CH 3 - L2 part1

Transcription Factors

Enhancers

Enhancer structure

Individual TF Binding Sites are short motifs: 4-15 bp

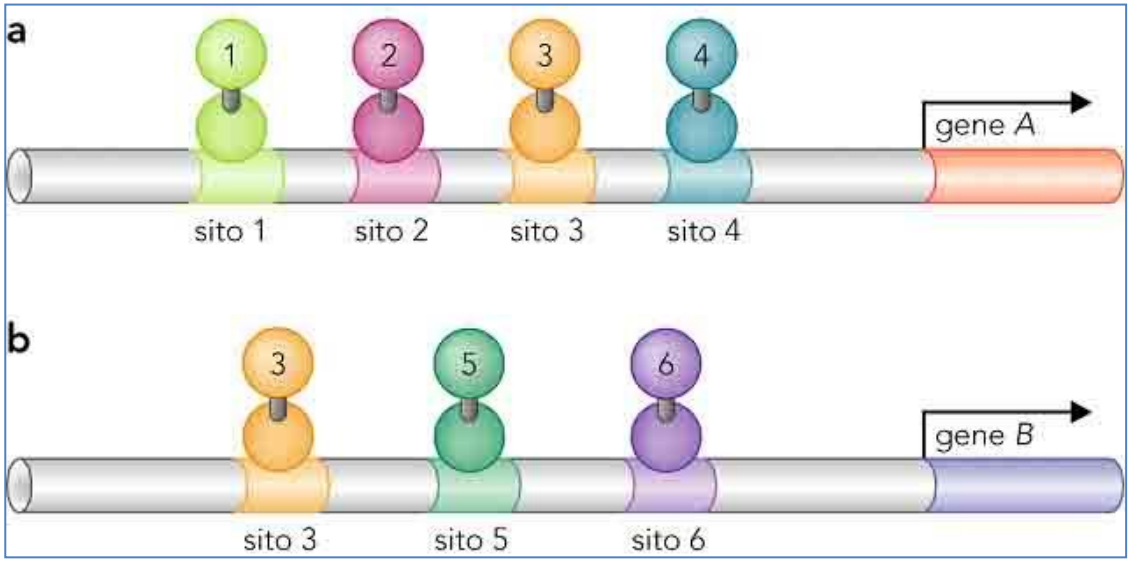
How can «small» cis-elements guide the **specificity** of DNA binding ?

First, almost all TFs bind DNA either as dimers or trimers or higher, and with more complex patterns (e.g. heterodimers of the same family)

Second, almost all the time TFBS are not isolated but found in *clusters*

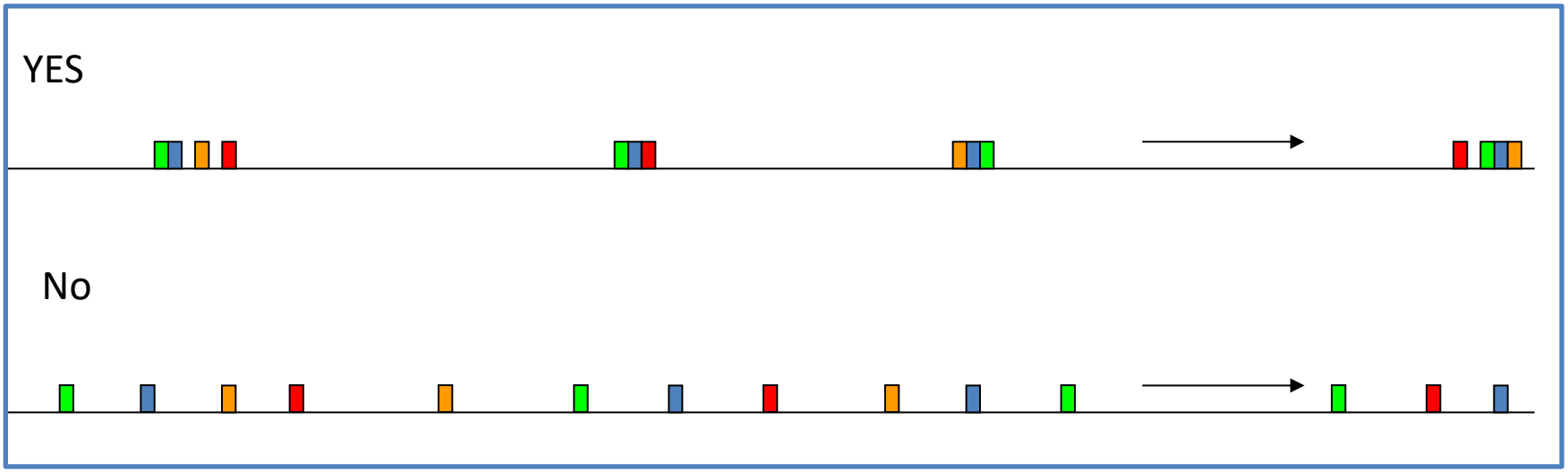
We will see in the following that TF function follow the co-co-co- rule:

- ✓ Combinatorial binding integrates multiple regulation
- ✓ Compositional binding increases fine-tuning
- ✓ Cooperativity will determine transcriptional outcome



Watson textbook

Enhancer structure

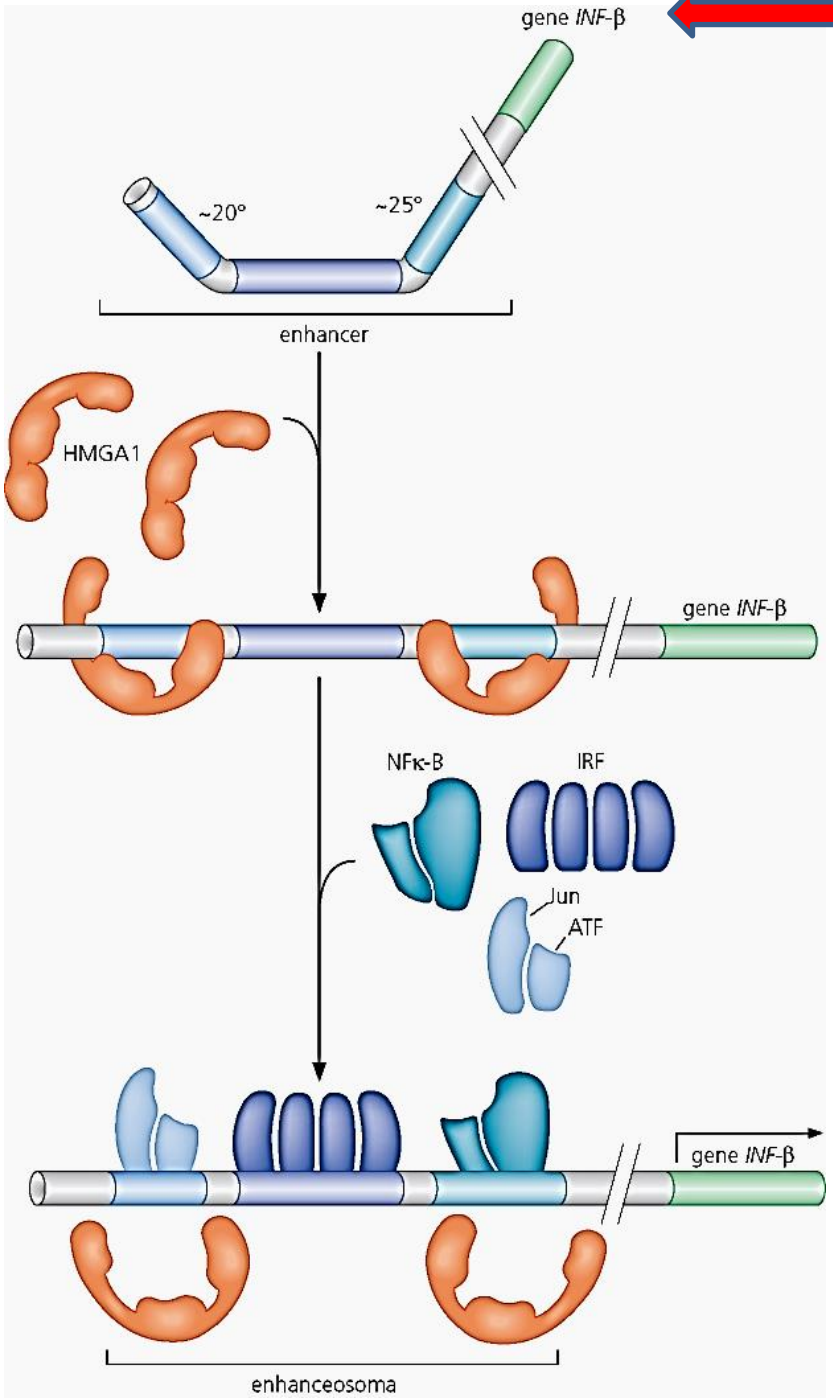


Compositionality

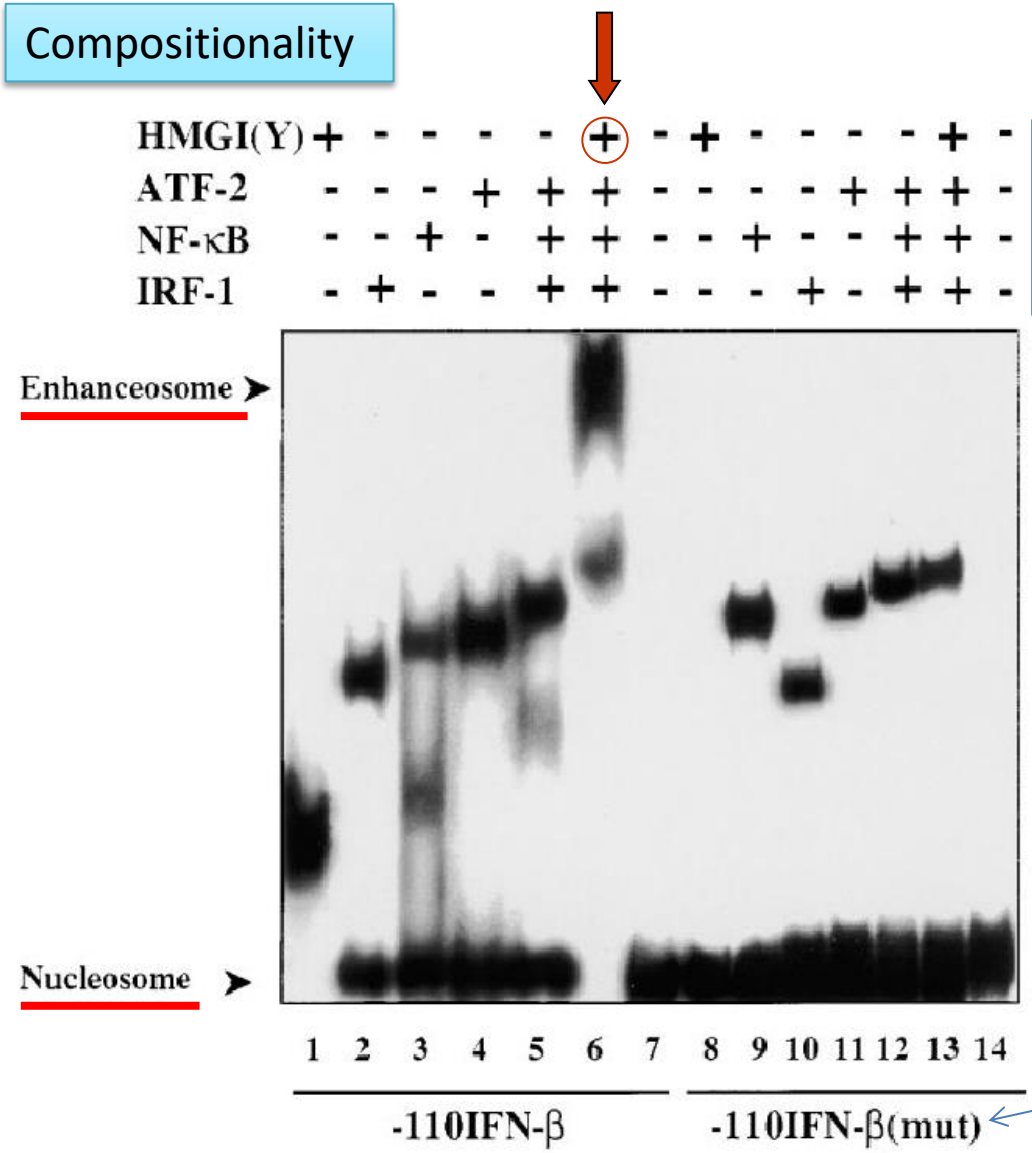
TFs binding may be favoured by the local 3D conformation

Old example

the INF-β enhancer:



Compositionality



How it is done:
Enhancer DNA with nucleosome is added of recombinant proteins.

Band-shift (**EMSA**) analysis shows binding of ATF-2, NF-kB and IRF-1 to the enhancer.

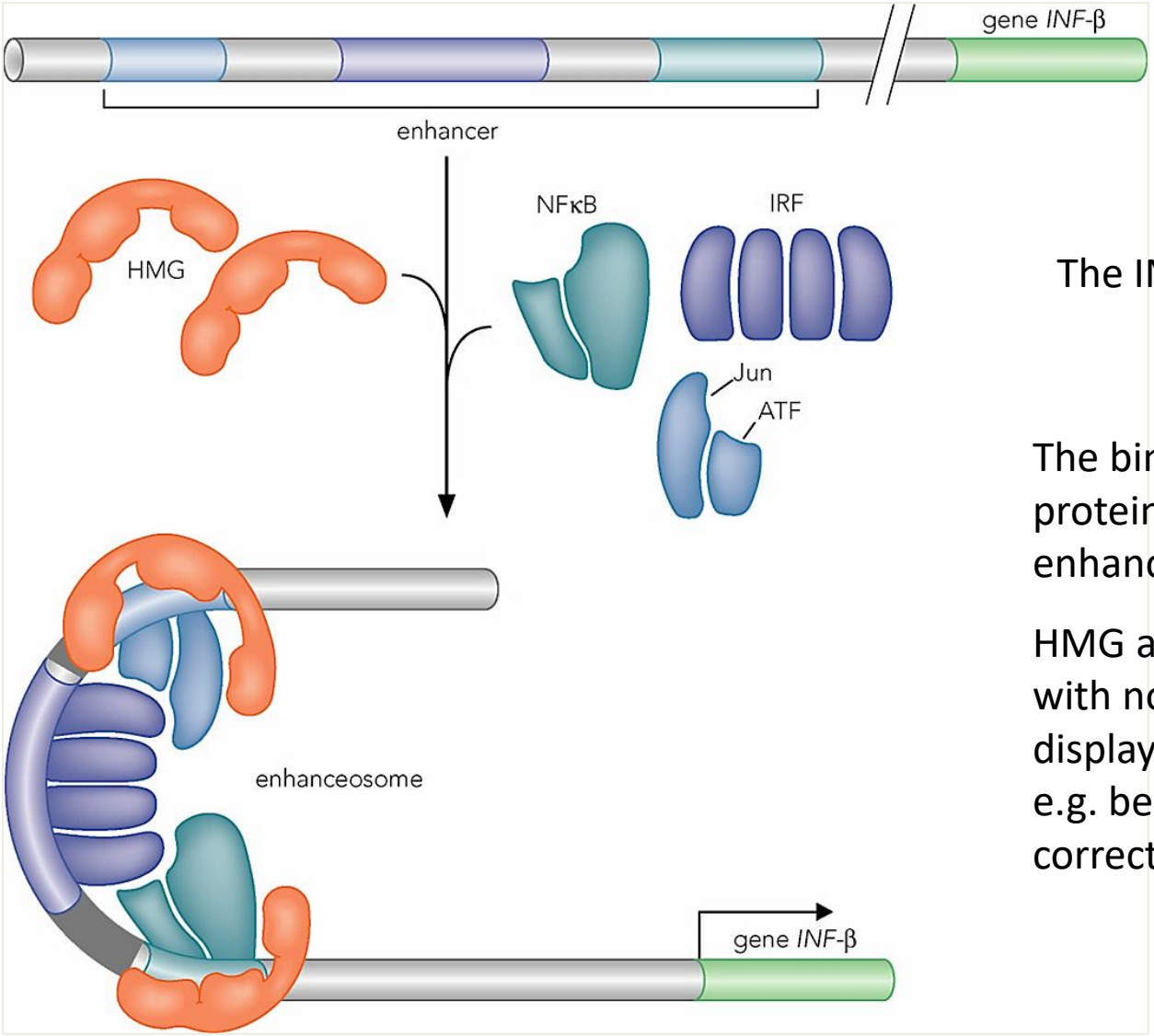
The enhanceosome is formed only in the presence of the architectural protein HMGI(Y) to 4 specific sites.

from: Agalioti et al. (2000) Cell 103: 667-678.

Mutated in HMGI(Y) site

(D) Assembly of the IFN-β enhanceosome on nucleosomal IFN-β promoter fragments. An IFN-β promoter fragment (−143 to +183) (lanes 1–7) or an identical-sized fragment bearing mutations in all HMG I(Y) binding sites (lanes 8–14) were reconstituted into a nucleosome, gel purified, and used in EMSA experiments along with recombinant IFN-β activators in the presence or in the absence of HMG I(Y). The following amounts of recombinant proteins were used: HMG I(Y) 10 ng, IRF-1 30 ng, NF-κB 20 ng, ATF-2/c-Jun 50 ng. The bottom part of the Figure depicts a diagrammatic illustration of the enhanceosome bound to the IFN-β nucleosomal promoter fragment.

Compositionality



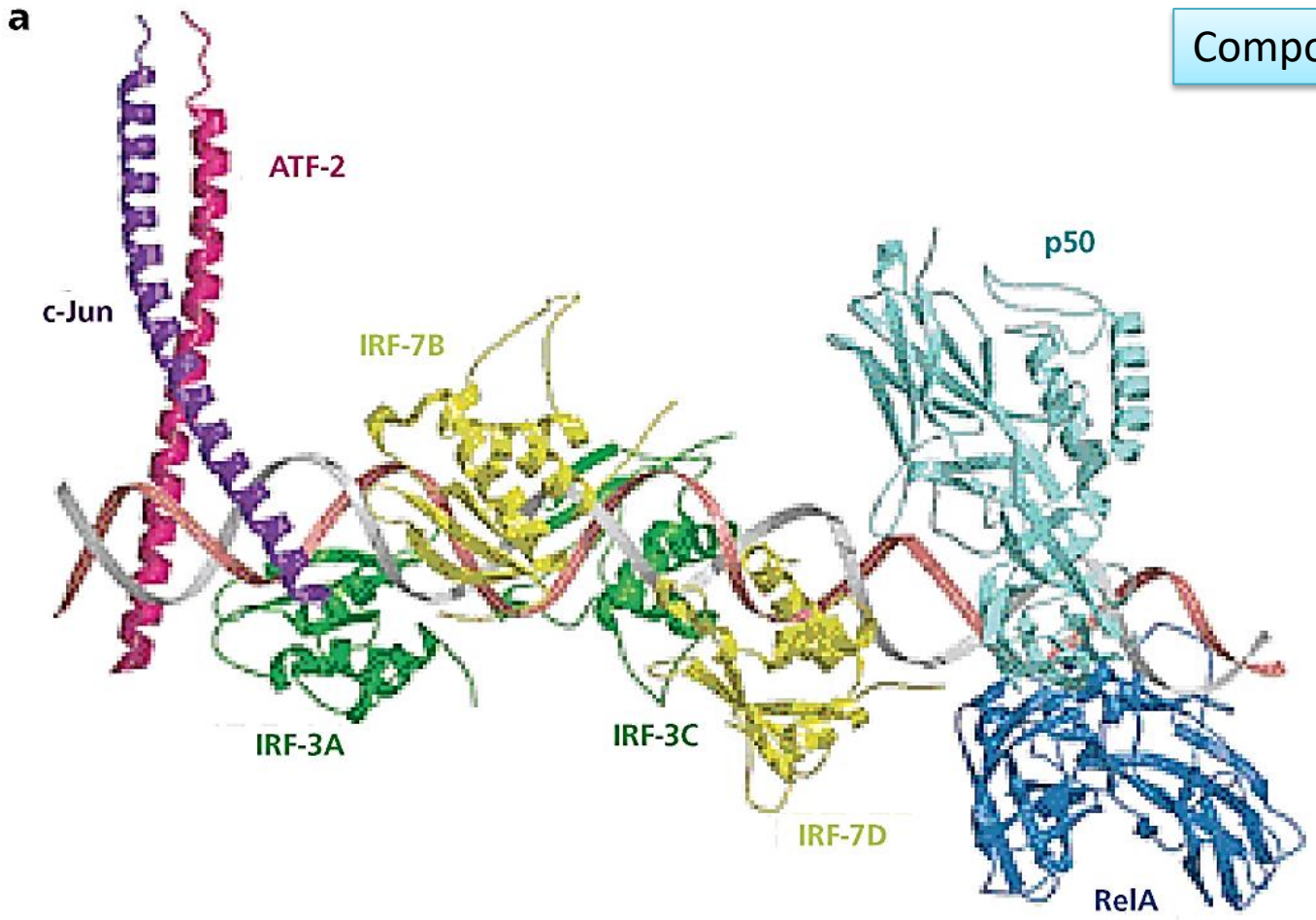
The INF-β "enhanceosome"

The binding of multiple different proteins to adjacent sites in enhancers is required.

HMG are DNA-binding proteins with no transactivating domain, but displaying "architectural" functions, e.g. bending the DNA and allowing correct interaction among TFs.

This old example illustrates «compositionality»

Compositionality



b

	ATF	Jun	IRF	IRF	IRF	IRF	NF-κB	
Uomo	1:	AAATGTAAATGACATAGGAAAACTGAAAGGGAGAAGTGAAAGTGGGAAATTCCTCTGAAT						:60
Topo	1:	......AAATGACAGAGGAAAACTGAAAGGGAGAACTGAAAGTGGGAAATTCCTCTGA..						:52
Ratto	1:	......AAATGACGGAGGAAAAGTGAAAGGGAGAACTGAAAGTGGGAAATTCCTCTGA..						:52
Suino	1:	......AAATGACATAGGAAAACTGAAAGGGAGAACTGAAAGTGGGAAATTCCTCTGAA.						:53
Cavallo	1:	.AATGTAAATGACATAGGAAAACAGAAAGGGAGAACTGAAAGTGGGAAATTCCTCTGAA.						:58
Bovino2	1:	.....TAAATGACAAAGGAAAACTGAAAGGGAGAACTGAAAGTGGGAAATCTCTCC....						:45
Bovino	1:	.....TAAATGACATGGGAAAAATGAAAGCGAGAACTGAAAGTGGGAAATTCCTCT....						:51

100 milioni di anni

Combinatorial control of gene expression

Attila Reményi[1,2,4], Hans R Schöler[1,3] & Matthias Wilmanns[2]

Revealing the molecular principles of eukaryotic transcription factor assembly on specific DNA sites is pivotal to understanding how genes are differentially expressed. By analyzing structures of transcription factor complexes bound to specific DNA elements we demonstrate how protein and DNA regulators manage gene expression in a combinatorial fashion.

TF-TF interaction may be mediated by DNA (adjacent elements) or by simple protein-protein contacts

812 VOLUME 11 NUMBER 9 SEPTEMBER 2004 NATURE STRUCTURAL & MOLECULAR BIOLOGY

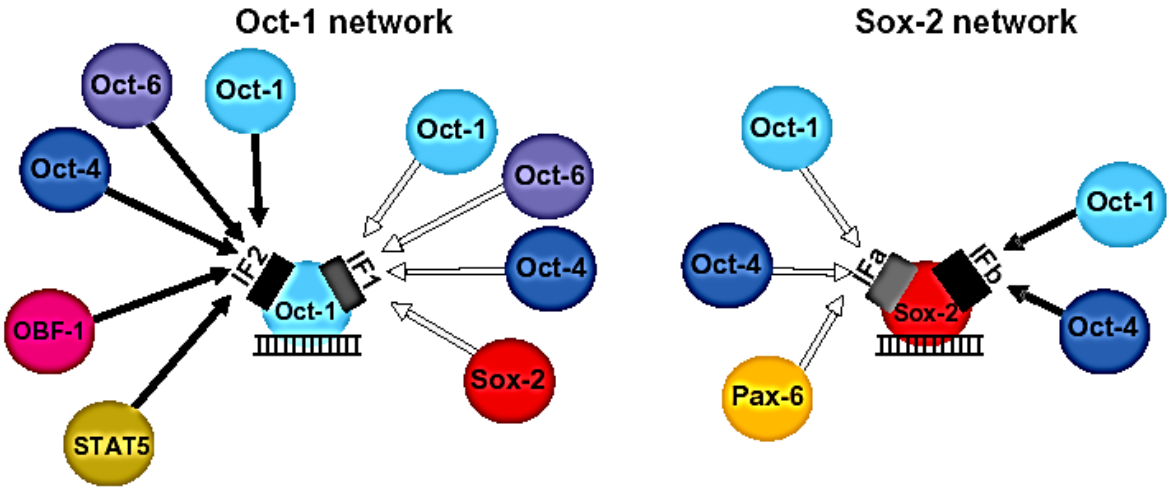
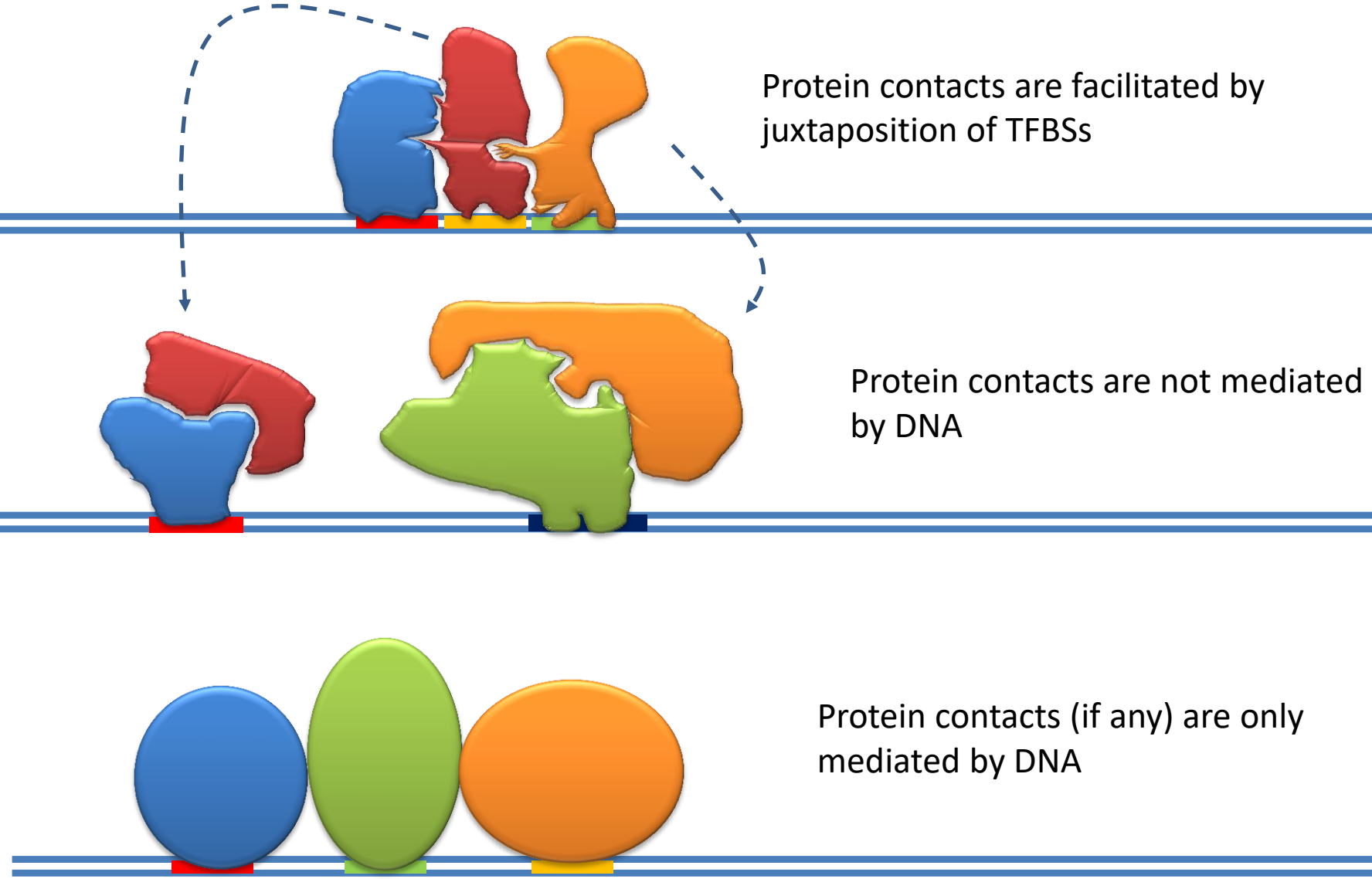


Figure 3 Interaction diagram of Oct-1 and Sox-2. Transcription factors are depicted as protein molecules with surface patches that can interact with a whole array of different partners provided that the protein is bound to a specific DNA element. DNA-bound Oct-1 and Sox-2 are depicted schematically with protein-protein interaction surface patches that are instrumental in binding to other partners. IF1 and IF2 on the Oct-1–DNA complex denote two interfaces of Oct-1 that are accessible and used for interaction on various DNA. Similarly, IFa and IFb designate interfaces of Sox-2 that are used for interaction on different DNA sites.



Protein contacts are facilitated by juxtaposition of TFBSs

Protein contacts are not mediated by DNA

Protein contacts (if any) are only mediated by DNA

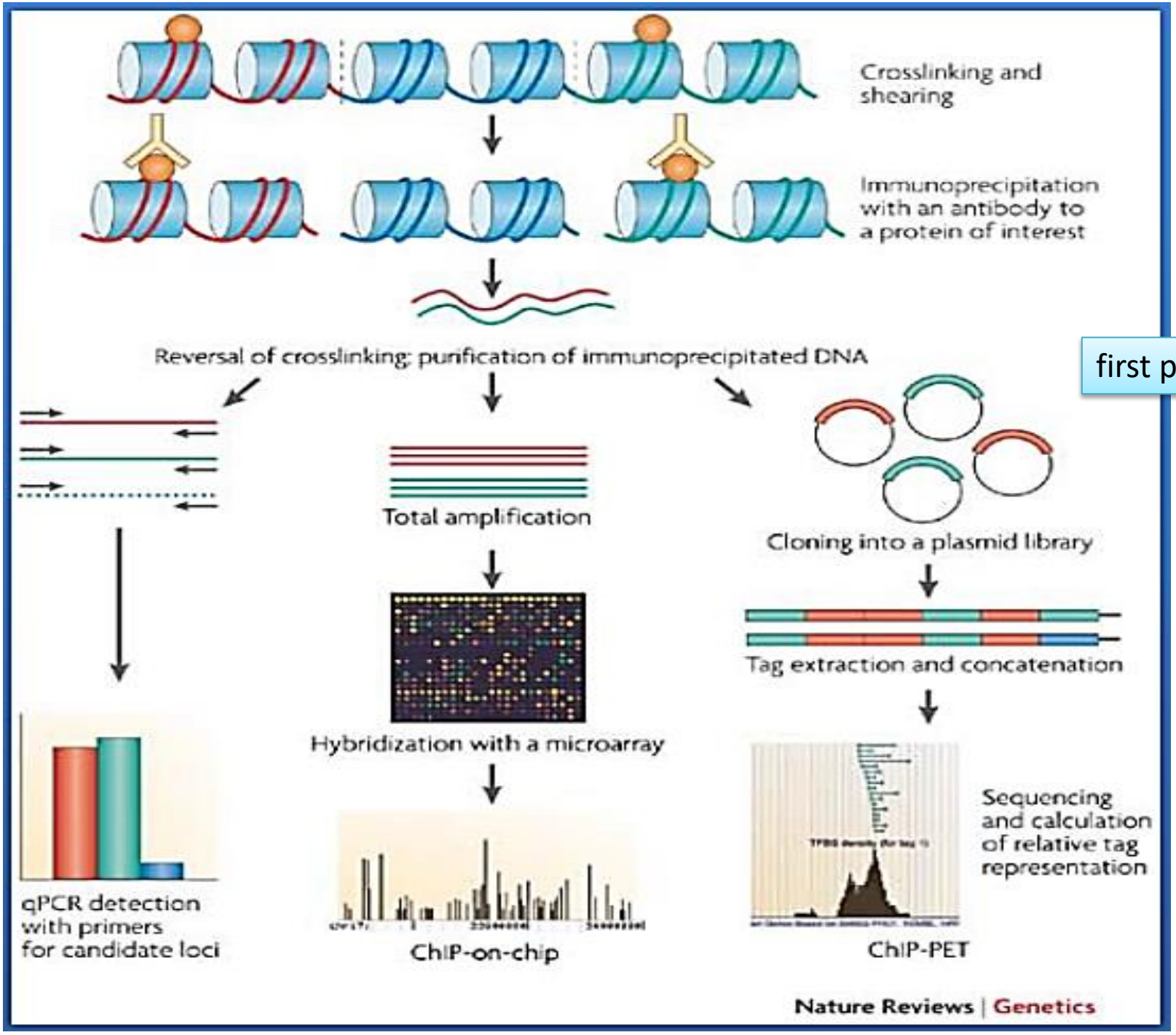
cooperativity

The binding of one Transcription Factor increases the probability of binding for a second TF and so forth

This is due often to chromatin «opening» that facilitates following TF binding

Factor added	Transcription
None	1
TF-A	2
TF-A + TF-B	3
TF-C	2
TF-A + TF-B + TF-C	25

Genome-wide identification of TFBS


Crosslinking and shearing
Immunoprecipitation with an antibody to a protein of interest
Reversal of crosslinking; purification of immunoprecipitated DNA
Total amplification
Cloning into a plasmid library
Tag extraction and concatenation
Hybridization with a microarray
qPCR detection with primers for candidate loci
ChIP-on-chip
Sequencing and calculation of relative tag representation
ChIP-PET
Nature Reviews | Genetics

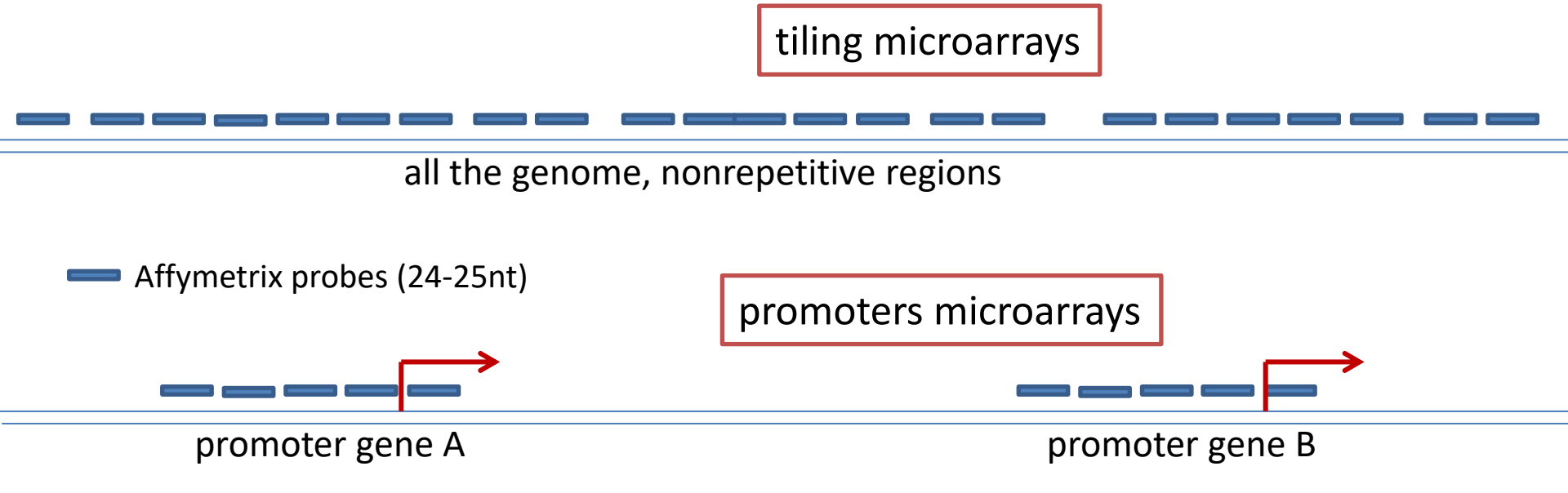
first phase

Caution !

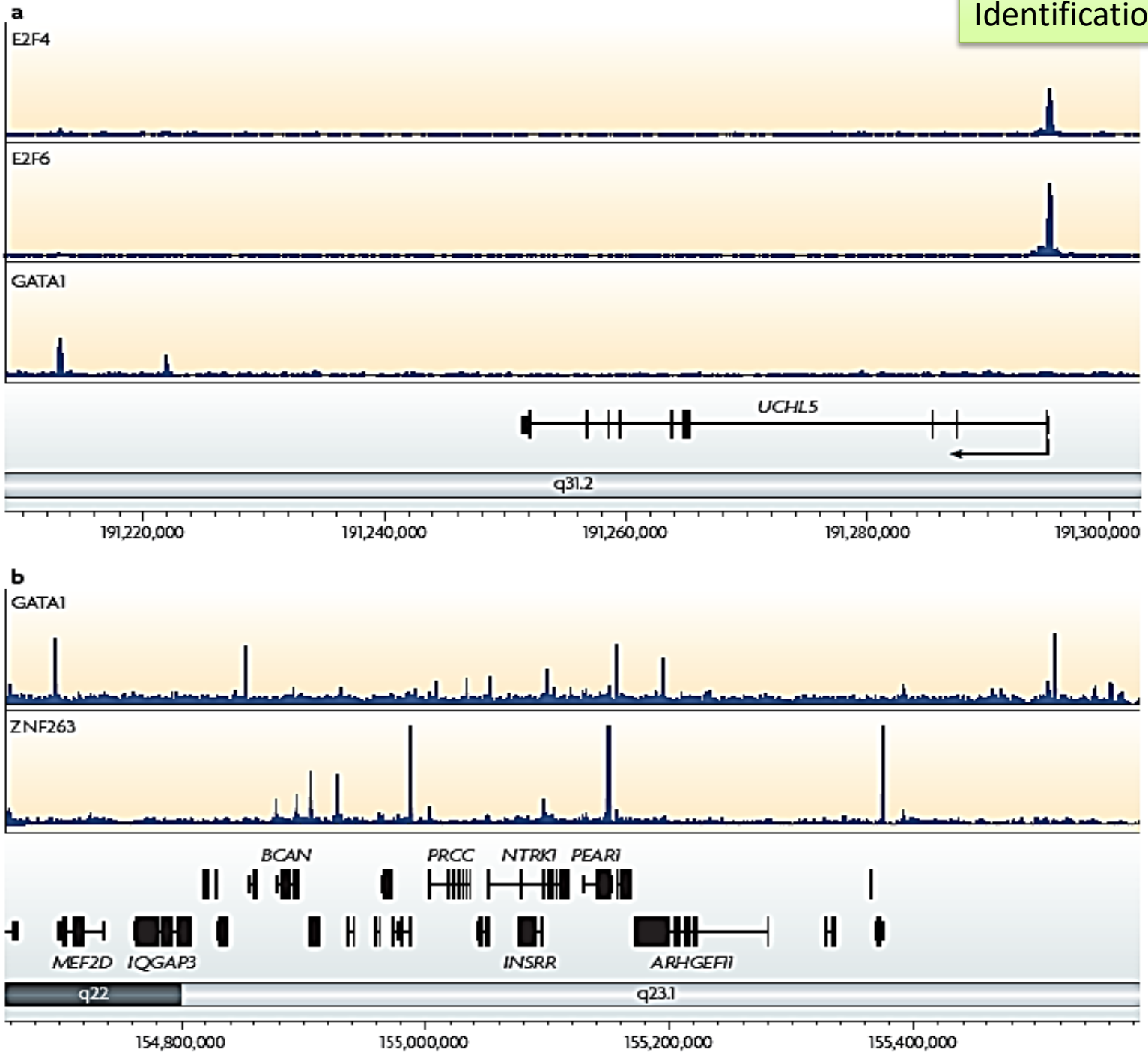
for TF mapping you can not use «expression microarrays»

(usually contain only expressed sequence probes)

Scientists used «tiling microarrays», i.e. probes covering the entire (nonripetitive) genome, or arrays containing only probes for known regulatory elements (e.g. known gene promoters).

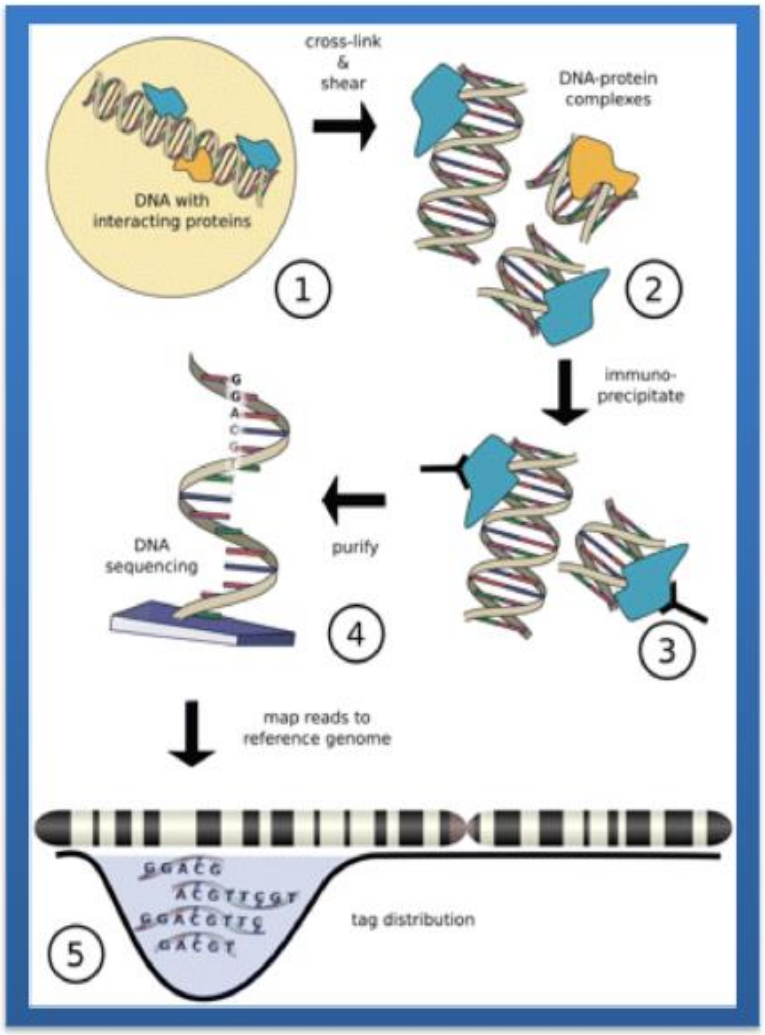


Identification of TFBS



Farnham 2009

Global Screening Methods



2. ChIP-seq

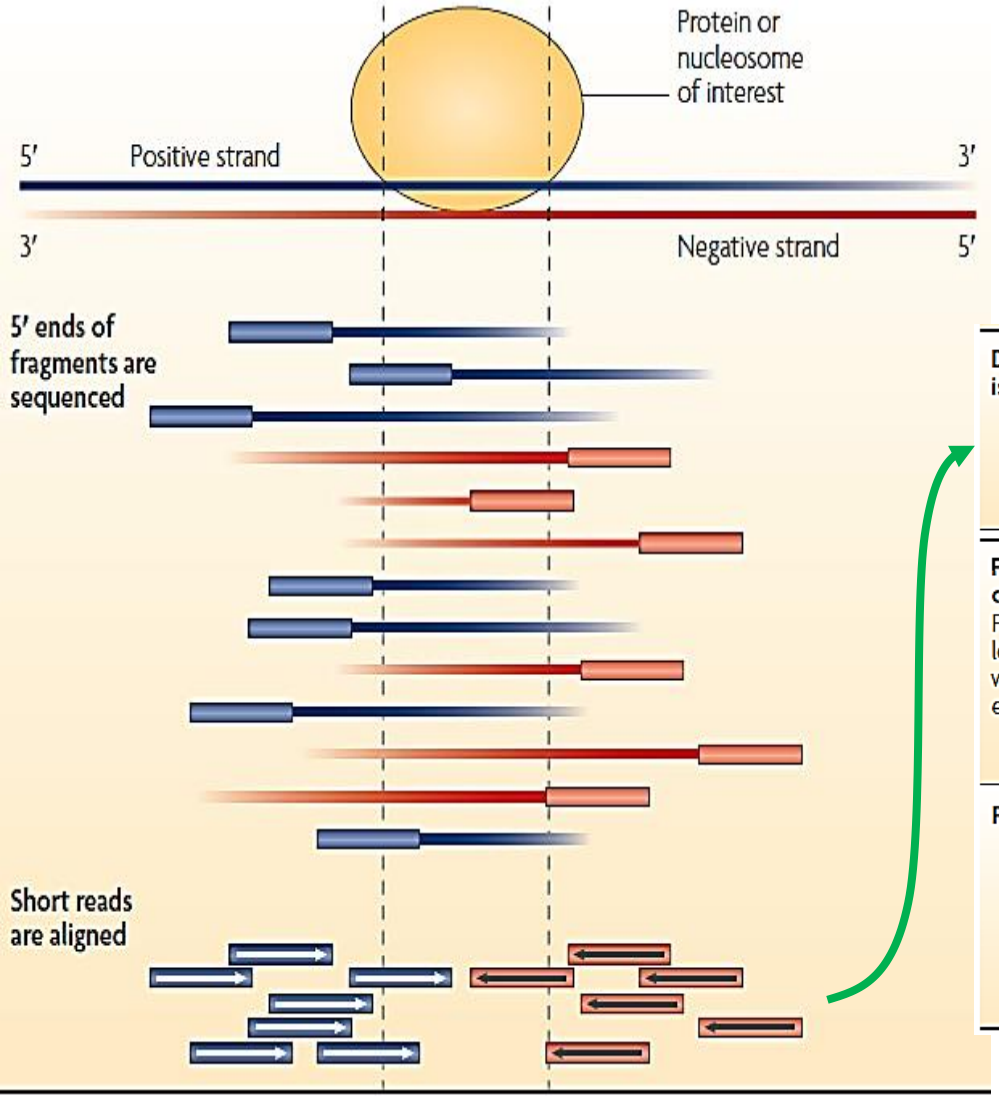
Perform regular ChIP, then sequence every DNA fragment immunoprecipitated (next-generation sequencing, Illumina or ABI-SOLID platforms)

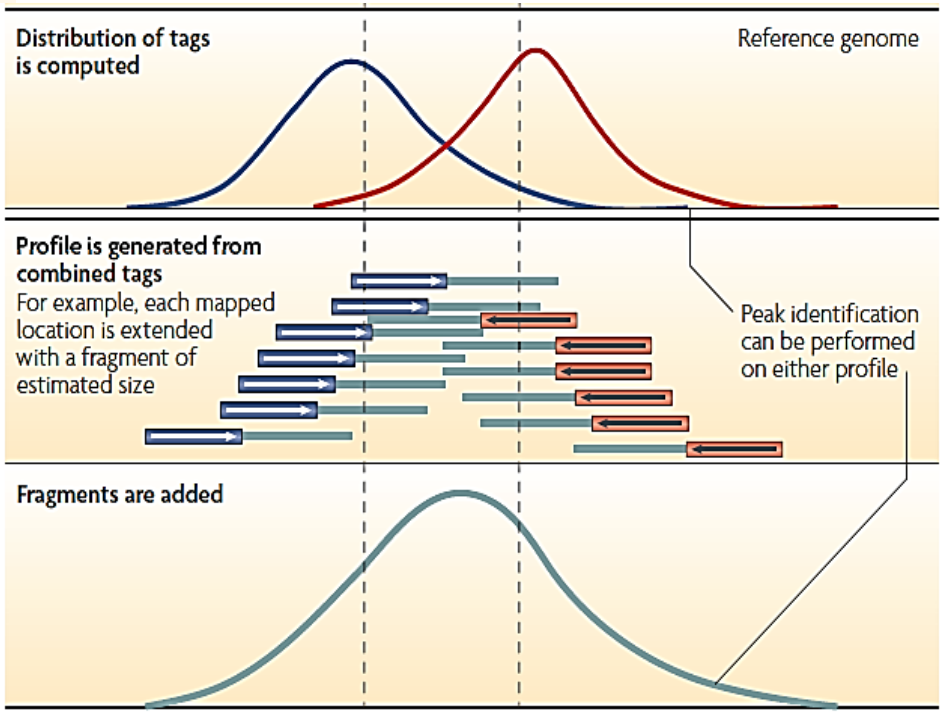
Advantages: little material required, higher resolution, fully open end approach, spatial resolution, less artifacts due to PCR amplification, possibility to multiplex, can do custom

Disadvantages: expensive (particularly if controls included), need large computer storage capacity, requires complex bioinformatics analysis.

Control sample: "Input" or "IgG"

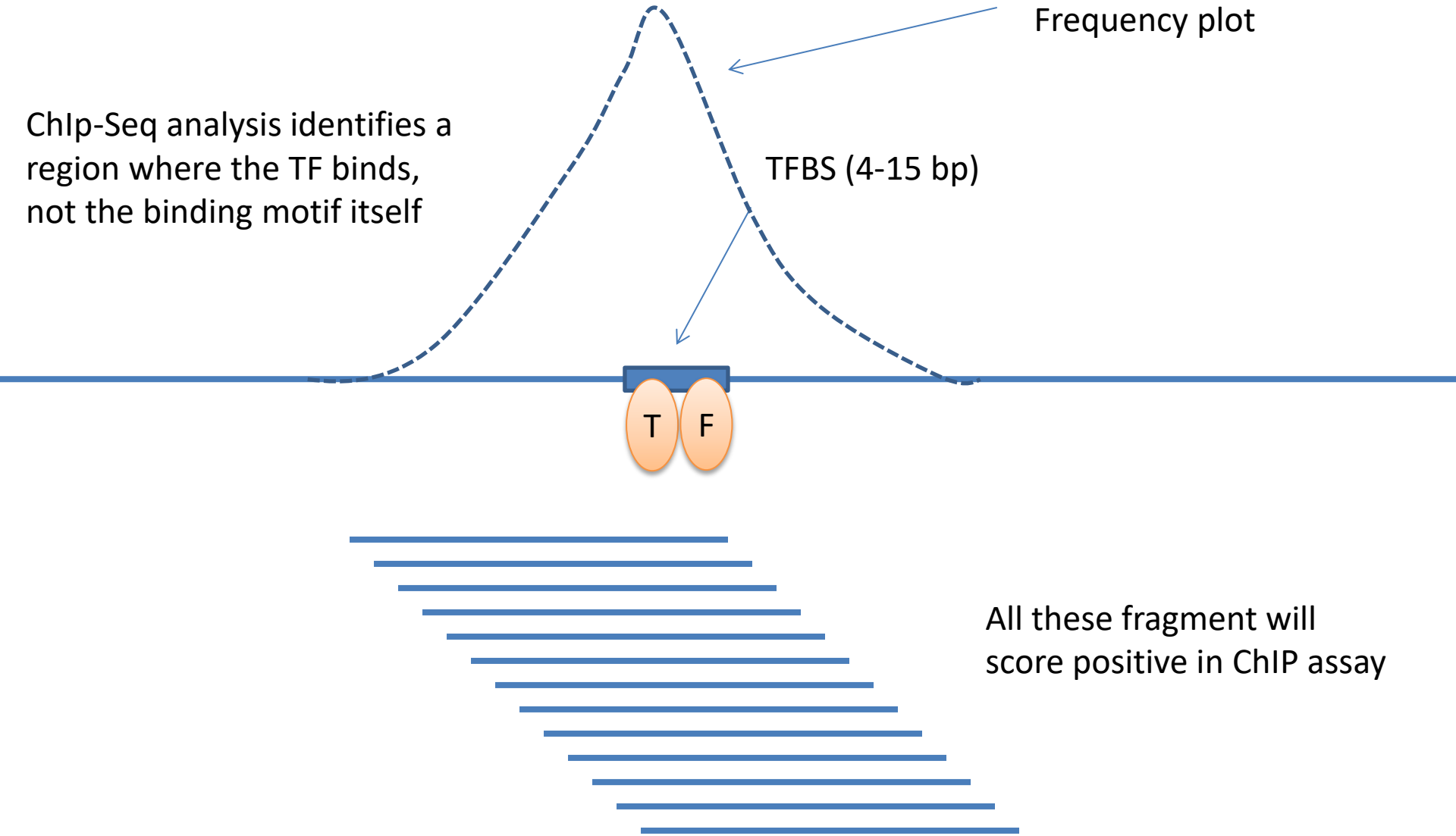
- Input: sonicated chromatin without immunoprecipitation
- IgG: "unspecific" IP





from Park, 2009, Nat Rev Genetics

Let us simplify a bit...



As a matter of fact, we have reads from <u>both strands</u>

join

peak call

peak annotation

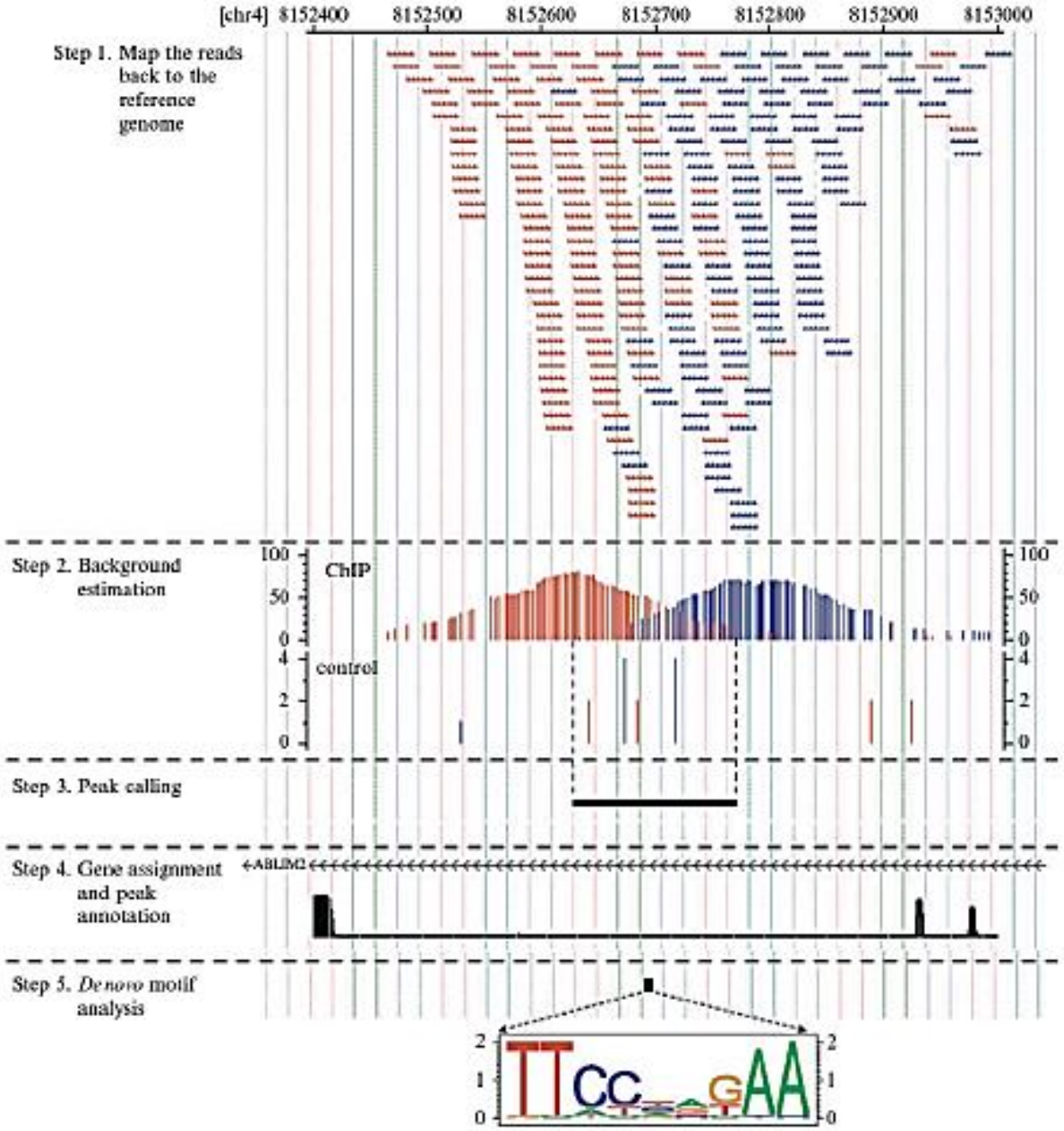
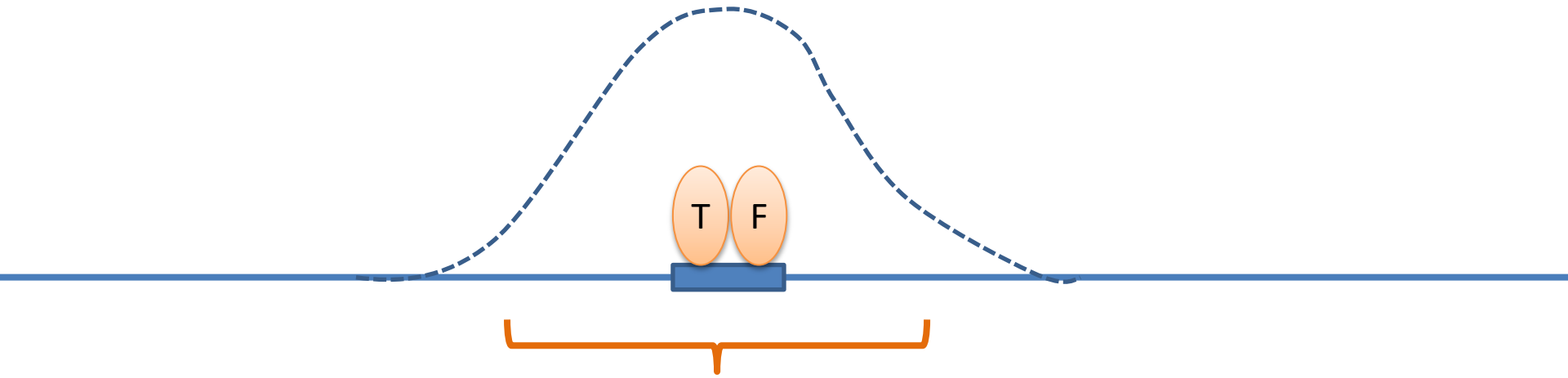


Figure 3.1 Work flow of ChIP-seq data processing and analyzing pipeline.

Statistical analysis defines
the binding region



a genome fragment around the center of the peak

(or defined by statistics)

can now be explored to predict the binding sequence

Algorithms exist for two different purposes:

1. to statistically evaluate the presence of a given known TF-binding motif in the list of «chipped» sequences

2. to evaluate, in the list of chipped sequences, the most represented «words» as compared to control sequences

Please consider that, even when a TF-binding motif is determined using direct biochemical techniques such as SELEX, you will never get a single, unvariable sequence motif

This is because the protein can adapt to certain base change at certain positions of the recognition sequence (less determinant contacts).

Example:
JASPAR database at http://jaspar.genereg.net/

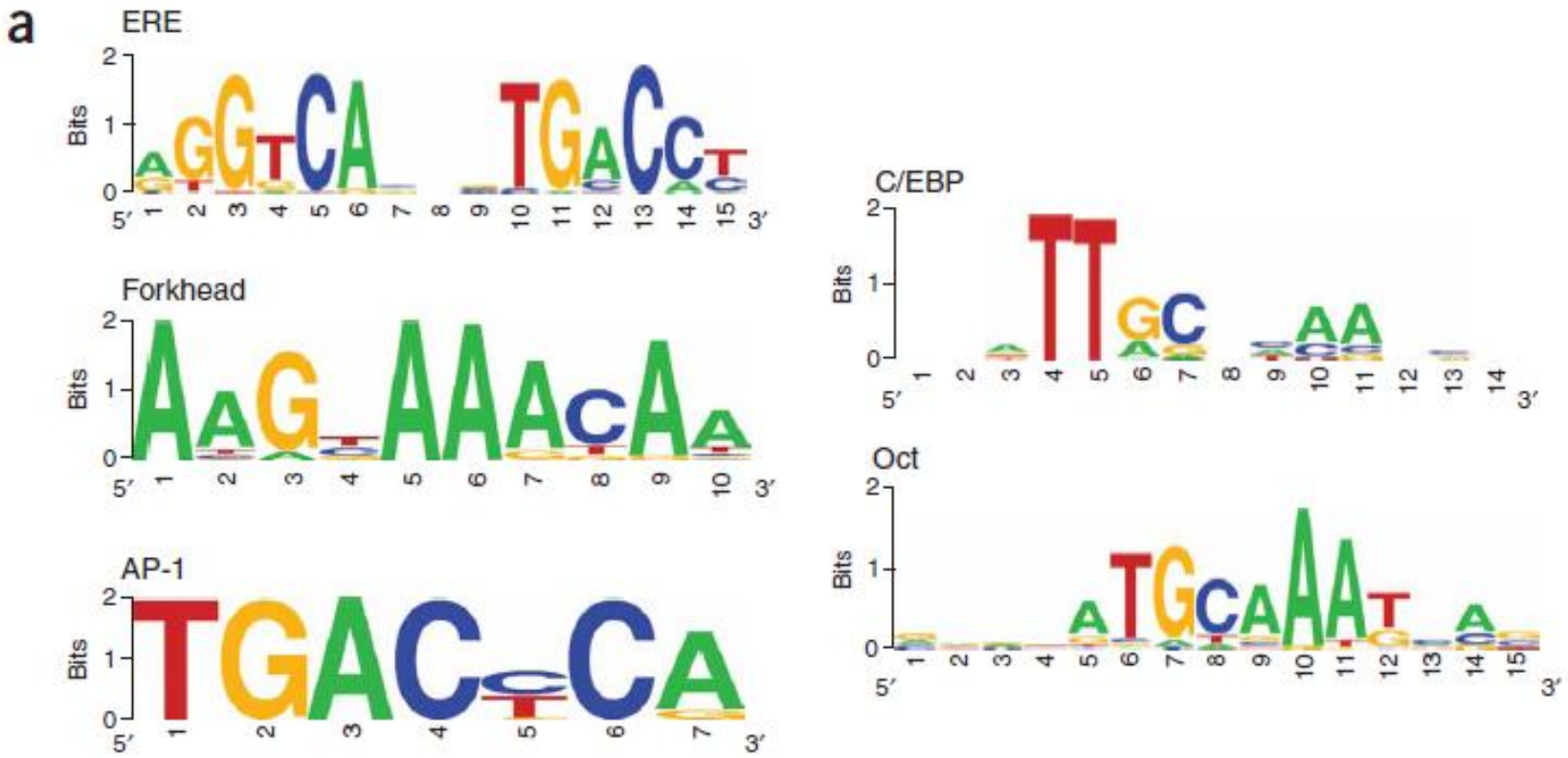
What is a PWM ?

https://en.wikipedia.org/wiki/Position_weight_matrix

For example take **-500, +500** interval around binding peak summits: algorithms exist to find unbiased overrepresented motifs, or known motifs, based on **positional weight matrices**.

see Bioinformatics

Examples of sites for different TFs:

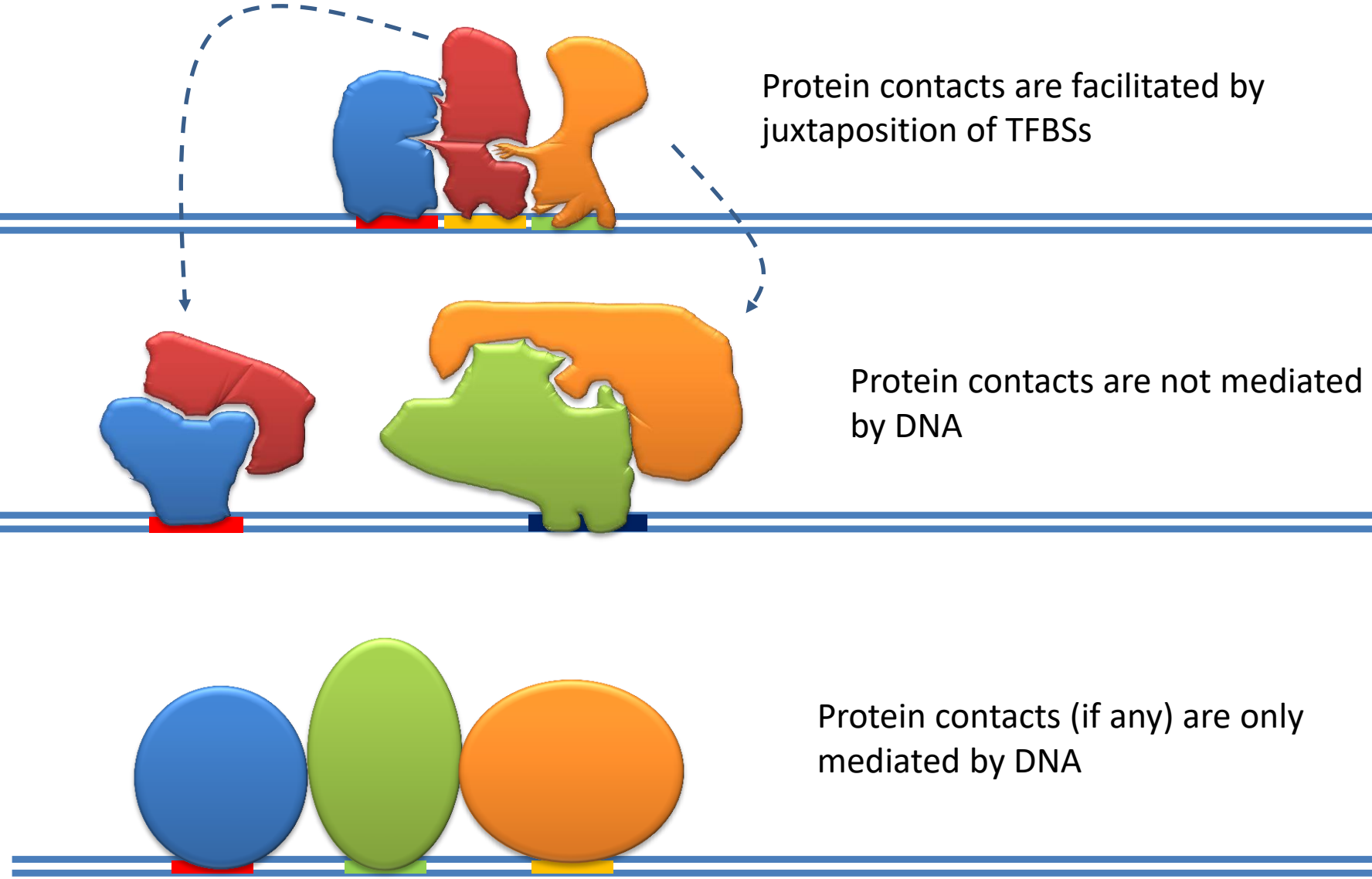


Caution !

ChIP-seq experiments identify TFBS NOT DNA/TF interaction

TFs show a complex protein-protein interaction network among themselves and with co-regulators (coactivators, corepressors, remodellers, other) (next lesson) and

formaldehyde will cross-link everything



Protein contacts are facilitated by juxtaposition of TFBSs

Protein contacts are not mediated by DNA

Protein contacts (if any) are only mediated by DNA

One important question is specificity

In the case seen as examples, the «order» of TFs is dictated by DNA sequence

In higher eukaryotic genomes, do we find any kind of combination of TFBS ? In other words, any kind of TF cooperation is allowed ?

Example of a TF analysis by ChIP-Seq → paper

ENCODE studies (in the following)

ARTICLE

doi:10.1038/nature11245

Architecture of the human regulatory network derived from ENCODE data

Mark B. Gerstein[1,2,3]*, Anshul Kundaje[4]*, Manoj Hariharan[5]*, Stephen G. Landt[5]*, Koon-Kiu Yan[1,2]*, Chao Cheng[1,2]*, Xinmeng Jasmine Mu[1]*, Ekta Khurana[1,2]*, Joel Rozowsky[2]*, Roger Alexander[1,2]*, Renqiang Min[1,2,6]*, Pedro Alves[1]*, Alexej Abyzov[1,2], Nick Addleman[5], Nitin Bhardwaj[1,2], Alan P. Boyle[5], Philip Cayting[5], Alexandra Charos[7], David Z. Chen[3], Yong Cheng[5], Declan Clarke[8], Catharine Eastman[5], Ghia Euskirchen[5], Seth Frietze[9], Yao Fu[1], Jason Gertz[10], Fabian Grubert[5], Arif Harmanci[1,2], Preti Jain[10], Maya Kasowski[5], Phil Lacroute[5], Jing Leng[1], Jin Lian[11], Hannah Monahan[7], Henriette O'Geen[12], Zhengqing Ouyang[5], E. Christopher Partridge[10], Dorrelyn Patacsil[5], Florencia Pauli[10], Debasish Raha[7], Lucia Ramirez[5], Timothy E. Reddy[10]†, Brian Reed[7], Minyi Shi[5], Teri Slifer[5], Jing Wang[1], Linfeng Wu[5], Xinqiong Yang[5], Kevin Y. Yip[1,2,13], Gili Zilberman-Schapira[1], Serafim Batzoglou[4], Arend Sidow[14], Peggy J. Farnham[9], Richard M. Myers[10], Sherman M. Weissman[11] & Michael Snyder[5]

Transcription factors bind in a combinatorial fashion to specify the on-and-off states of genes; the ensemble of these binding events forms a regulatory network, constituting the wiring diagram for a cell. To examine the principles of the human transcriptional regulatory network, we determined the genomic binding information of 119 transcription-related factors in over 450 distinct experiments. We found the combinatorial, co-association of transcription factors to be highly context specific: distinct combinations of factors bind at specific genomic locations. In particular, there are significant differences in the binding proximal and distal to genes. We organized all the transcription factor binding into a hierarchy and integrated it with other genomic information (for example, microRNA regulation), forming a dense meta-network. Factors at different levels have different properties; for instance, top-level transcription factors more strongly influence expression and middle-level ones co-regulate targets to mitigate information-flow bottlenecks. Moreover, these co-regulations give rise to many enriched network motifs (for example, noise-buffering feed-forward loops). Finally, more connected network components are under stronger selection and exhibit a greater degree of allele-specific activity (that is, differential binding to the two parental alleles). The regulatory information obtained in this study will be crucial for interpreting personal genome sequences and understanding basic principles of human biology and disease.

Transcription Factors + transcription-related factors by ChIP-Seq.

The sum of chromatin sites bound by a given TF in a given cell type under a specific experimental condition is called «**cistrome**»

Very often, single TF bind to different chromatin sites in different cell types (i.e. TFs display cell type-specific cistromes).

Whenever TFs collapse (+other histone and cofactor marks) → enhancer

Using data from 5 cell lines, the ENCODE project has identified:

- Total 7.5 million «peaks» (40% of these within 2.5Kbp from TSS).

This allowed to estimate around 400,000 putative enhancers in the human genome (Gerstein et al., 2012). (some recent estimates reach up to one million).

Therefore, **enhancers largely outnumber promoters**.

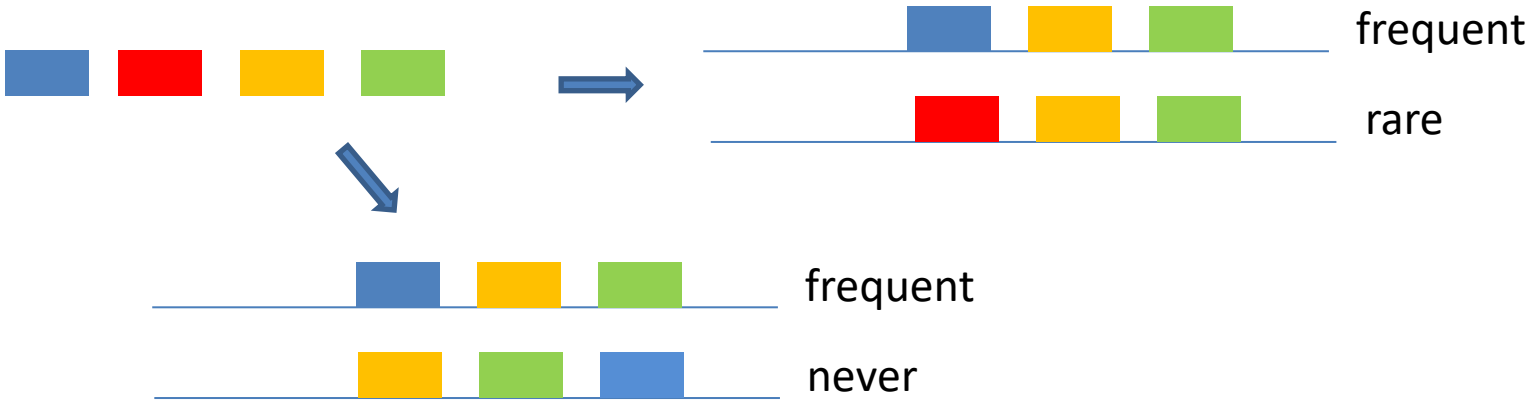
The combinatorial rule of Transcription Factor Binding Sites at enhancers

Questions:

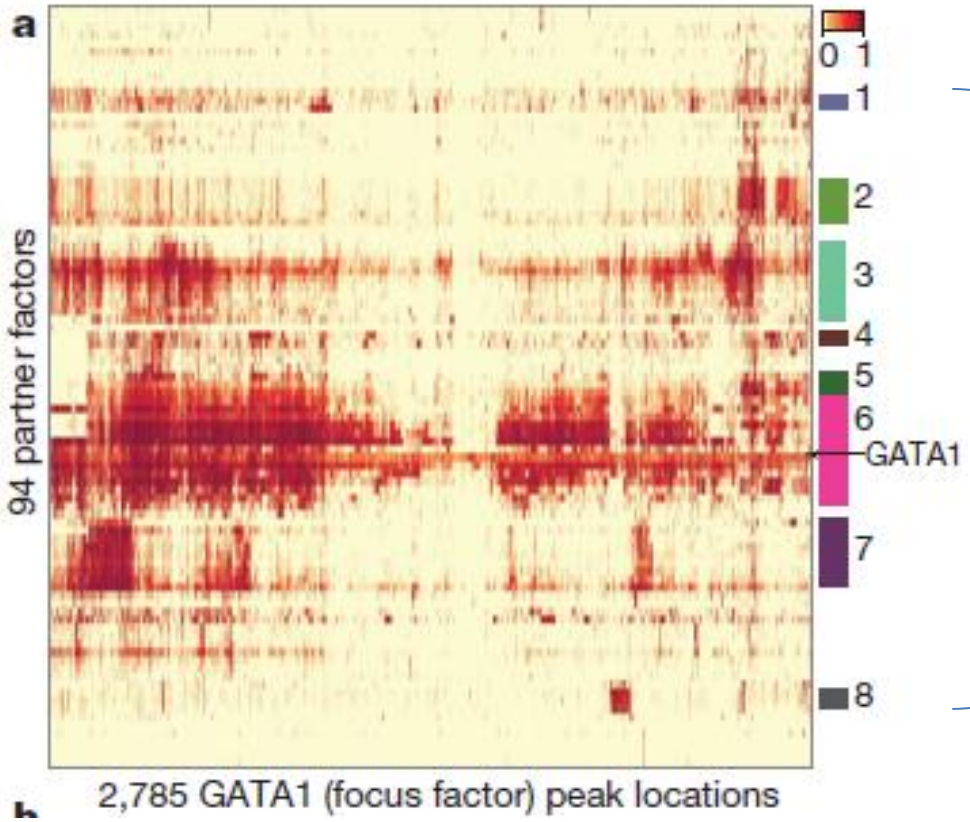
- Is the «combinatorial» rule for TFs at enhancers true ?
- Can any TF combine with any other TF ?

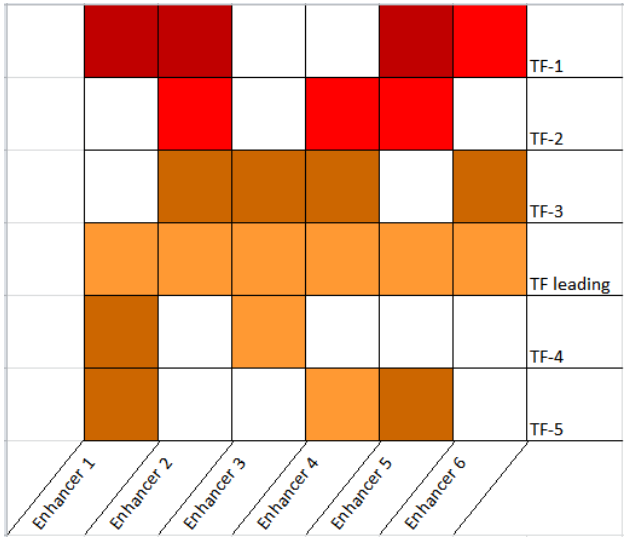
Gerstein paper clearly indicates that different «combinatorial» groups exist or, in other words, not all the possible combinations are seen.

This implies co-evolution of regulatory modules. This conclusion is further emphasized by the obervation that in a given conserved module, different TFBS are often arranged (ordered) in the same way.

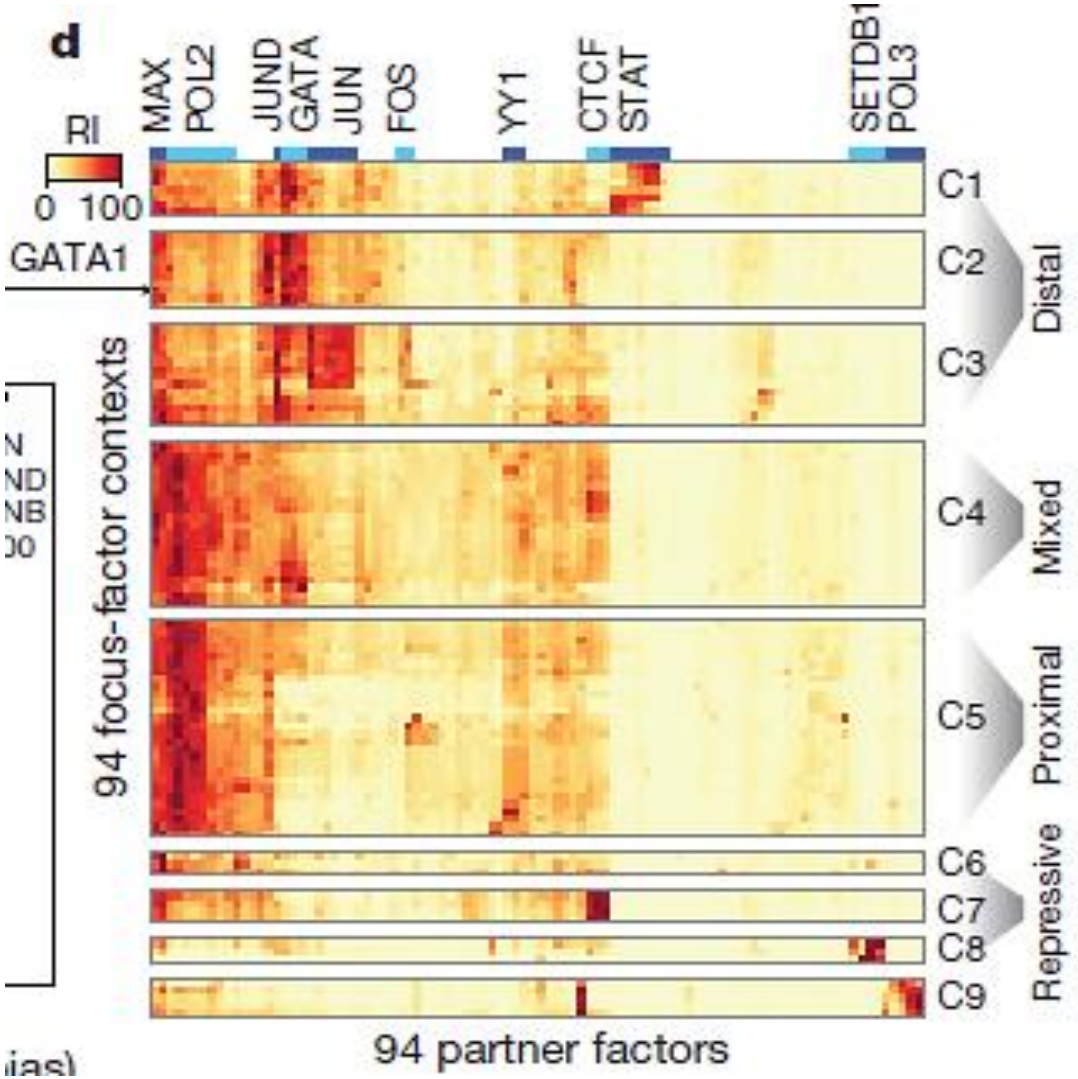


One example from Gernstein paper: GATA1





The «focus» factor here is GATA1. Merging all the experiments,
94 partner factors were found.
Taking into account co-presence, several clusters are defined.



Coassociation shows groups that are quite **strictly delimited in one type of regulatory element** (e.g. Distal or proximal).
Coassociation with REST or HDAC2 also define some «repressive» elements

This «clustering» effect also suggests that enhancers have spread in the genome by duplication events (transposition ?)

There is also evidence that in some cases Transposable Elements have been domisticated to act as enhancers

Testori et al. BMC Genomics 2012, 13:400
http://www.biomedcentral.com/1471-2164/13/400

BMC Genomics

RESEARCH ARTICLE Open Access

The role of Transposable Elements in shaping the combinatorial interaction of Transcription Factors

Alessandro Testori[1,3*], Livia Caizzi[1,2], Santina Cutrupi[1,5], Olivier Friard[1], Michele De Bortoli[1,3], Davide Cora'[1,4*†] and Michele Caselle[1,6*†]

Abstract

Background: In the last few years several studies have shown that Transposable Elements (TEs) in the human genome are significantly associated with Transcription Factor Binding Sites (TFBSs) and that in several cases their expansion within the genome led to a substantial rewiring of the regulatory network. Another important feature of the regulatory network which has been thoroughly studied is the combinatorial organization of transcriptional regulation. In this paper we combine these two observations and suggest that TEs, besides rewiring the network, also played a central role in the evolution of particular patterns of combinatorial gene regulation.

Results: To address this issue we searched for TEs overlapping Estrogen Receptor α (ERα) binding peaks in two publicly available ChIP-seq datasets from the MCF7 cell line corresponding to different modalities of exposure to estrogen. We found a remarkable enrichment of a few specific classes of Transposons. Among these a prominent role was played by MIR (Mammalian Interspersed Repeats) transposons. These TEs underwent a dramatic expansion at the beginning of the mammalian radiation and then stabilized. We conjecture that the special affinity of ERα for the MIR class of TEs could be at the origin of the important role assumed by ERα in Mammalians. We then searched for TFBSs within the TEs overlapping ChIP-seq peaks. We found a strong enrichment of a few precise combinations of TFBS. In several cases the corresponding Transcription Factors (TFs) were known cofactors of ERα, thus supporting the idea of a co-regulatory role of TFBS within the same TE. Moreover, most of these correlations turned out to be strictly associated to specific classes of TEs thus suggesting the presence of a well-defined "transposon code" within the regulatory network.

Conclusions: In this work we tried to shed light into the role of Transposable Elements (TEs) in shaping the regulatory network of higher eukaryotes. To test this idea we focused on a particular transcription factor: the Estrogen Receptor α (ERα) and we found that ERα preferentially targets a well defined set of TEs and that these TEs host combinations of transcriptional regulators involving several of known co-regulators of ERα. Moreover, a significant number of these TEs turned out to be conserved between human and mouse and located in the vicinity (and thus candidate to be regulators) of important estrogen-related genes.

Keywords: Transposable elements, ChIP-seq, Transcription factors, ERα, Combinatorial interaction

Lesson 2 will continue with Videos from

Eric Wieschaus
https://www.youtube.com/watch?v=Ncxs21KEj0g

and

Michael Levine
https://www.youtube.com/watch?v=Cfgze3BItDY
+
https://www.youtube.com/watch?v=FLa3oRPxNLM

These Videos are availabe in our Moodle site, Chapter 3, Book: Ch 3 Lesson streaming