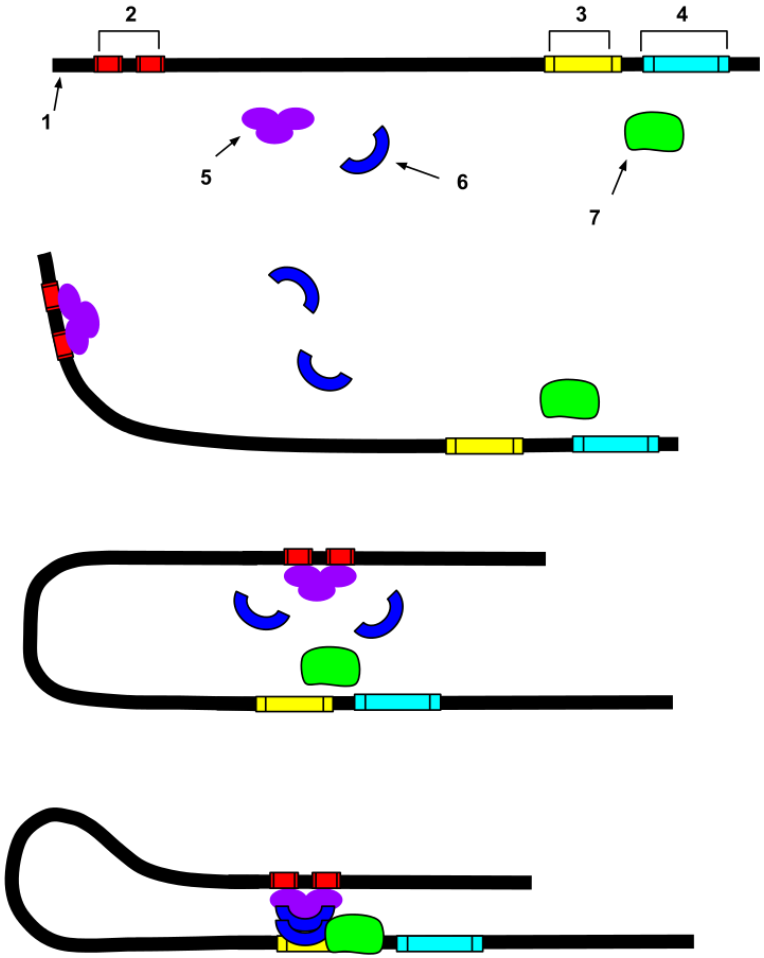
Enhancer (genetics)

From Wikipedia, the free encyclopedia

- In genetics, an **Enhancer** is a short (50–1500 bp) region of DNA that can be bound by proteins (activators) to increase the likelihood that transcription of a particular gene will occur.
- These proteins are usually referred to as transcription factors.
- Enhancers are cis-acting. They can be located up to 1 Mbp (1,000,000 bp) away from the gene, upstream or downstream from the start site.
- There are hundreds of thousands of enhancers in the human genome.
- Enhancers are found in both prokaryotes and eukaryotes.

info@genexplain.com

Usage of an Enhancer

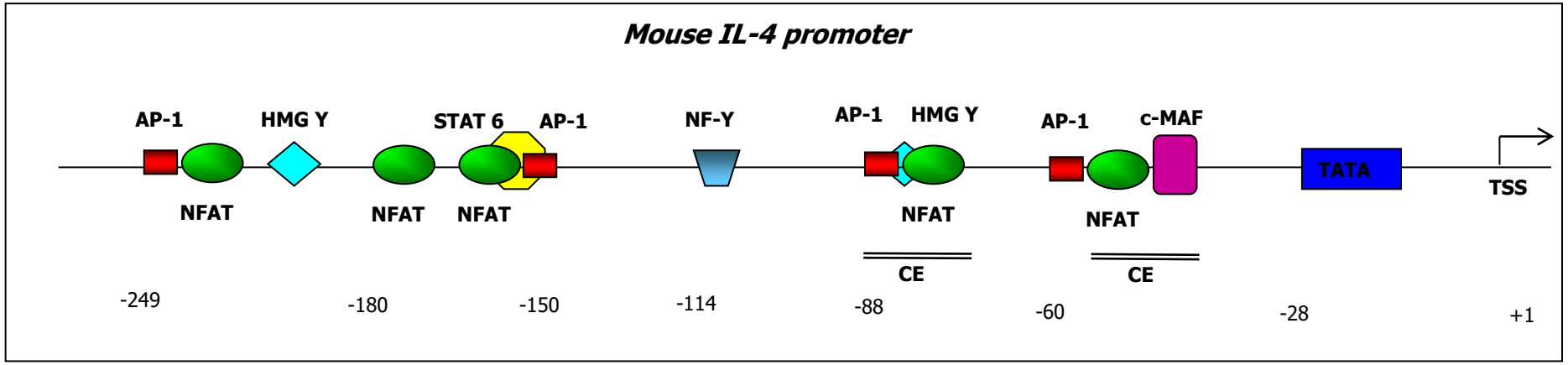


1. DNA
2. Enhancer
3. Promoter
4. Gene
5. Transcription Activator Protein
6. Mediator Protein
7. RNA Polymerase

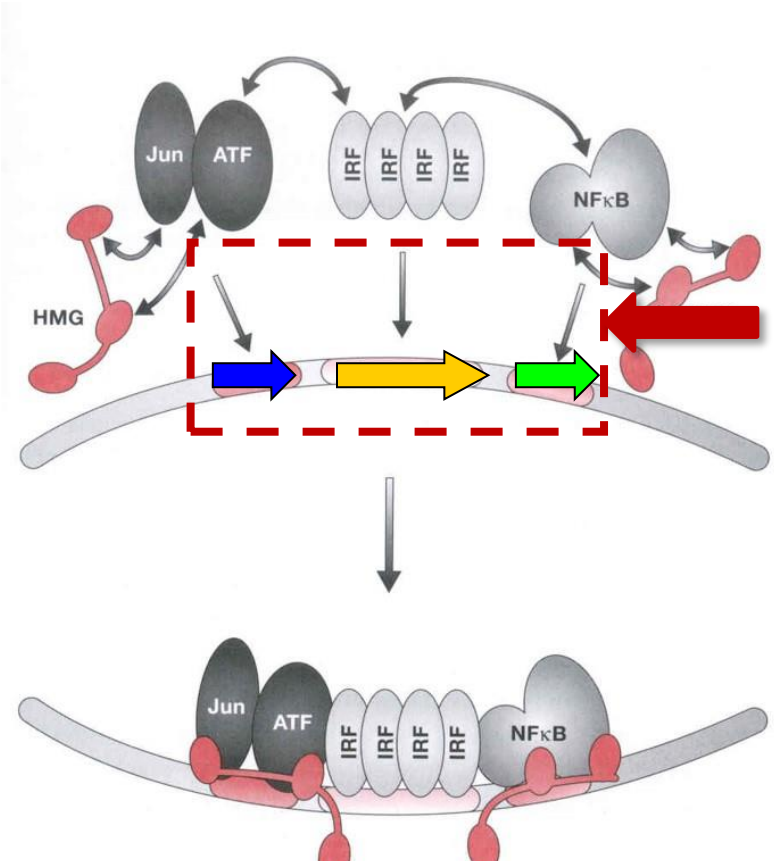
- Seen here is a four-step diagram depicting the usage of an enhancer.
- Within this DNA sequence, proteins known as transcription factors bind to the enhancer and increase the activity of the promoter.

info@genexplain.com

Promoter and enhancer: Mouse IL-4 as example



Promoter and enhancer: Human Interferon-ß as example



Composite Module (CM)

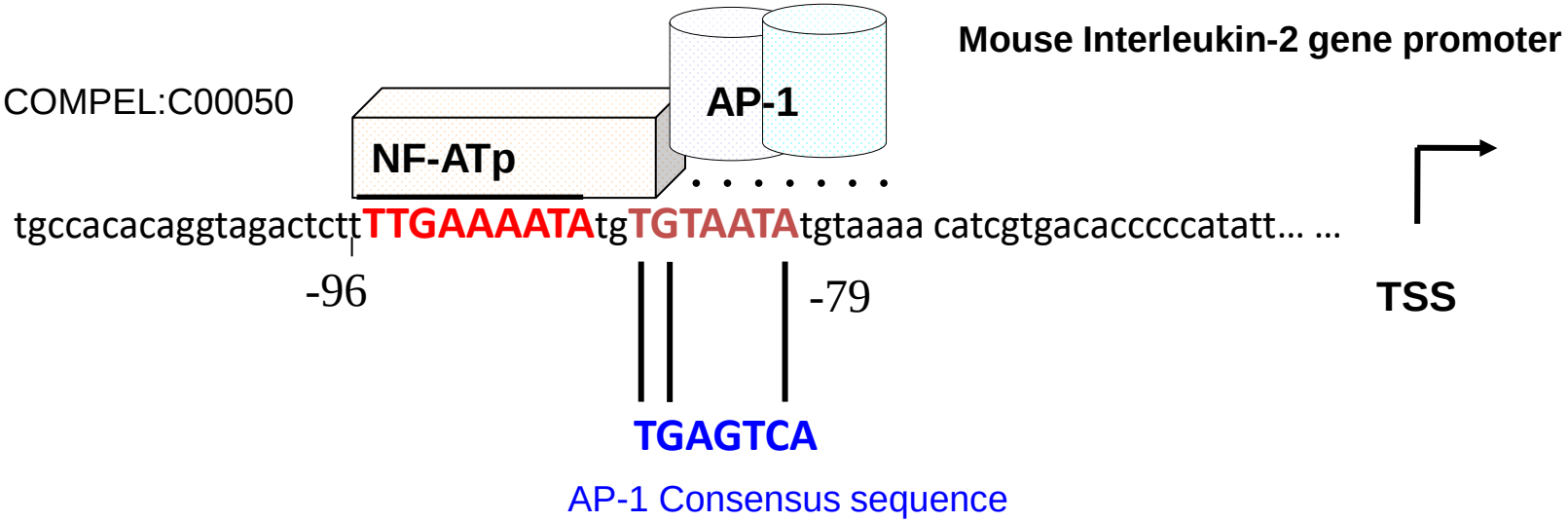
FIGURE 3.3. The human interferon-β enhanceosome. HMG represents HMGI/Y, a ubiquitous protein that binds cooperatively with the three activators. HMGI/Y both bends the DNA and contacts the activators. Each of the transcription factors shown is a member of a family of related activators.

(Mark Ptashne, Alexander Gann Genes and Signals, 2002)

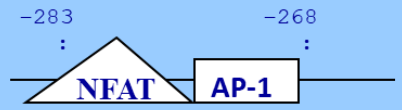
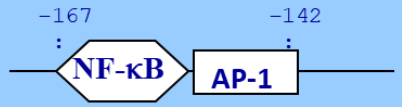
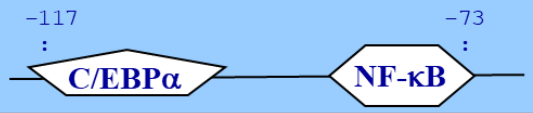
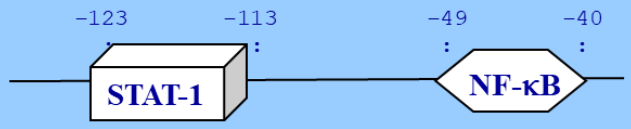
info@genexplain.com

Enhancers are site compositions = Composite Modules (CM)

- One of the TF binding sites in a composite module (CM) can be rather weak.
- Weak DNA-protein interactions are stabilized by protein-protein interactions.

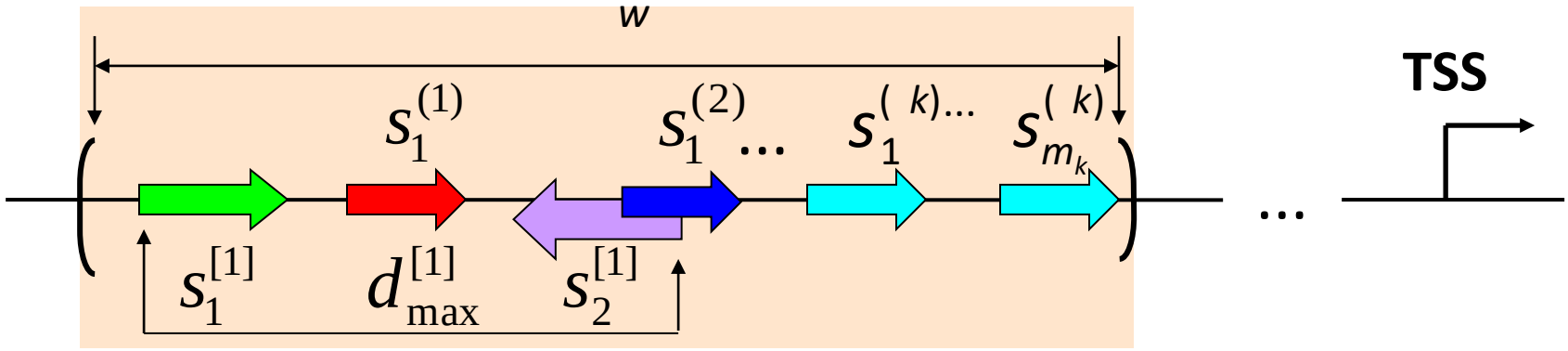


Examples of Composite Modules (CM)

N	Gene	Scheme of CM
1.	IgH , *Mus musculus*	
2.	IL-2, *Homo sapiens*	
3.	IL-2, *Homo sapiens*	
4.	Il-2, *Mus musculus*	
5.	IgH , *Homo sapiens*	
6.	Serum amyloid A1, *Rattus norv.*	
7.	IRF-1, *Mus musculus*	

Search for Composite Modules (CM)

Composite Modules Analysis (CMA):

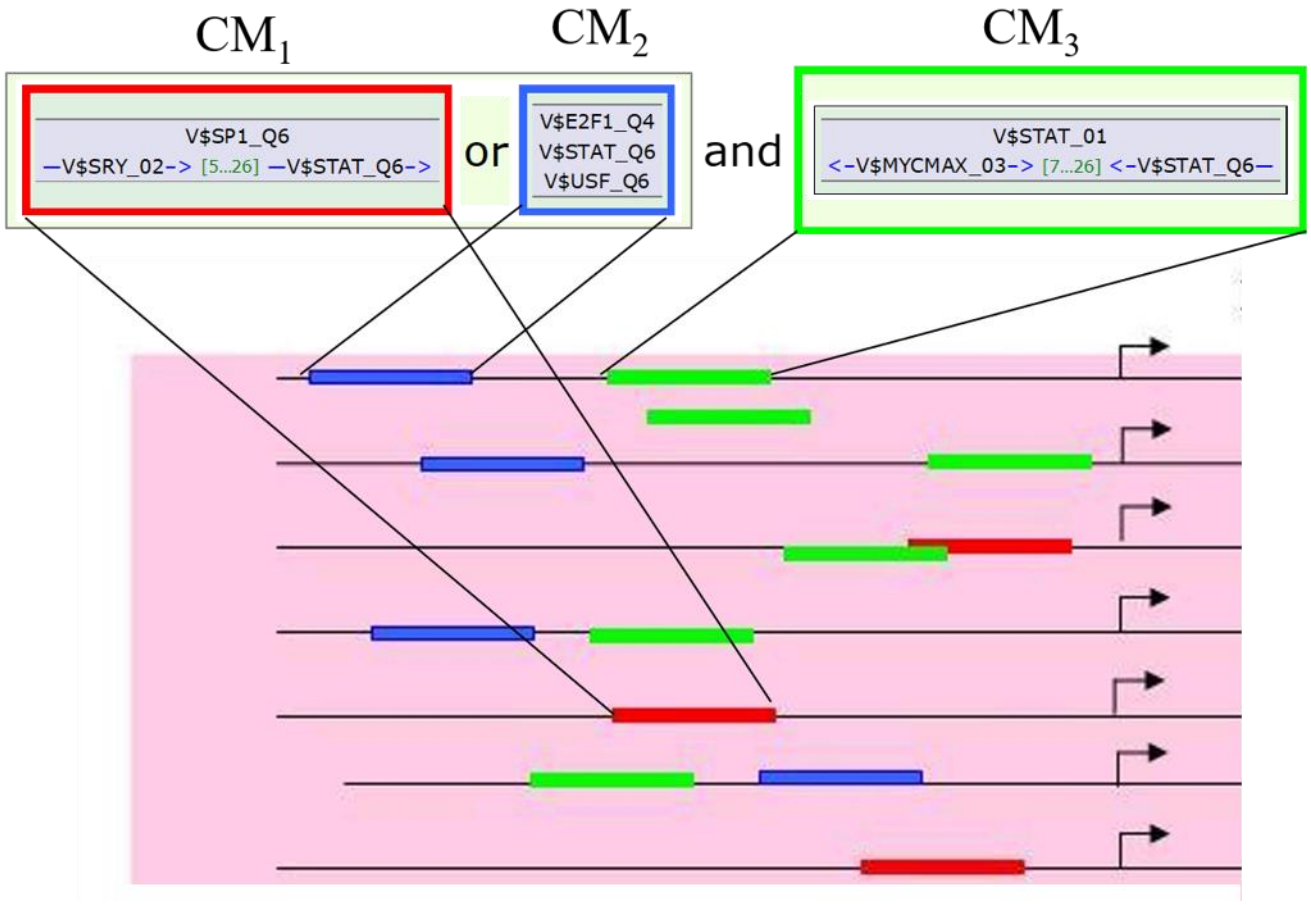


$$\begin{matrix} d_{\max}^{[1]} & d_{\max}^{[1]} & \dots & d_{\max}^{[R]} \\ q_{cut-off}^{(1)} & q_{cut-off}^{(2)} & \dots & q_{cut-off}^{(k)} \\ \phi^{(1)} & \phi^{(2)} & \dots & \phi^{(k)} \end{matrix}$$

Parameters of the model to be estimated by CMA

TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 34:D108-D110. PubMed.

Promoter model (example)



info@genexplain.com

Search for Composite Modules (CM)

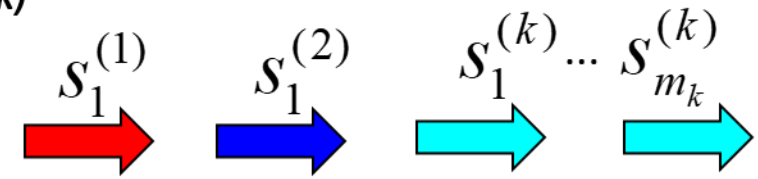
Composite Module Score (CMS):

K, the number of individual PWMs in the module, ***(k=1,K)***

Matrix cut-off values: $q_{cut-off}^{(k)}$

Relative impact values: $\phi^{(k)}$

Maximal number of best matches: m_k

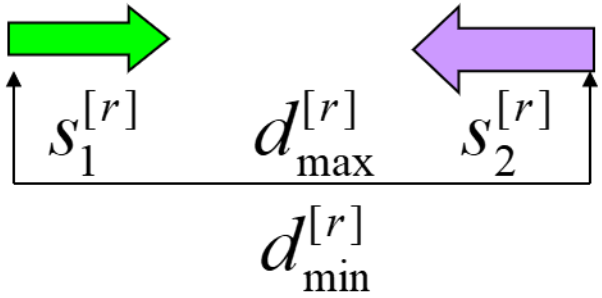


R, the number of pairs of PWMs ***(r=1,R)***

Matrix cut-off values: $q_{1,cut-off}^{[r]}$ $q_{2,cut-off}^{[r]}$

Relative impact values: $\phi^{[r]}$

Maximal and minimal distances: $d_{\max}^{[r]}$ $d_{\min}^{[r]}$

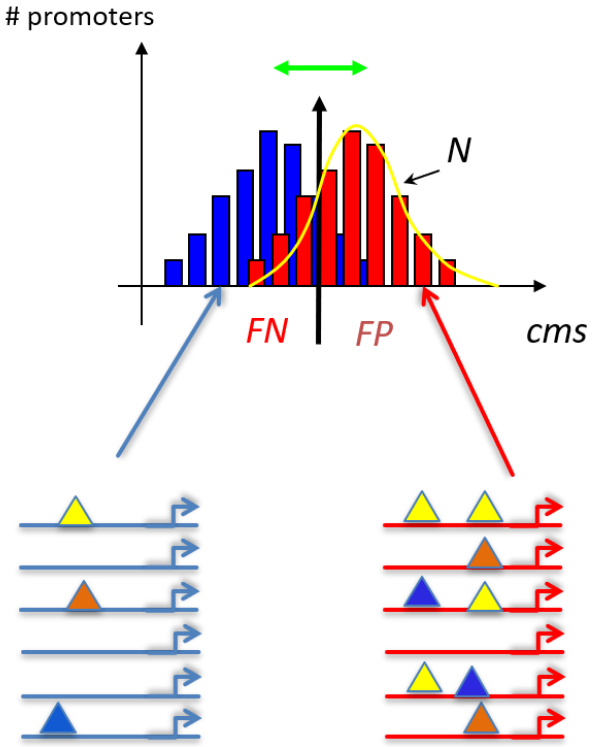


$$cms = \sum_{k=1,K} \phi^{(k)} \times \sum_{i=1}^{m_k} q_i^{(k)} + \sum_{r=1,R} \phi^{[r]} \times (q_1^{[r]} + q_2^{[r]}) / MAX$$

Search for Composite Modules (CM)

Fitness function of the Genetic-Regression Algorithm (GRA):

$$F = \alpha \cdot R + \beta \cdot (1 - FN) + (1 - \beta) \cdot (1 - FP) + \gamma \cdot T + \delta \cdot N - \mu \cdot k$$



R – linear regression

FN – false negatives

FP – false positives

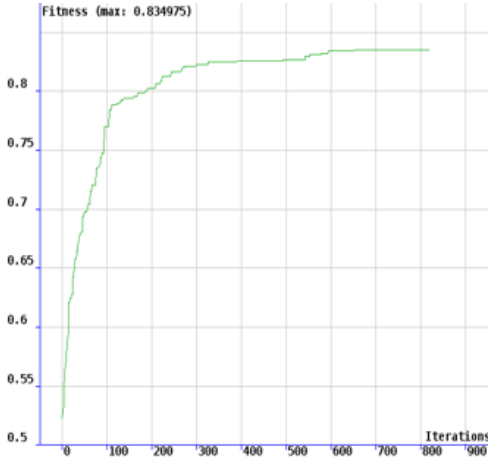
T – T-test (difference between mean values)

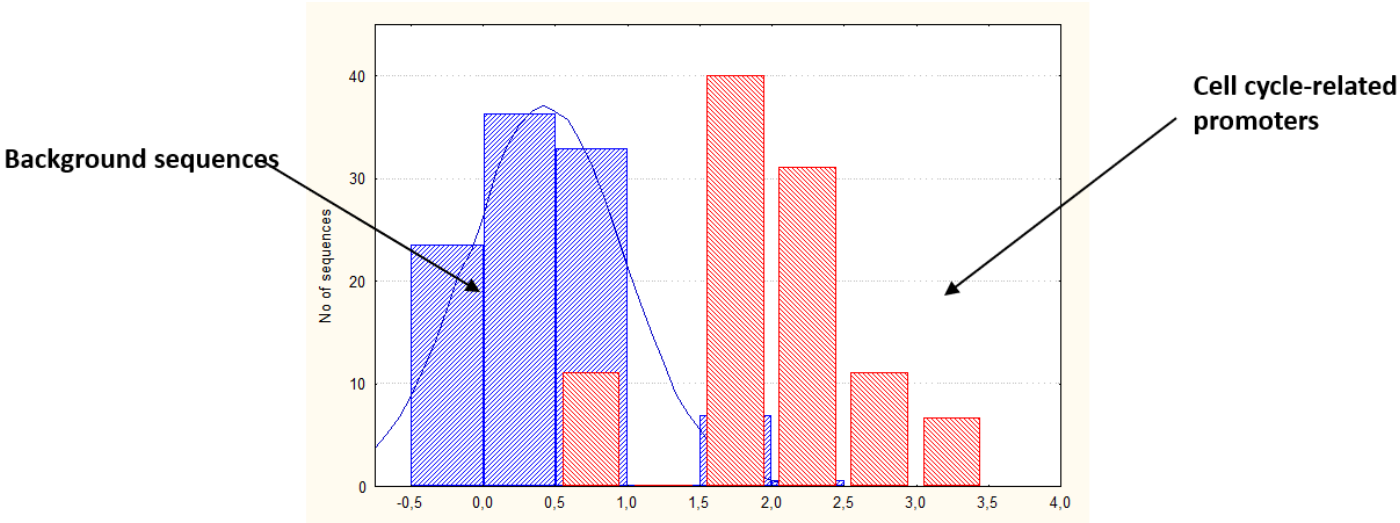
N – normal likeness

k – number of free parameters

Example of Composite Modules (CM) in promoters of cell cycle-related genes

Weight: ϕ	$q_{cut-off}$	TF matrix
1.000000	0.840072	V$E2F_19
0.954483	0.737637	V$TATA_01
0.888064	0.939687	V$CREB_01
0.816179	0.941583	V$SP1_Q6
0.039746	0.839702	V$TAL1BETAE47_01





Useful links

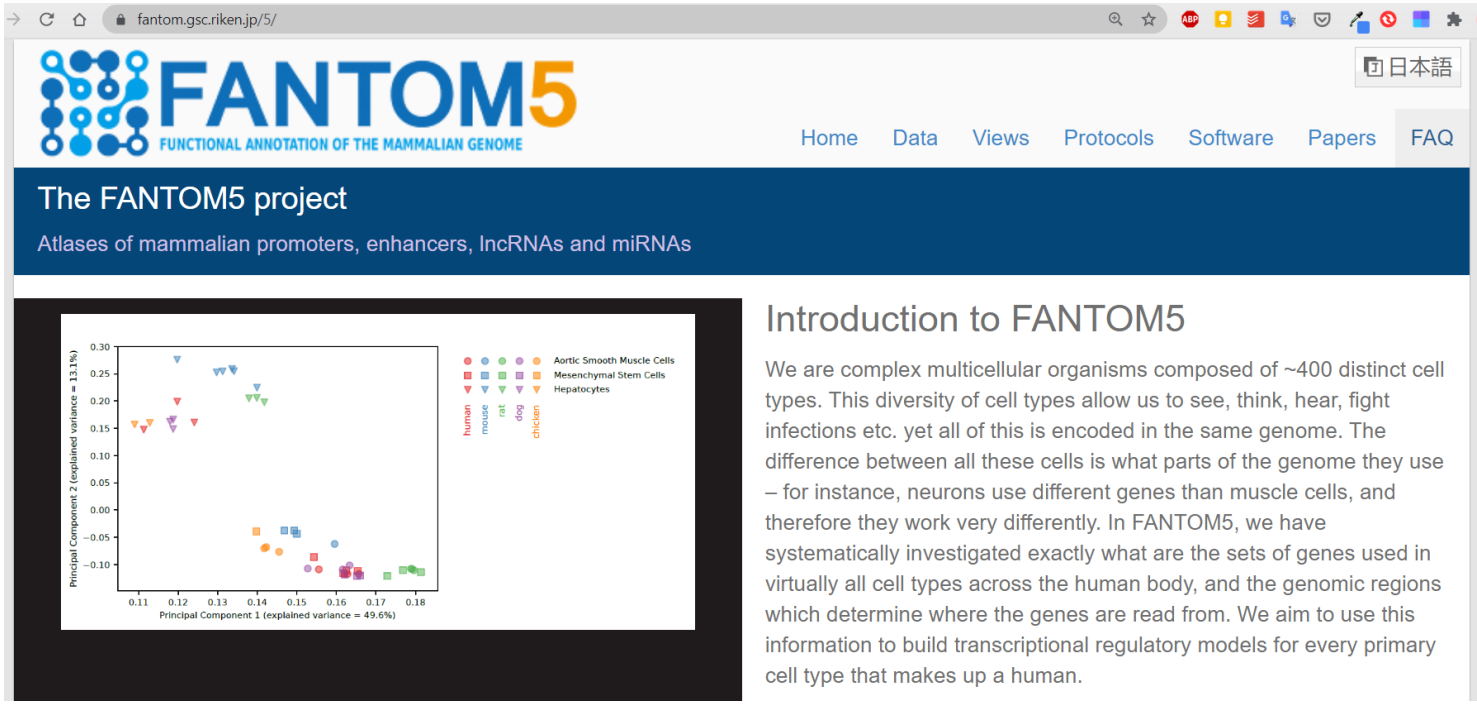
https://genexplain-platform-documentation.readthedocs.io/en/latest/Chip_seq.html?highlight=CMA#search-for-composite-modules-with-transfac

https://genexplain-platform-documentation.readthedocs.io/en/latest/Analysis_method_description.html?highlight=CMA#visualization-and-interpretation-of-the-results

info@genexplain.com

Fantom5 Database

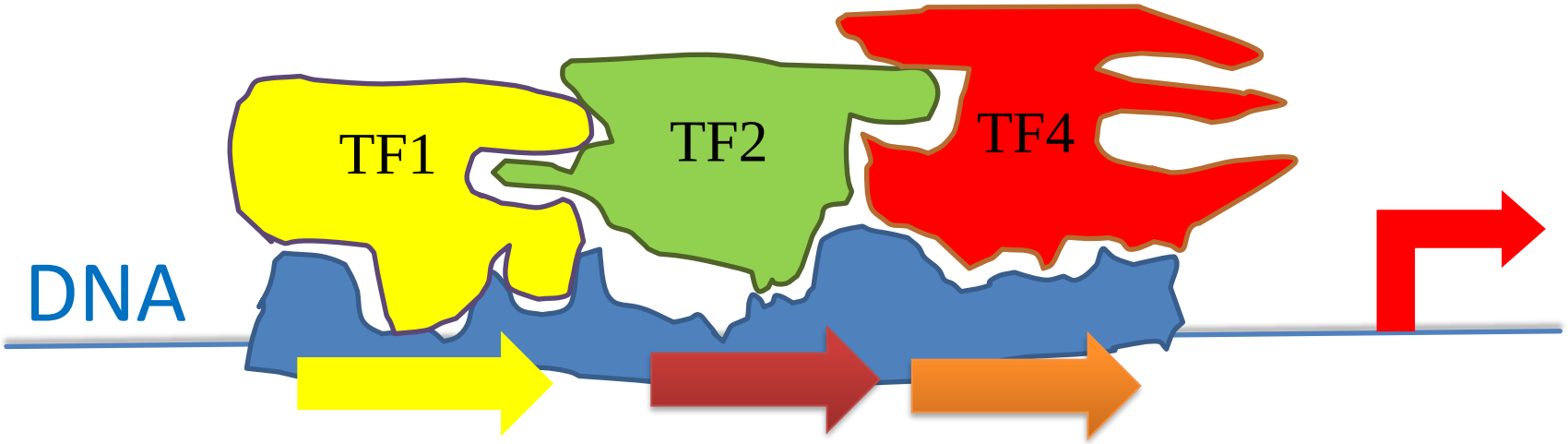
An atlas of active enhancers across human cell types and tissues



[paper link]

[software link]

Fuzzy puzzle

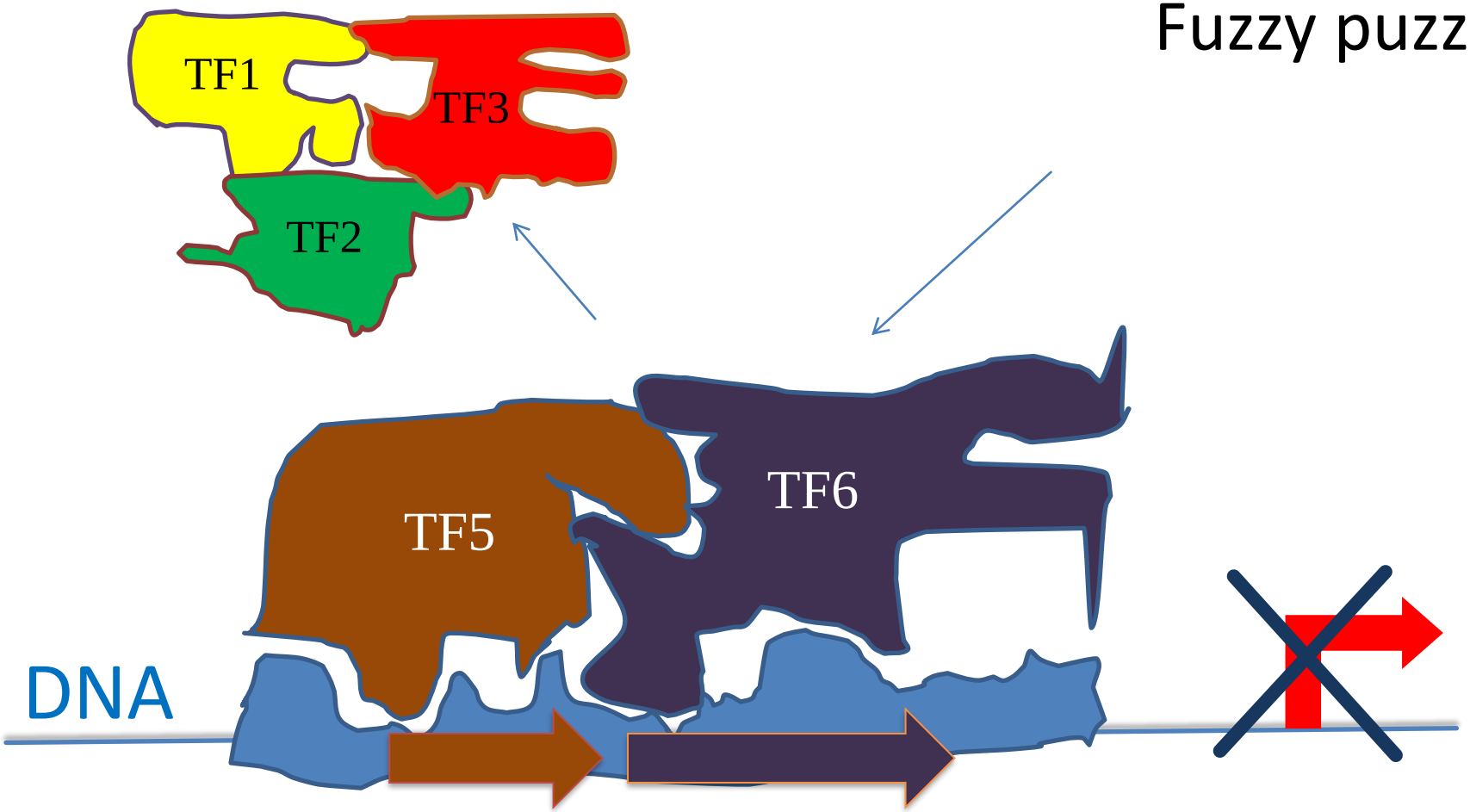


Activation of transcription

Cell type 1

info@genexplain.com

Fuzzy puzzle



Repression of transcription

Cell type 2

info@genexplain.com

Demo

- Part 1: Identifying regulation specific combinations of transcription factor binding sites with the help of Composite Module Analysis (CMA)
- Part 2: Useful tools for working with transcripts, transcript regions, transcript tracks
- Part 3: Using Fantom5 database to analyze tissue- or cell-type specific promoters
- MTB report

info@genexplain.com