

Enhancers

Day 4

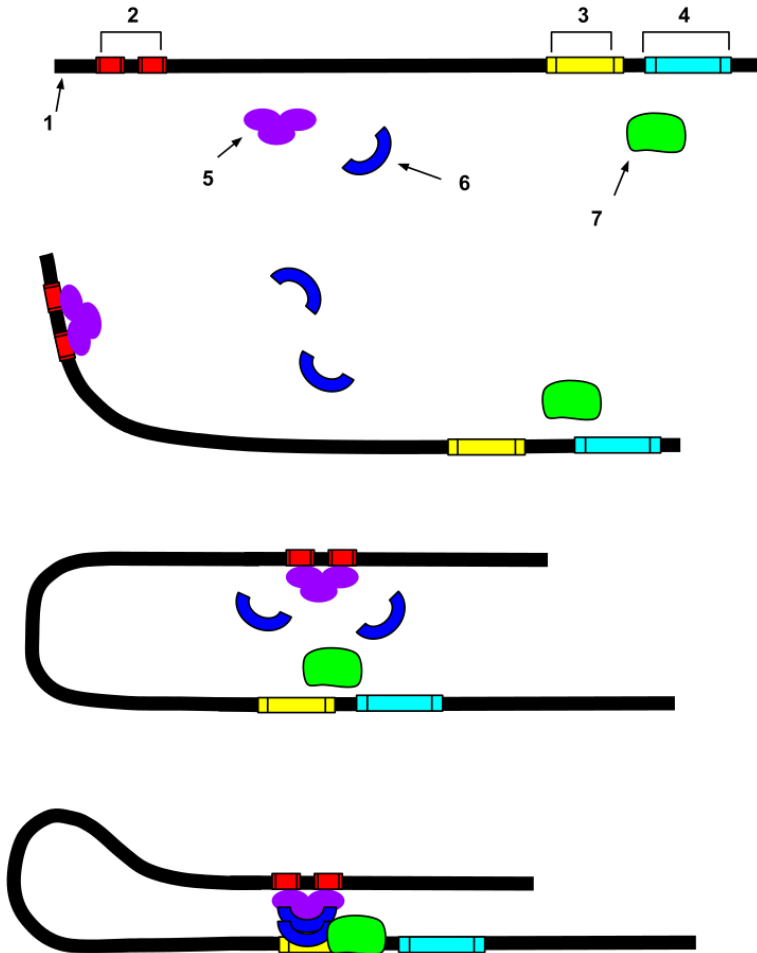
Enhancer

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.

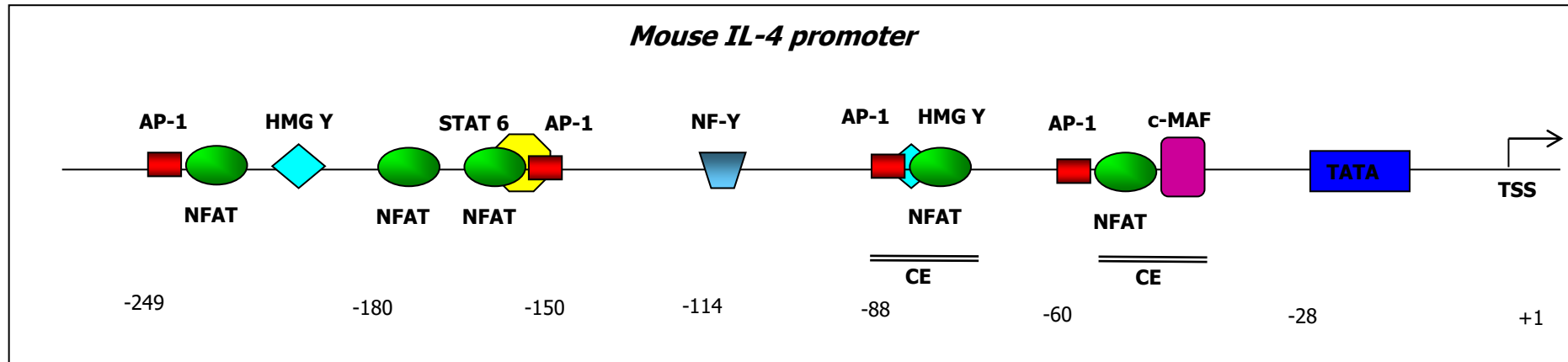
Usage of an Enhancer



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.

Promoter and enhancer: Mouse IL-4 as example



Promoter and enhancer: Human Interferon- β as example

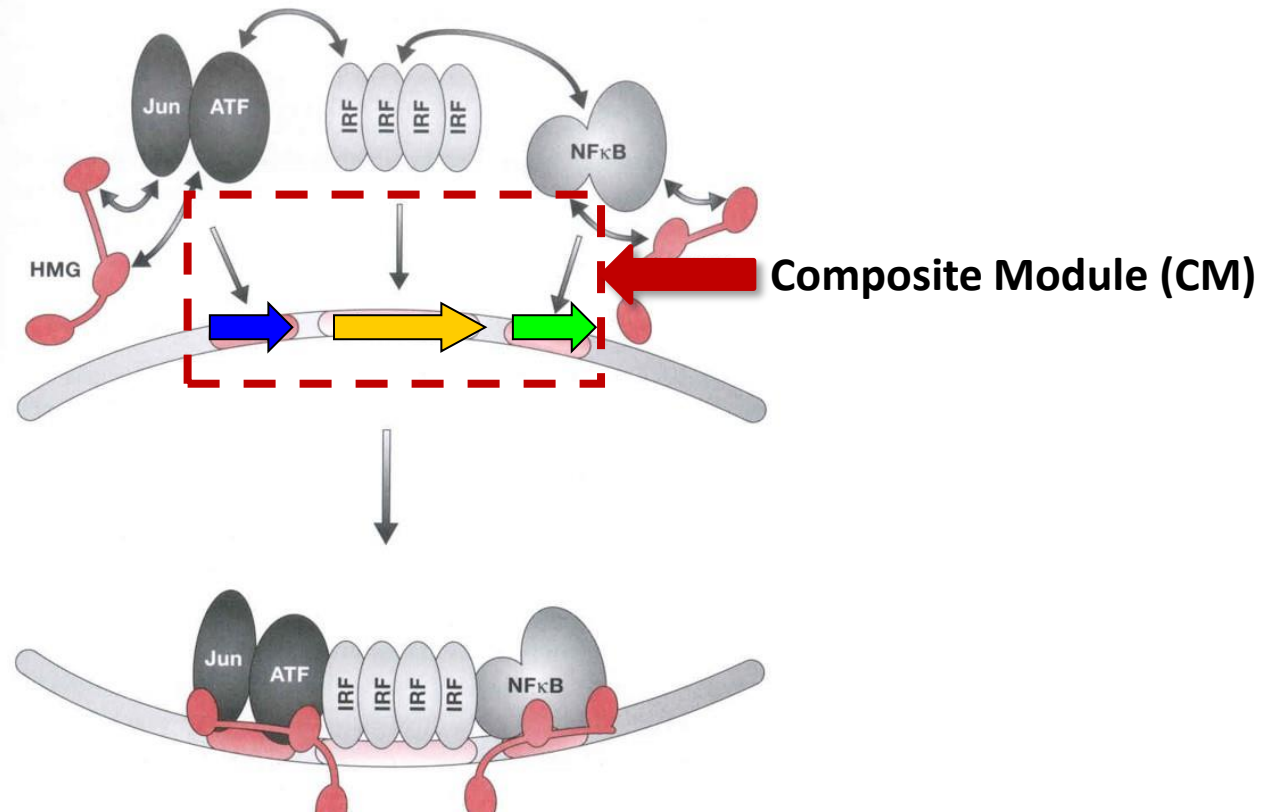
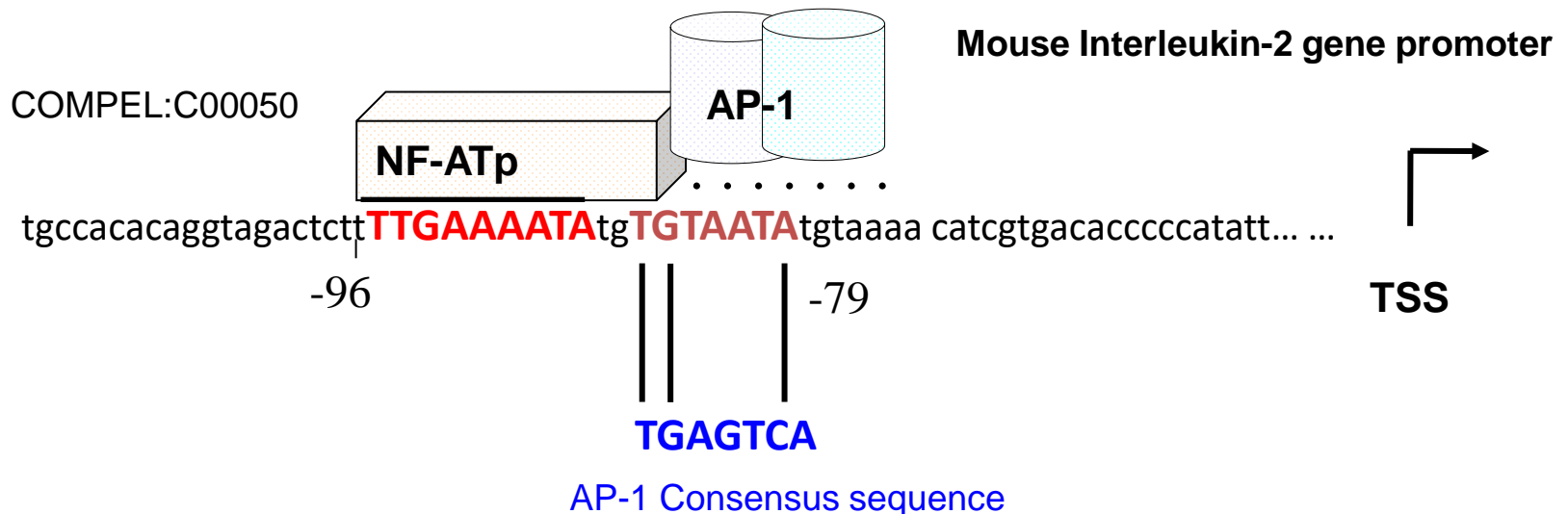


FIGURE 3.3. The human interferon- β enhanceosome. HMG represents HMGI/Y, a ubiquitous protein that binds cooperatively with the three activators. HMGI/Y both bends the DNA and contacts the activators. Each of the transcription factors shown is a member of a family of related activators.


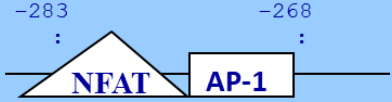
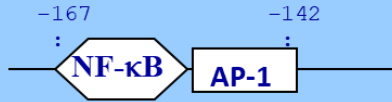
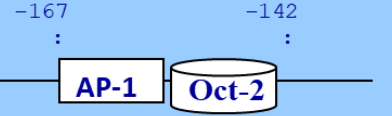

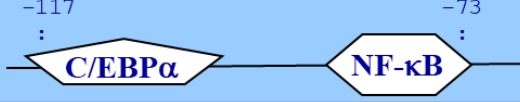
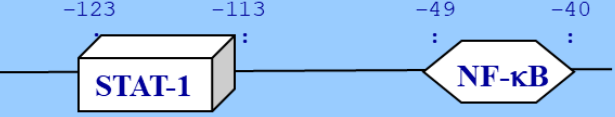
(Mark Ptashne, Alexander Gann [Genes and Signals](#), 2002)

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.

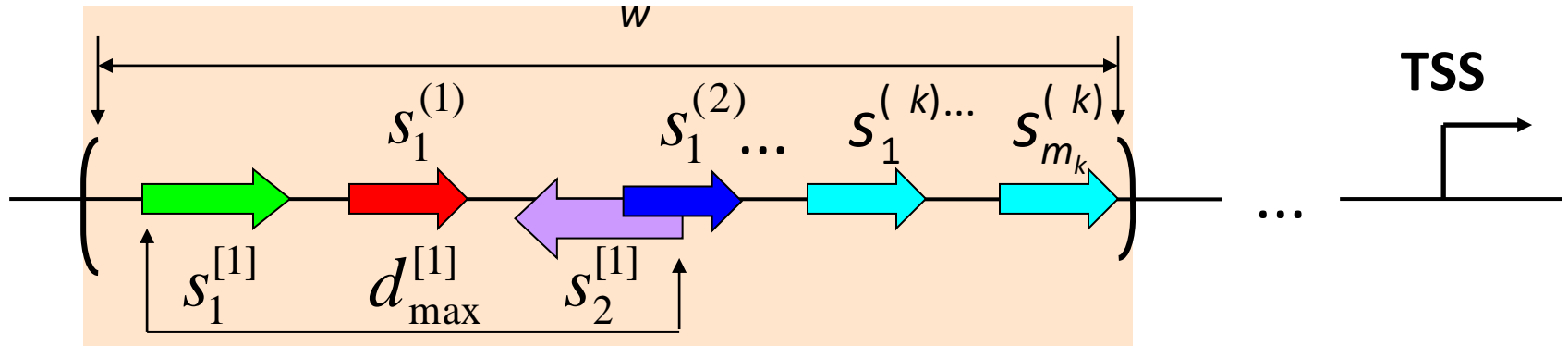


Examples of Composite Modules (CM)

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

Search for Composite Modules (CM)

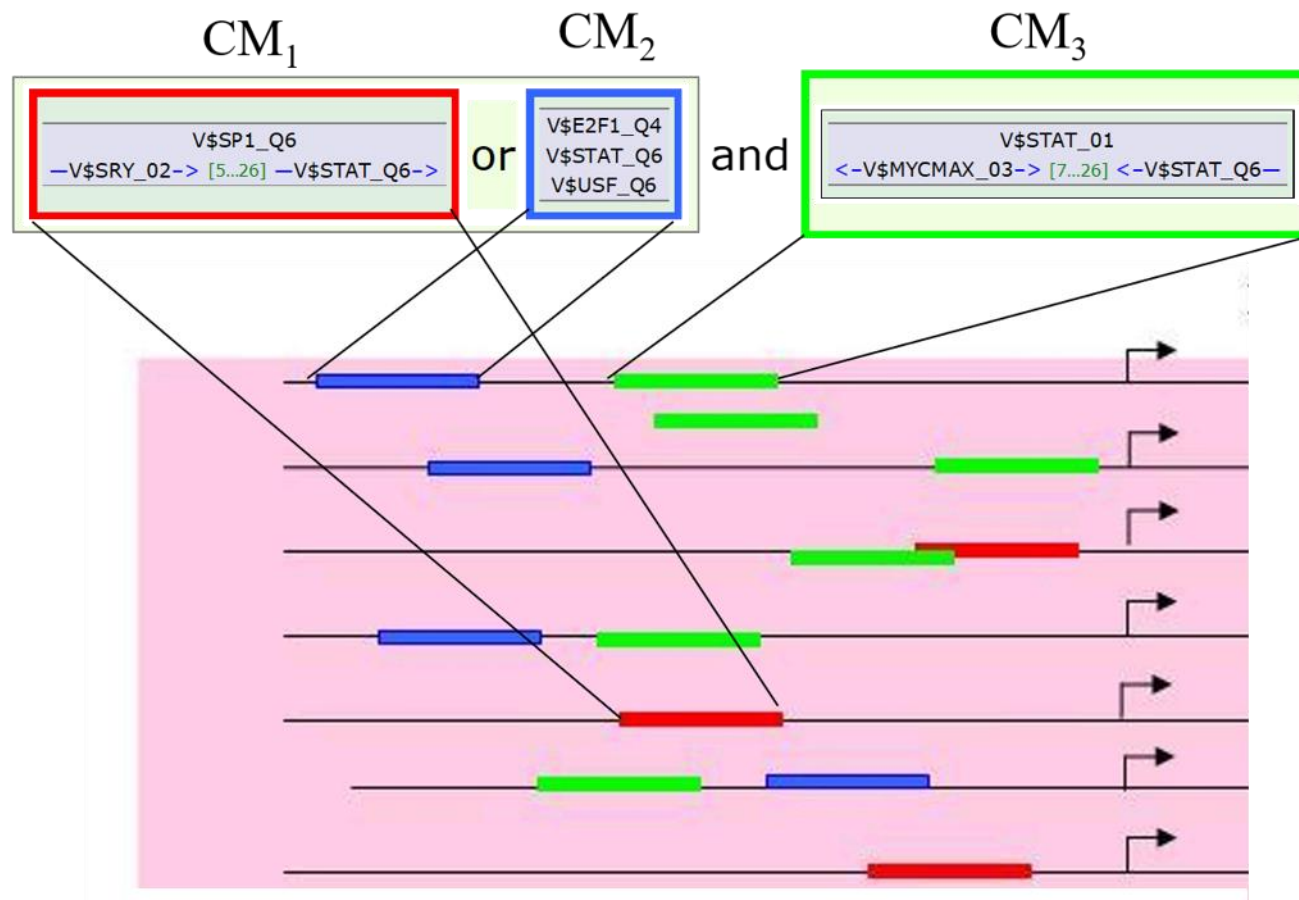
Composite Modules Analysis (CMA):



$d_{\max}^{[1]}$	$d_{\max}^{[1]}$...	$d_{\max}^{[R]}$	} Parameters of the model to be estimated by CMA
$q_{cut-off}^{(1)}$	$q_{cut-off}^{(2)}$...	$q_{cut-off}^{(k)}$	
$\phi^{(1)}$	$\phi^{(2)}$...	$\phi^{(k)}$	

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

Promoter model (example)



Search for Composite Modules (CM)

Composite Module Score (CMS):

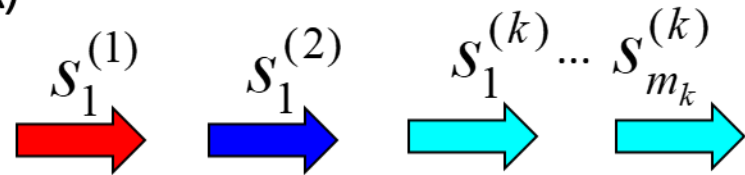
K, the number of individual PWMs in the module, ($k=1, K$)

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

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

Maximal number of best matches:

m_k

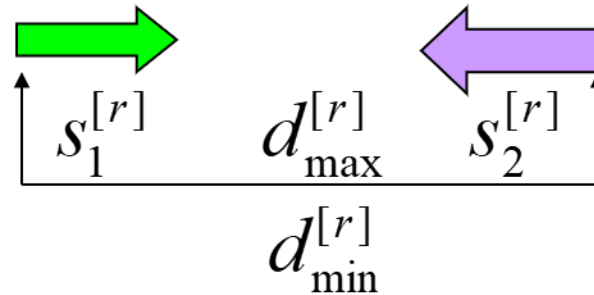


R, the number of pairs of PWMs ($r=1, R$)

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

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

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

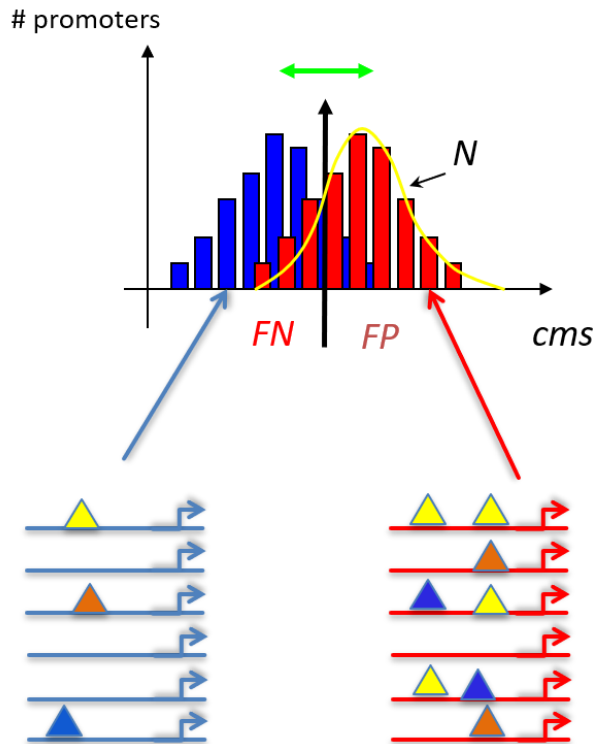


$$cms = \sum_{k=1, K} \phi^{(k)} \times \sum_{i=1}^{m_k} q_i^{(k)} + \sum_{r=1, R} \phi^{[r]} \times (q_1^{[r]} + q_2^{[r]}) / MAX$$

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

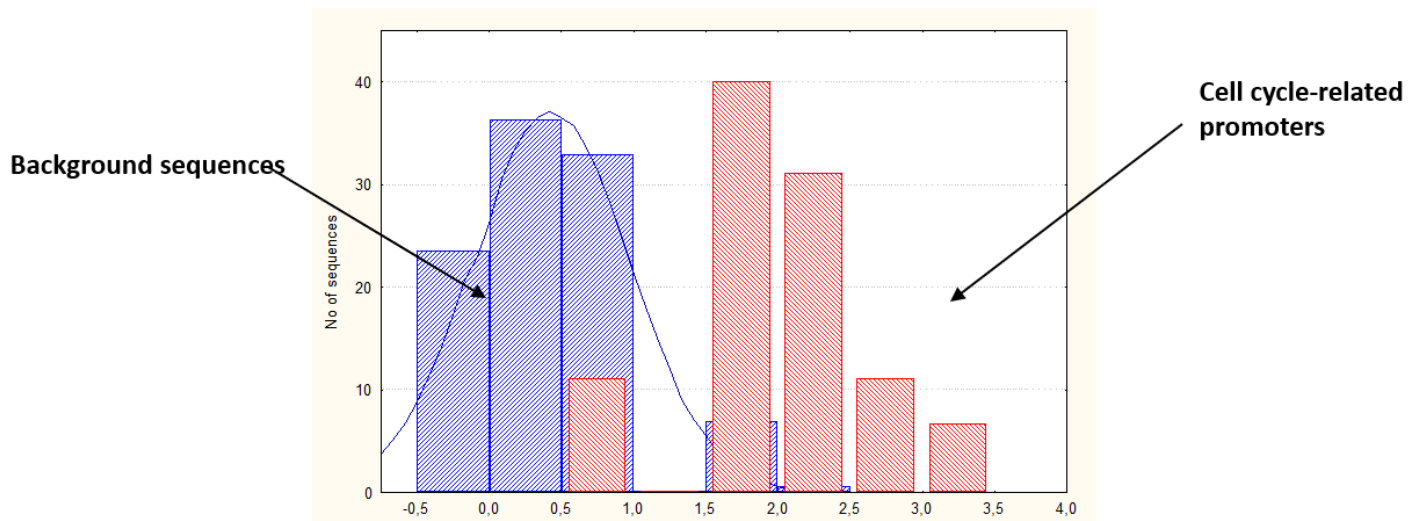
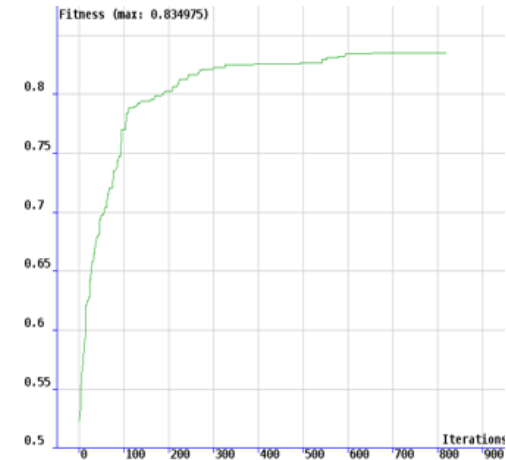
T – T-test (difference between mean values)

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



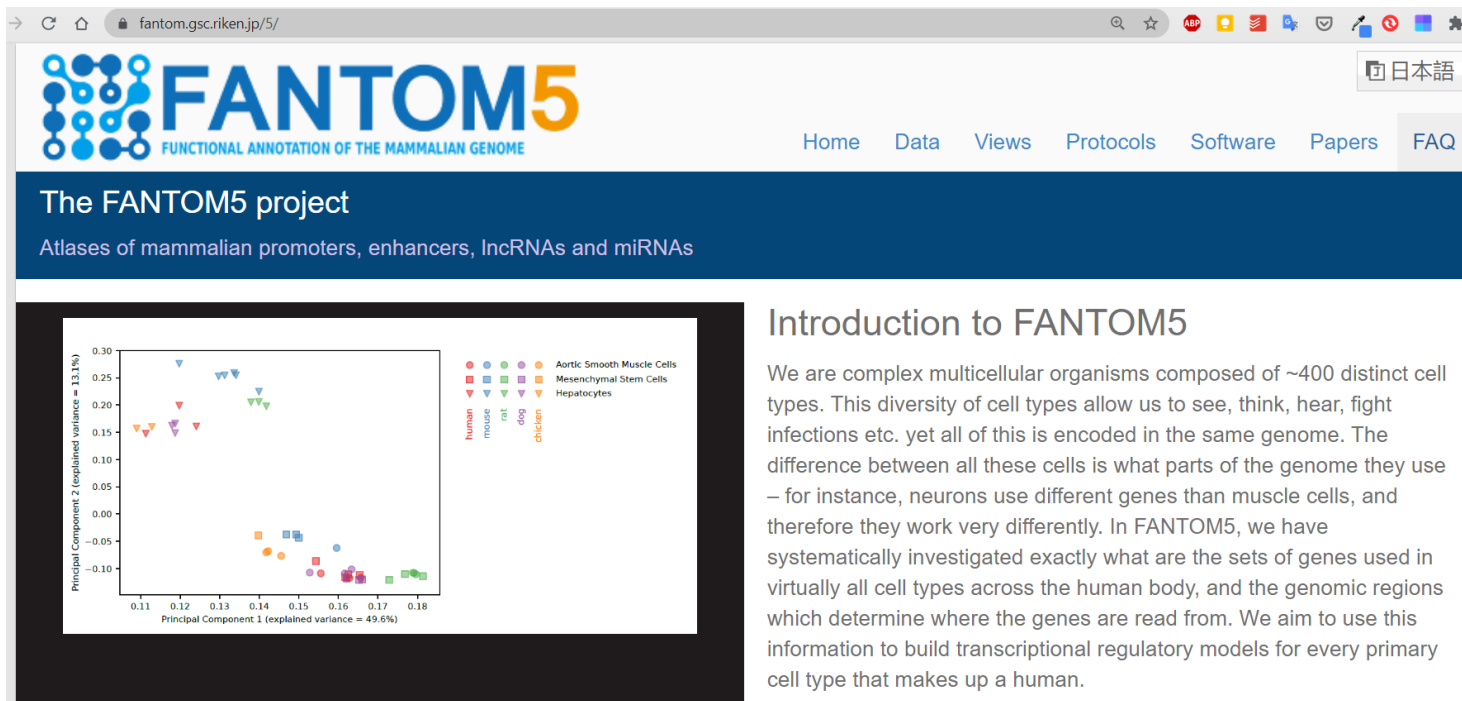
Useful links

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

Fantom5 Database

An atlas of active enhancers across human cell types and tissues



The screenshot shows the FANTOM5 website interface. At the top, there is a navigation menu with links for Home, Data, Views, Protocols, Software, Papers, and FAQ. Below the navigation is a dark blue banner with the text "The FANTOM5 project" and "Atlases of mammalian promoters, enhancers, lncRNAs and miRNAs". The main content area is divided into two columns. The left column features a PCA plot titled "Introduction to FANTOM5". The plot shows Principal Component 1 (explained variance = 49.6%) on the x-axis and Principal Component 2 (explained variance = 13.1%) on the y-axis. Data points are color-coded by species: human (red), mouse (blue), rat (green), dog (purple), and chicken (orange). The legend also lists cell types: Aortic Smooth Muscle Cells (red circle), Mesenchymal Stem Cells (blue square), and Hepatocytes (green triangle). The right column contains the text "Introduction to FANTOM5" and a paragraph describing the project's goal to investigate gene sets used in various cell types across the human body.

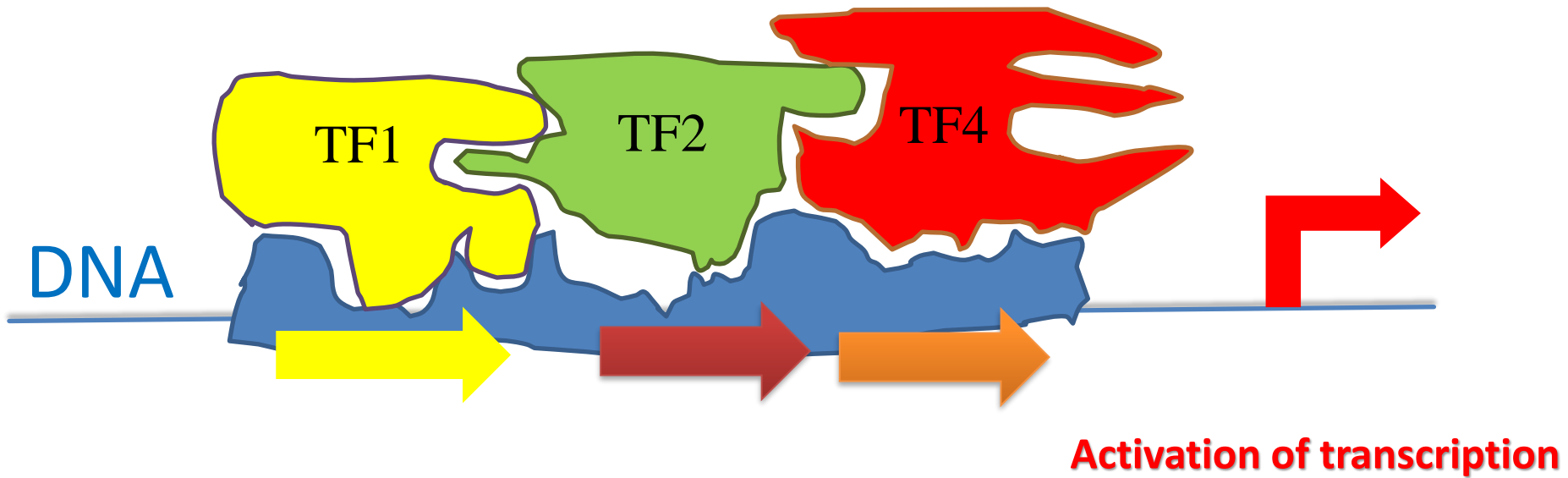
Introduction to FANTOM5

We are complex multicellular organisms composed of ~400 distinct cell types. This diversity of cell types allows us to see, think, hear, fight infections etc. yet all of this is encoded in the same genome. The difference between all these cells is what parts of the genome they use – for instance, neurons use different genes than muscle cells, and therefore they work very differently. In FANTOM5, we have systematically investigated exactly what are the sets of genes used in virtually all cell types across the human body, and the genomic regions which determine where the genes are read from. We aim to use this information to build transcriptional regulatory models for every primary cell type that makes up a human.

[paper link](#)

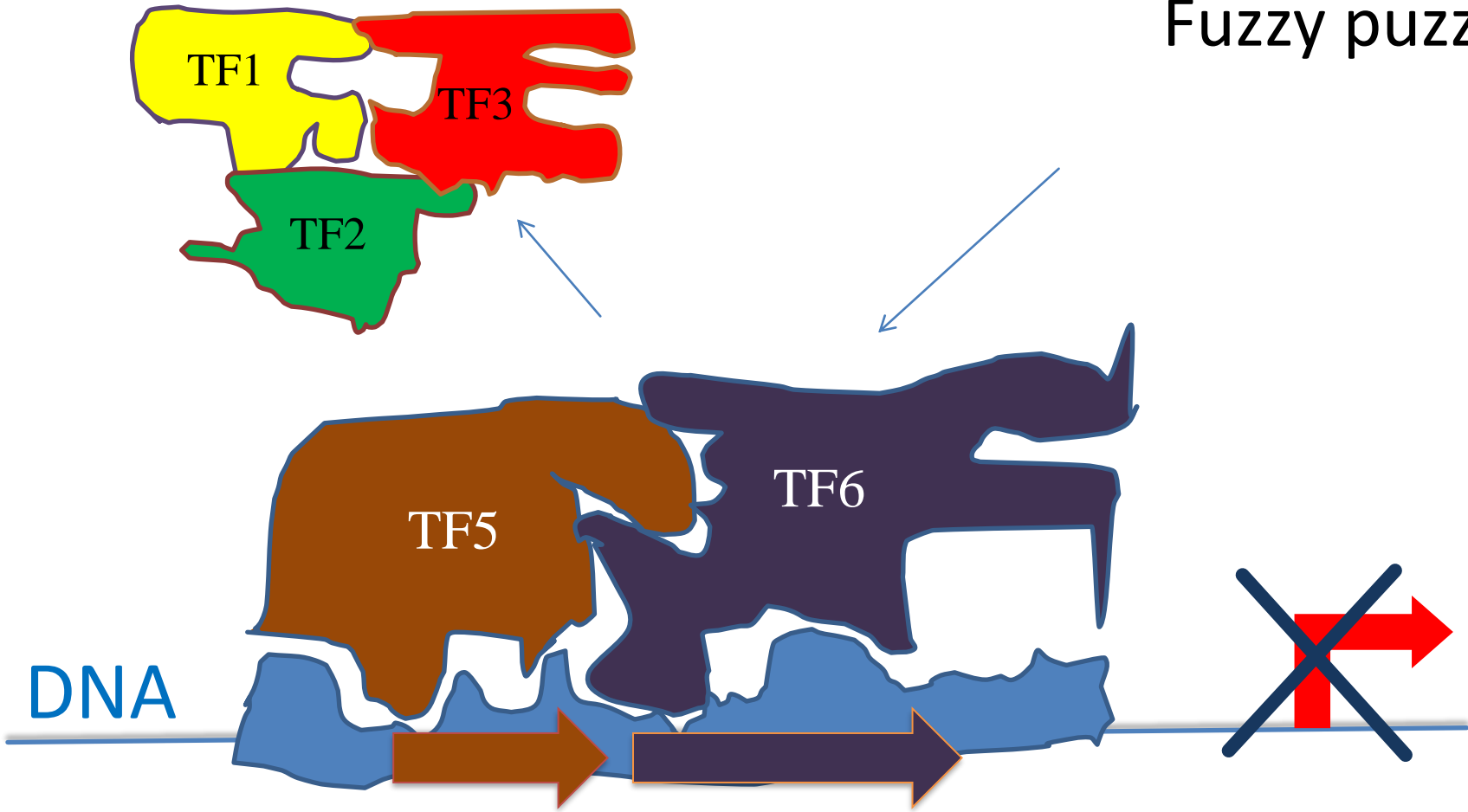
[software link](#)

Fuzzy puzzle



Cell type 1

Fuzzy puzzle



Repression of transcription

Cell type 2

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