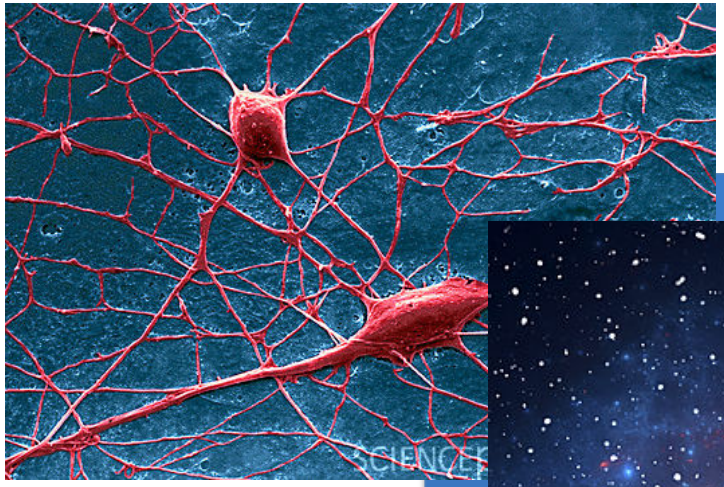
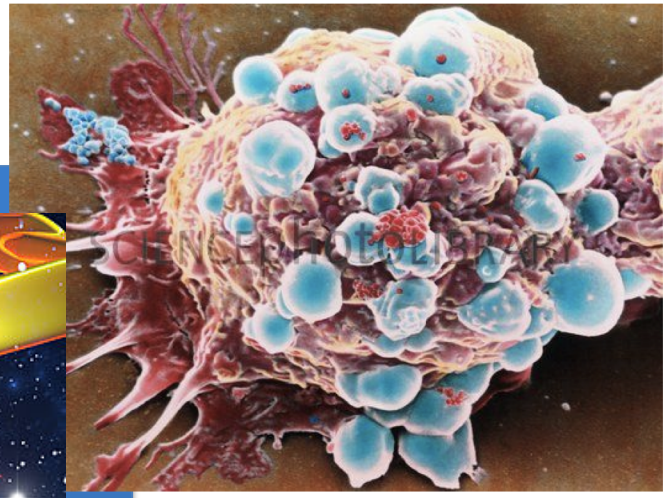
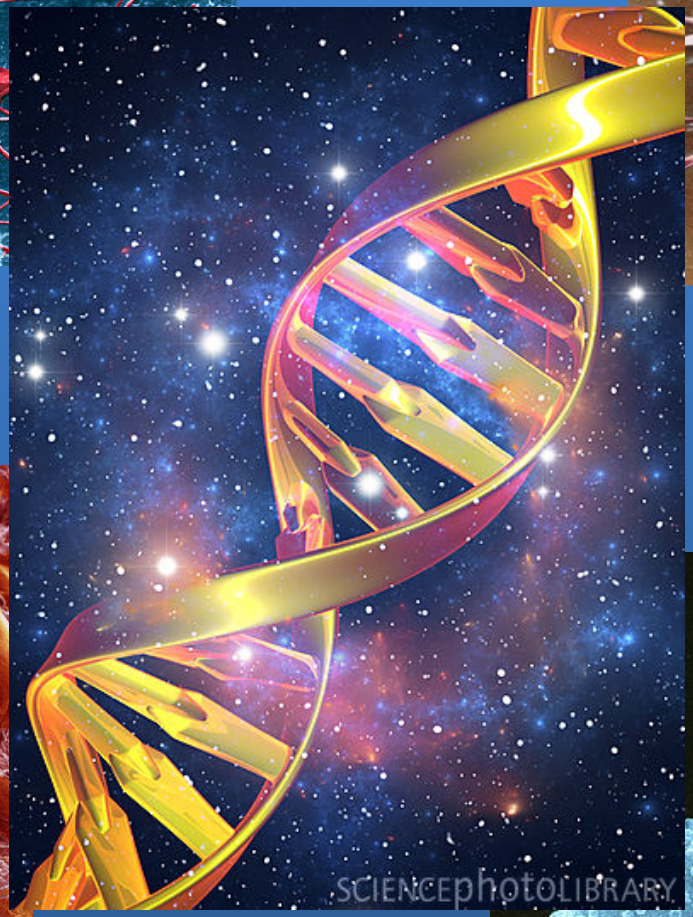
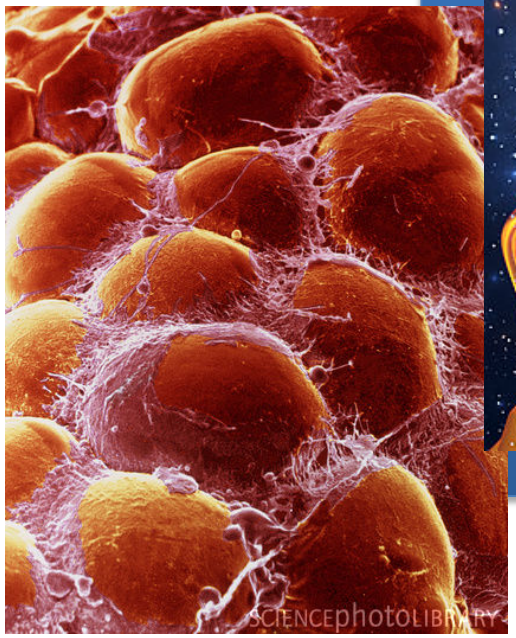
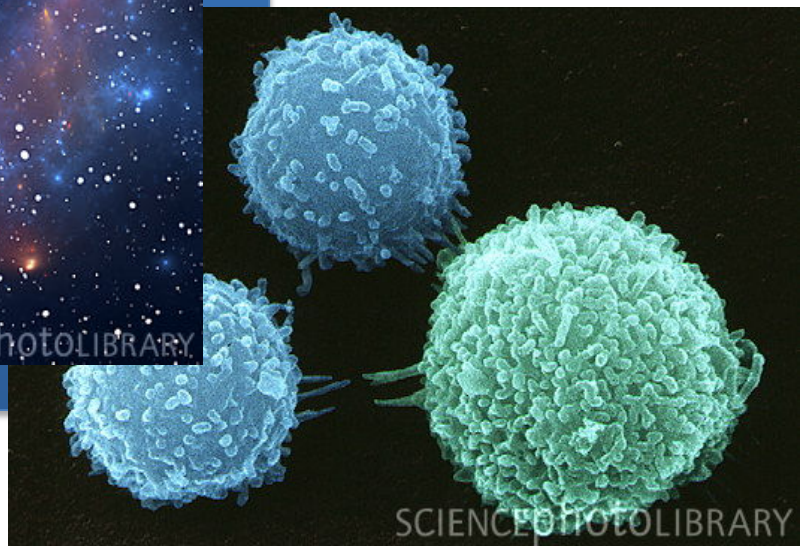
Part 4 – Transcriptional regulation



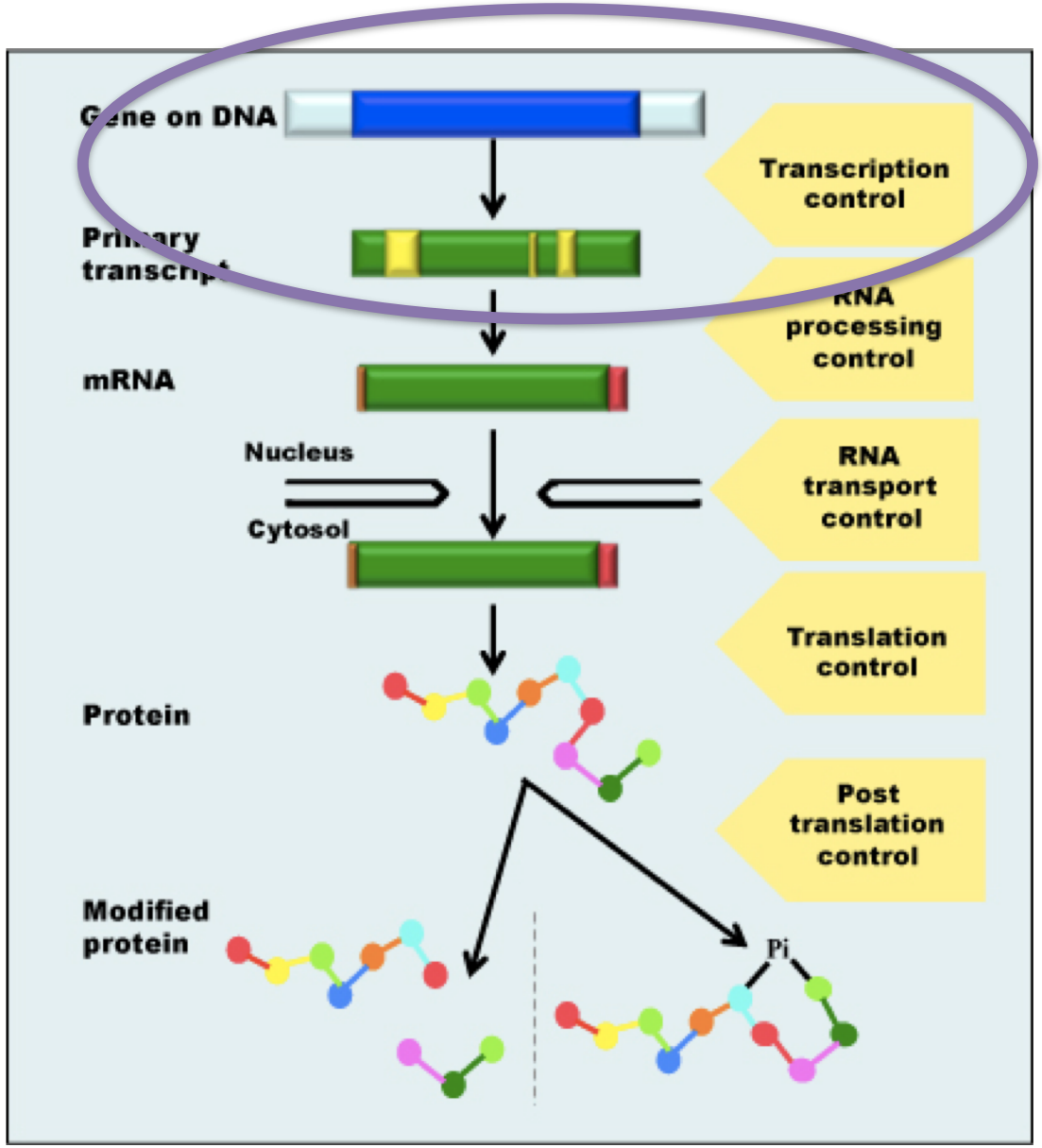



SCIENCEphotoLIBRARY


SCIENCEphotoLIBRARY


SCIENCEphotoLIBRARY

Levels of regulation of gene expression



Textbook

Leading Edge
Review

Cell

Looping Back to Leap Forward: Transcription Enters a New Era

Michael Levine,[1,*] Claudia Cattoglio,[1,2] and Robert Tjian[1,2,*]
[1]Department of Molecular and Cell Biology
[2]Howard Hughes Medical Institute, CIRM Center of Excellence, Li Ka Shing Center for Biomedical and Health Sciences
University of California, Berkeley, Berkeley, CA 94707, USA
*Correspondence: mlevine@berkeley.edu (M.L.), jmlim@uclink4.berkeley.edu (R.T.)
http://dx.doi.org/10.1016/j.cell.2014.02.009

Comparative genome analyses reveal that organismal complexity scales not with gene number but with gene regulation. Recent efforts indicate that the human genome likely contains hundreds of thousands of enhancers, with a typical gene embedded in a milieu of tens of enhancers. Proliferation of *cis*-regulatory DNAs is accompanied by increased complexity and functional diversification of transcriptional machineries recognizing distal enhancers and core promoters and by the high-order spatial organization of genetic elements. We review progress in unraveling one of the outstanding mysteries of modern biology: the dynamic communication of remote enhancers with target promoters in the specification of cellular identity.

Levine, 2014

Introduction

Transcription regulation is the premier mechanism underlying differential gene activity in animal development and disease.

differential gene activity

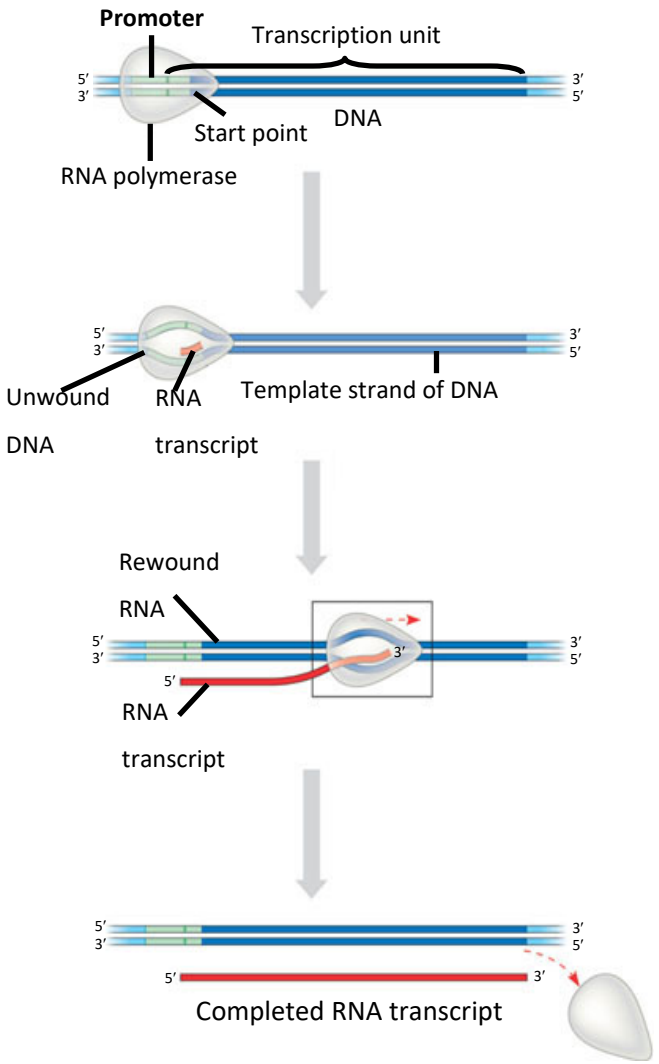
How is transcription of different genes regulated ?

Transcription steps

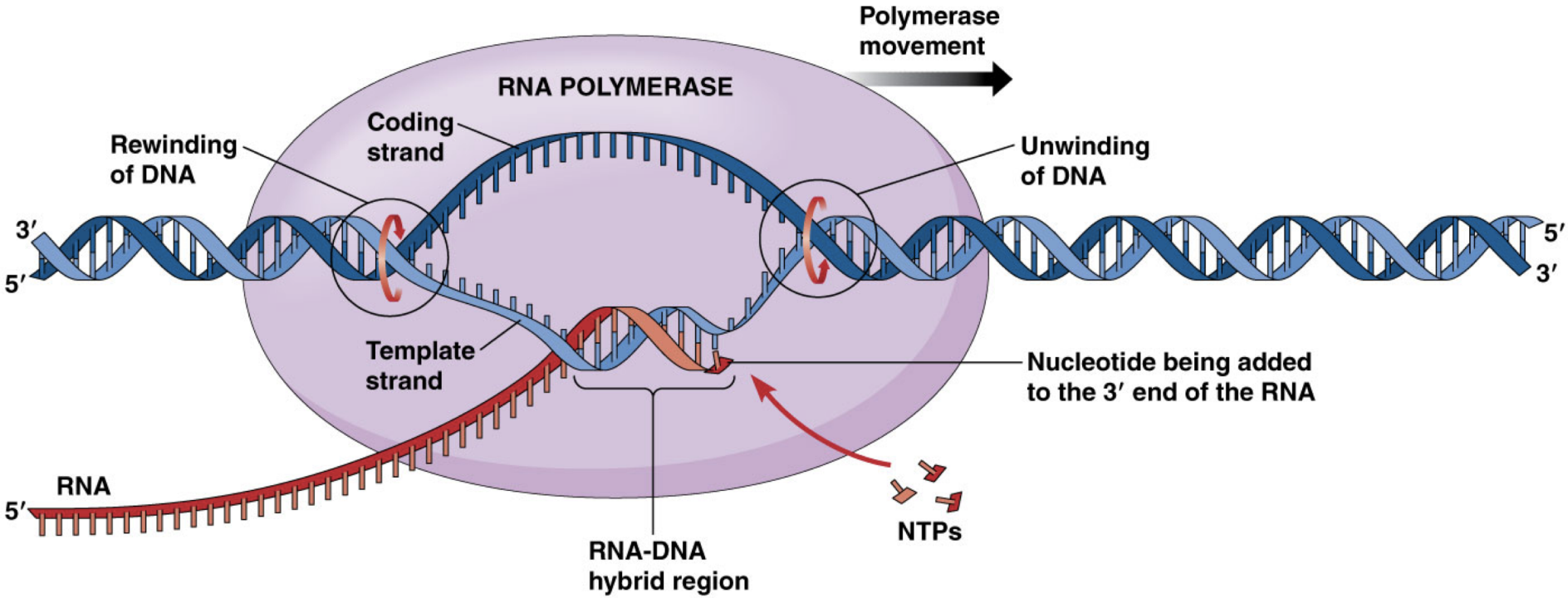
– Initiation

– Elongation

– Termination



RNA Polymerase

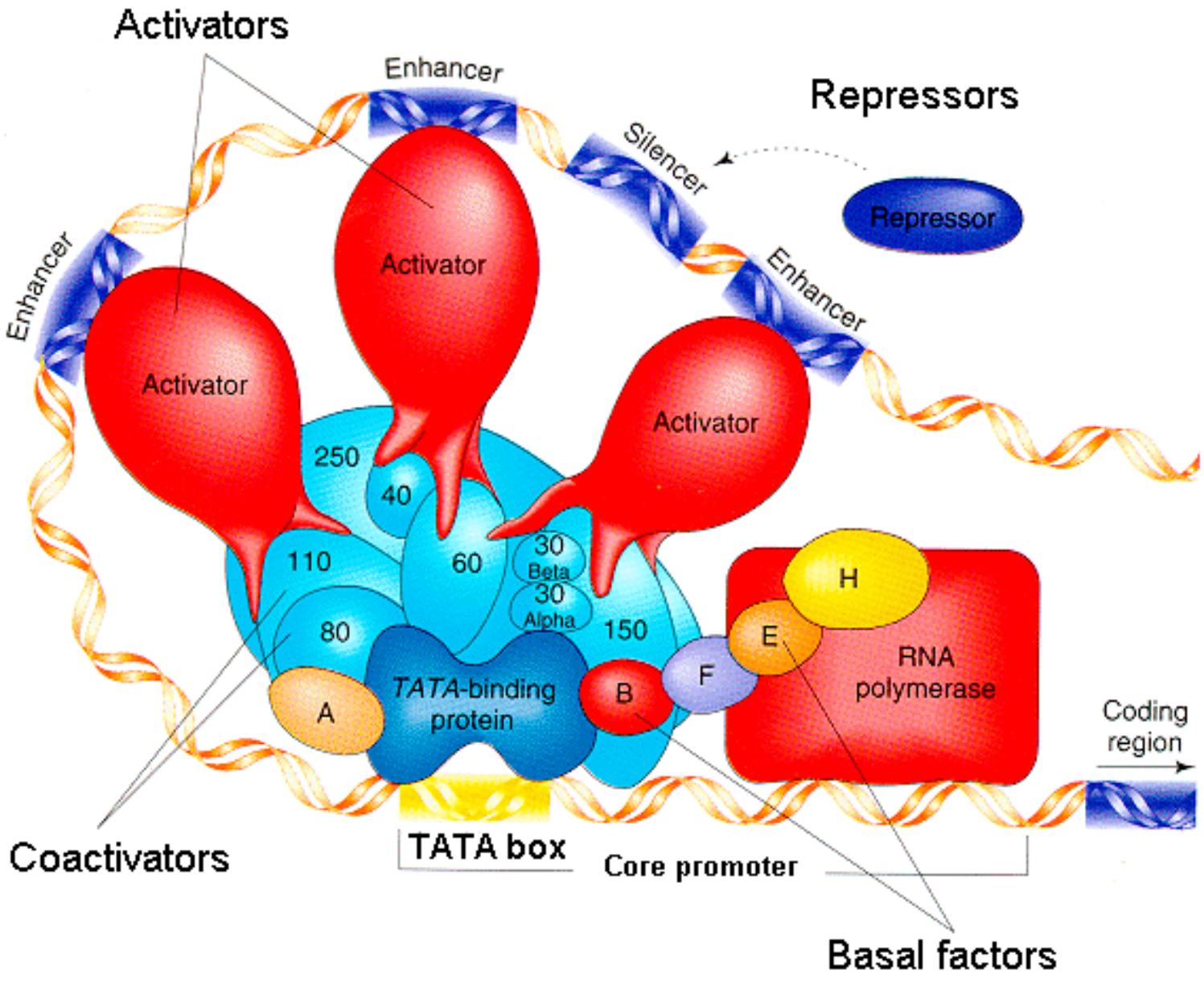


Eukaryotes RNA polymerase

1. **RNA Pol I:** Transcribes RNA found in ribosomes
   - 28s ribosomal RNA
   - 18s rRNA
   - 5.8s rRNA

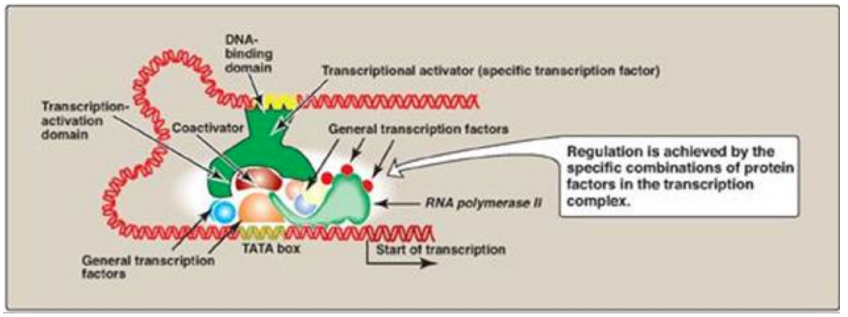
2. **RNA Pol II:** Transcribes
   - m-RNA
   - Some sn-RNA

3. **RNA Pol III:** Transcribes
   - t-RNA
   - 5s rRNA
   - Some sn-RNA

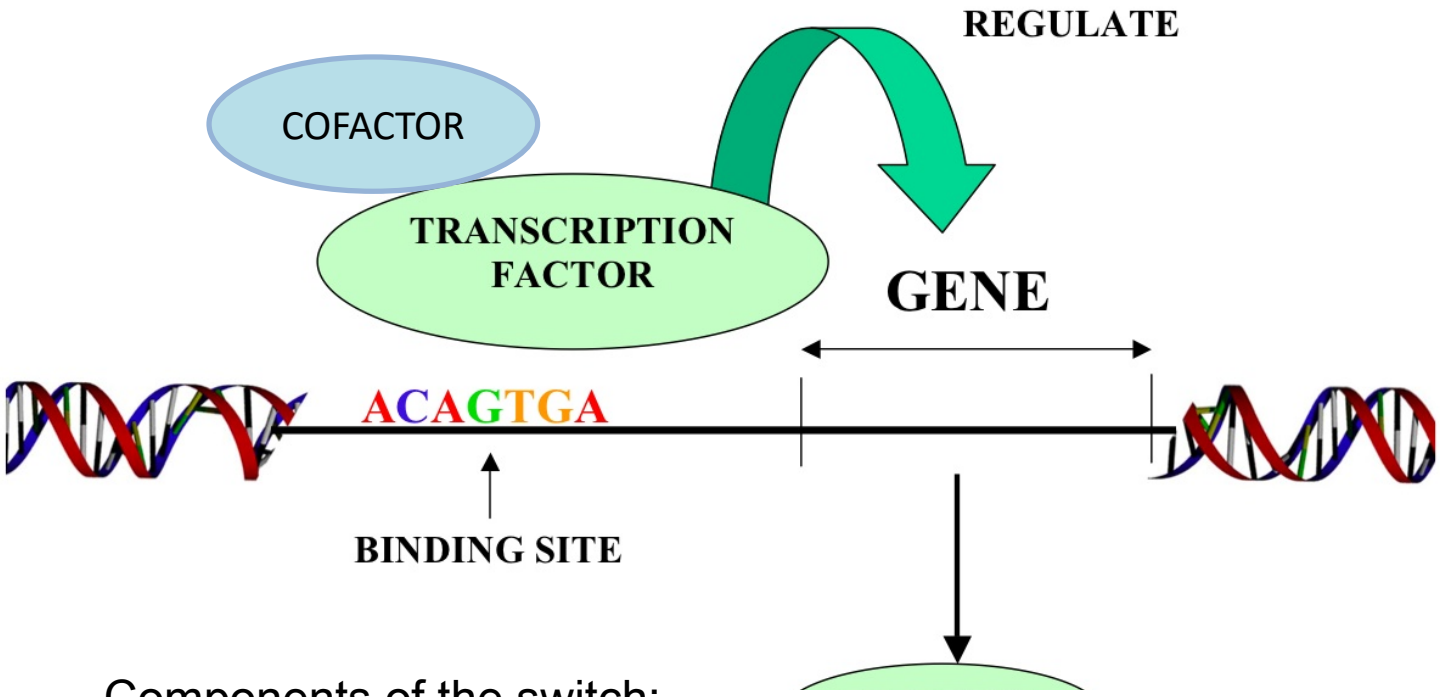

Activators
Enhancer
Repressors
Silencer
Repressor
Enhancer
Enhancer
Activator
Activator
Activator
250
40
110
60
30
Beta
30
Alpha
80
150
TATA-binding
protein
A
B
F
E
H
RNA
polymerase
Coding
region
Coactivators
TATA box
Core promoter
Basal factors

Factors required to regulate transcription

- Polymerase
- GTFs
- Mediator complex
- DNA-binding regulatory transcription factors
- Corepressors and Coactivators (including different classes of chromatin-modifying enzymes)



Regulated gene expression is achieved through on-off switches responding to environmental or developmental changes

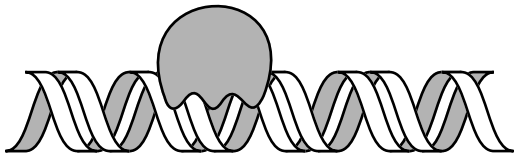


Components of the switch:

1- Short stretches of DNA regulatory sequences

2- Gene regulatory proteins

How to Study DNA Protein Interactions?



AGENDA

Different approaches to study Protein-DNA interactions

- Physical Interaction
  - Electrophoretic Mobility Shift Assay (EMSA)
  - DNA Footprinting
  - Chromatin Immunoprecipitation (ChIP)

- Functional readout of Protein-DNA Interactions
  - Reporter Assays
  - *In vitro* transcription

Different approaches to study Protein-DNA interactions

- Physical Interaction
  - Electrophoretic Mobility Shift Assay (EMSA)
  - DNA Footprinting
  - Chromatin Immunoprecipitation (ChIP)

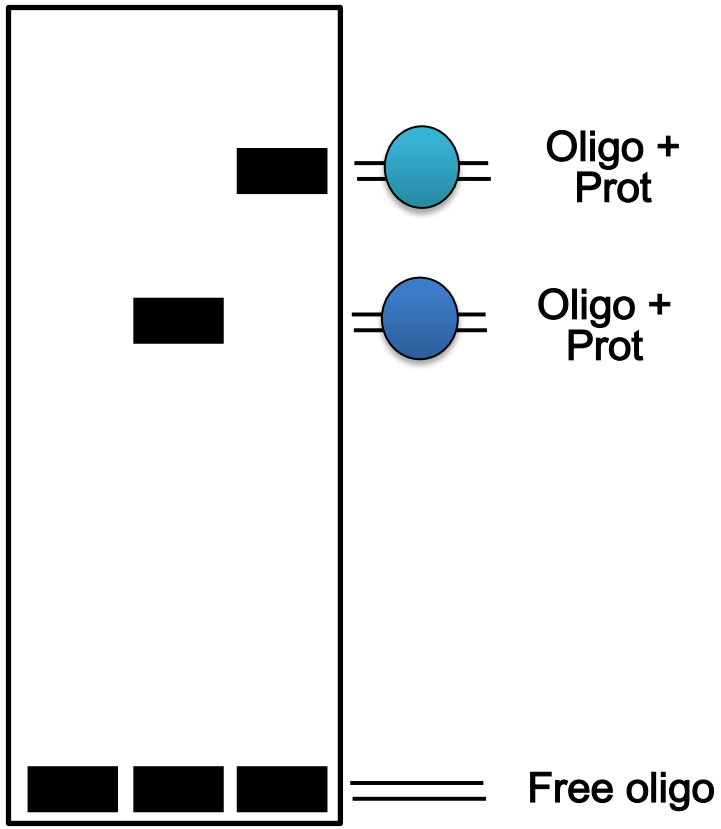
- Functional readout of Protein-DNA Interactions
  - Reporter Assays
  - *In vitro* transcription

Gel Shifts/Electrophoretic Mobility Shift Assays

- In vitro analysis of the transcriptional factor binding function
- Binding does not always correlate with transcriptional activity

Steps

1. Nuclear extracts from cells or tissues
2. Mix with ^{32}P-labeled ds-oligo
3. Run on Native acrylamide gels



Strengths of method

→detect low abundance DNA binding proteins from lysates

→test binding affinity with mutations

Limitations:

→in vitro, quantitation problems,

→need antibodies to identify complexes

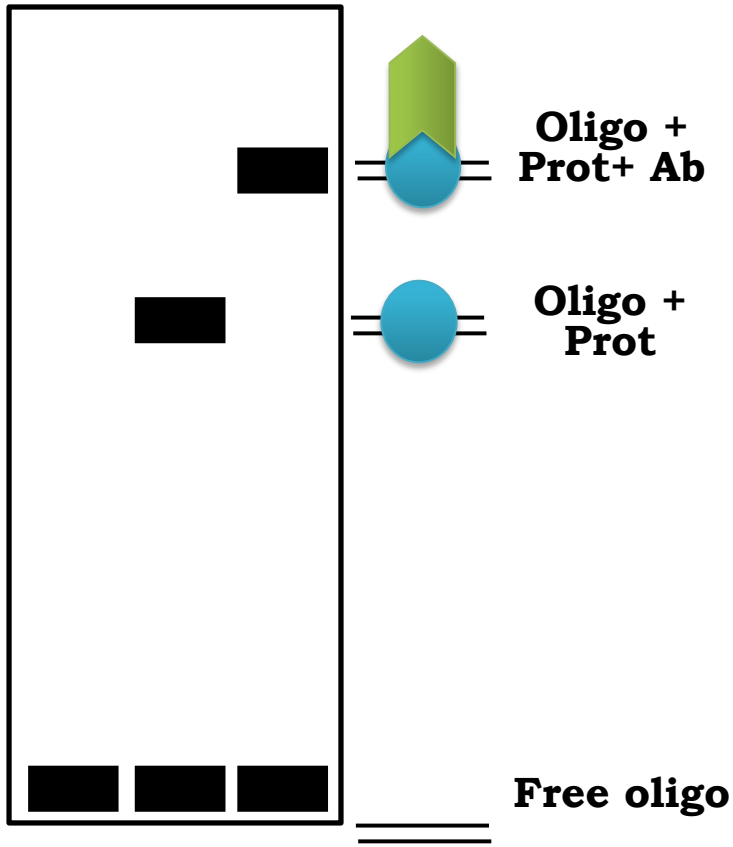
How do we determine the identity of complexes and if they are specific?

- **Competition assays**

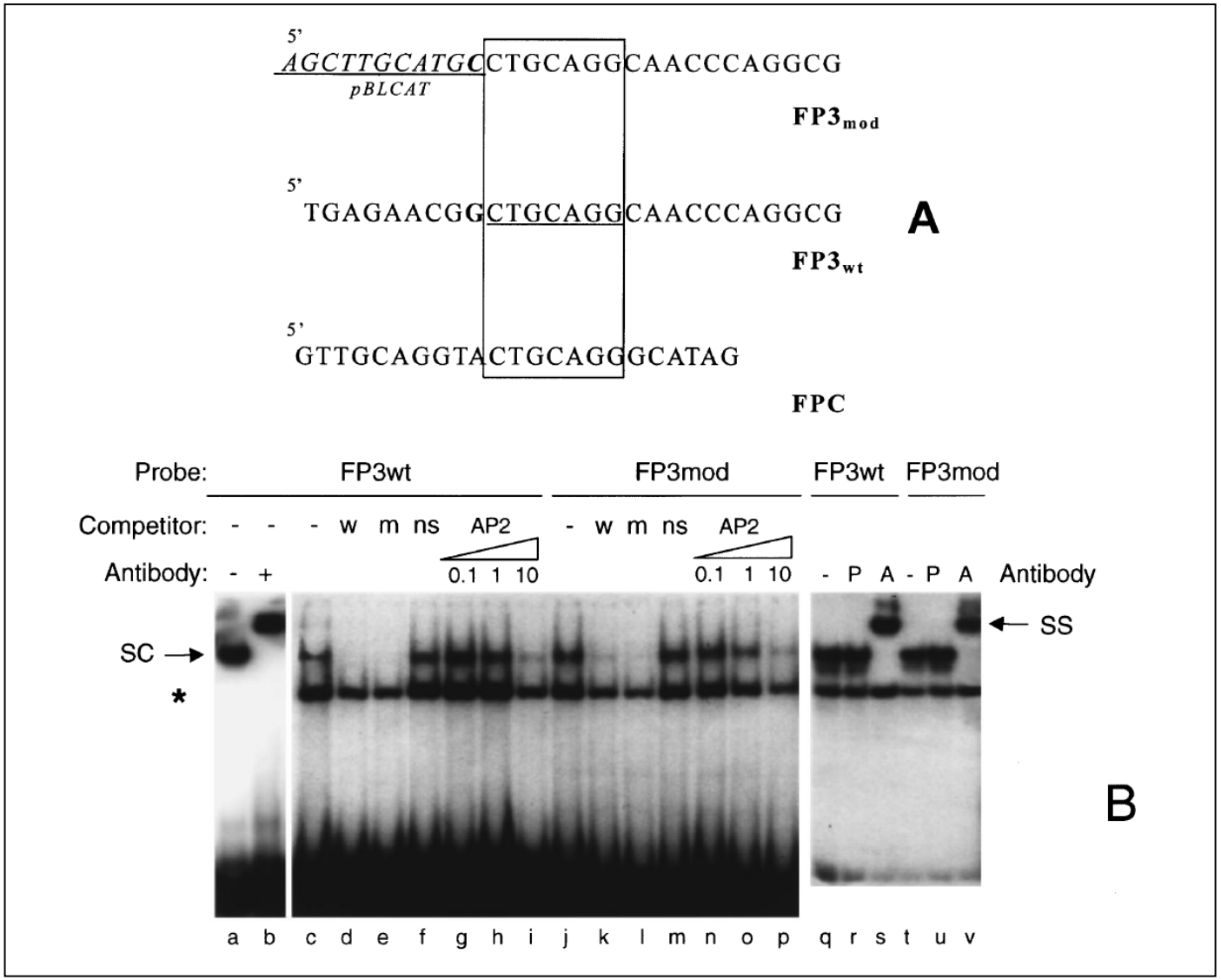
  Molar excess of identical, mutant, or consensus site

- **Supershift Assays**

  Add specific antibody to the binding assay



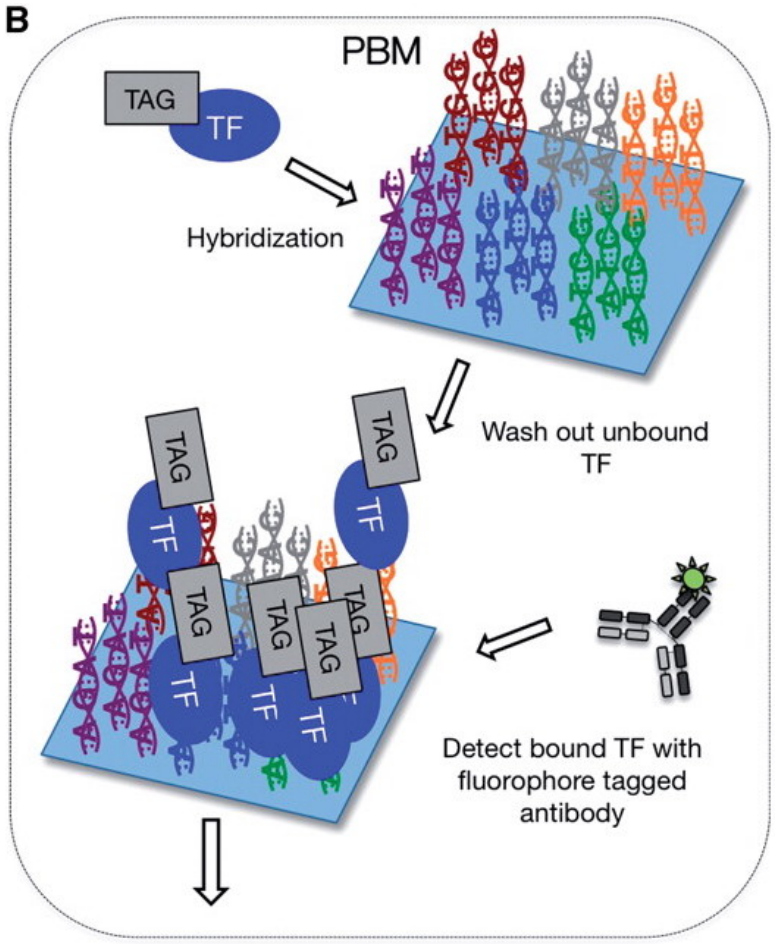
Examples: Gel Shifts/EMSA



Perissi et al., Oncogene, 2000

Protein Binding Microarrays

HT approach to investigate protein binding to double stranded DNA sequences



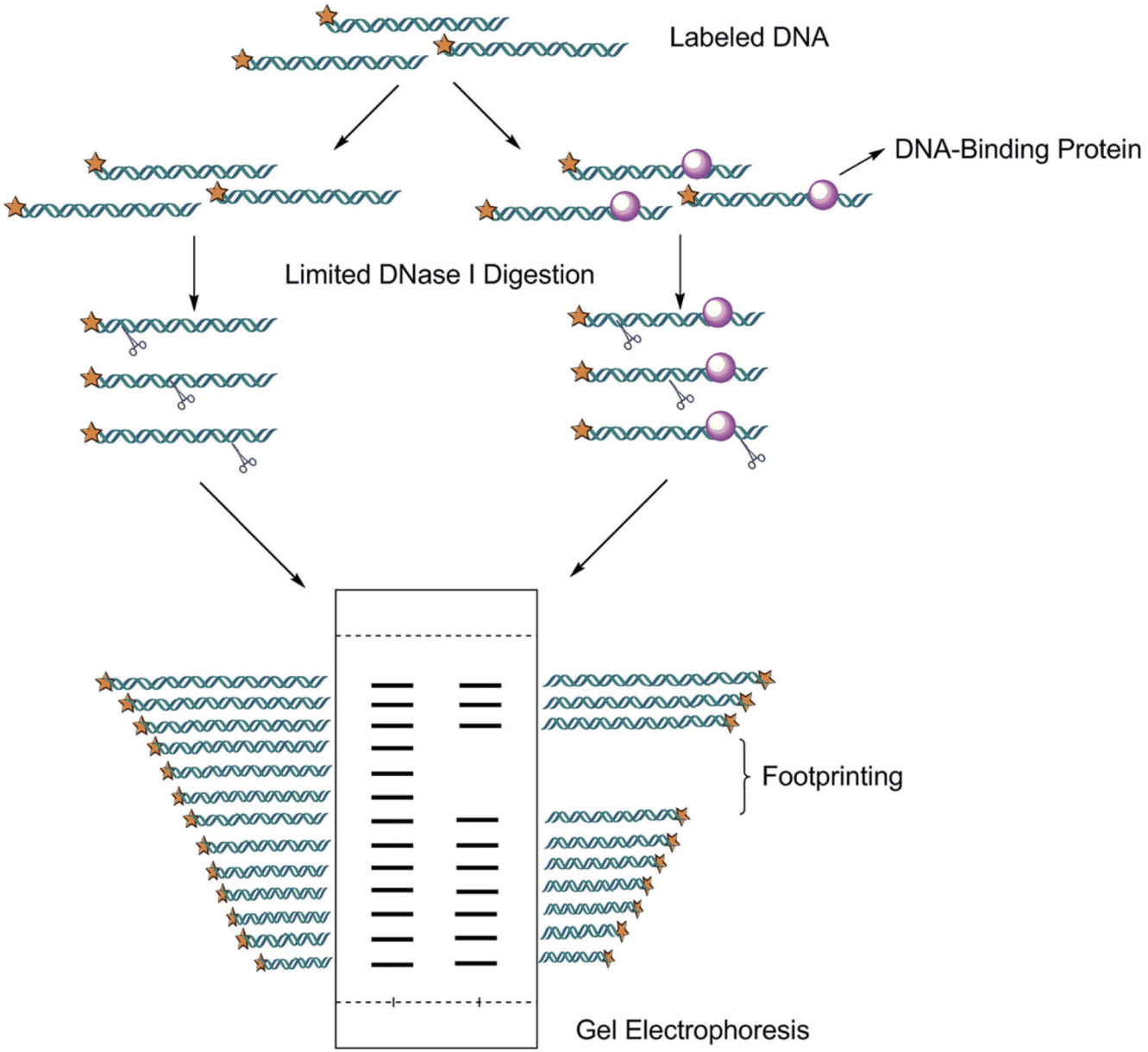
- Quantification of bound proteins is done using fluorescent labeled antibodies
- PBMs can be used for measuring protein binding to either synthetic or genomic sequences
- PBMs are useful for investigating binding specificity among homologs and for studying the binding of multi-protein complexes

Different approaches to study Protein-DNA interactions

- Physical Interaction
  - Electrophoretic Mobility Shift Assay (EMSA)
  - DNA Footprinting
  - Chromatin Immunoprecipitation (ChIP)

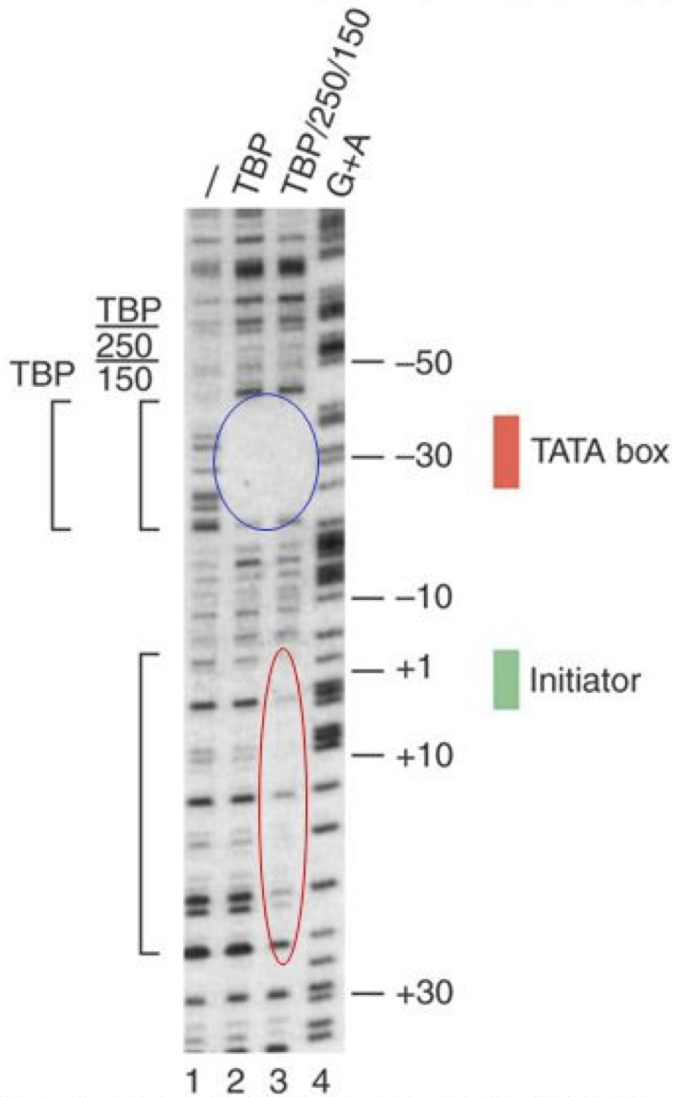
- Functional readout of Protein-DNA Interactions
  - Reporter Assays
  - *In vitro* transcription

DNaseI Footprinting



DNaseI Footprinting of hsp70 promoter

$TAF_{II}250$ and $TAF_{II}150$ cooperate in binding to the initiator & DPE



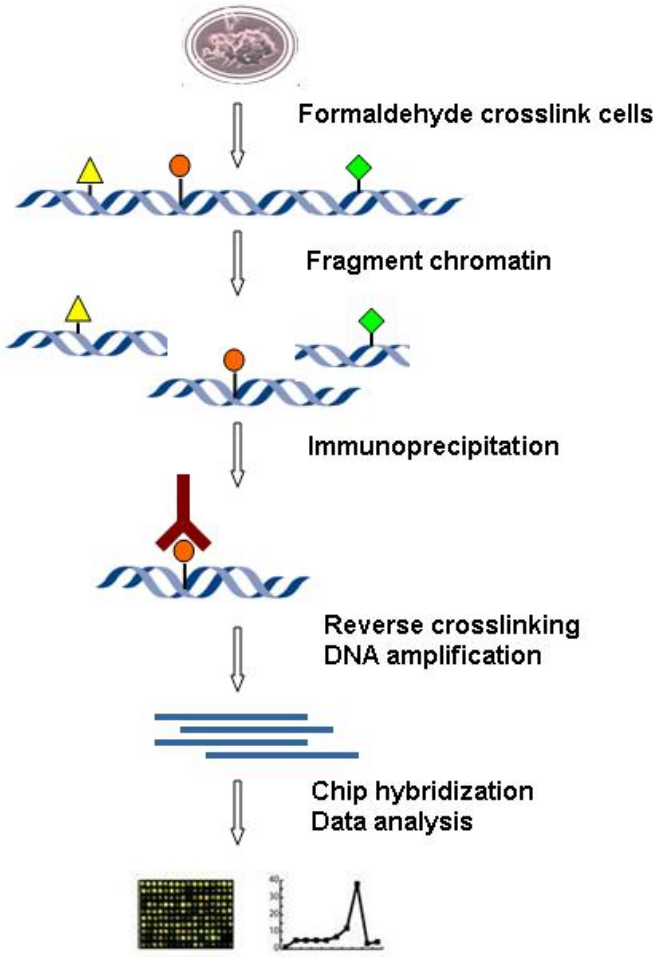
Verrijzer et al., Promoter recognition by TAFs. *Cell* 81 (30 Jun 1995) p. 1117, f. 2c.
Reprinted by permission of Elsevier Science.

Different approaches to study Protein-DNA interactions

- Physical Interaction
  - Electrophoretic Mobility Shift Assay (EMSA)
  - DNA Footprinting
  - Chromatin Immunoprecipitation (ChIP)

- Functional readout of Protein-DNA Interactions
  - Reporter Assays
  - *In vitro* transcription

Chromatin Immunoprecipitation



Chromatin Immunoprecipitation

Overview:

Strengths

•Allows you to detect transcription factor binding at specific sites within chromatin in vivo in cells or tissues.

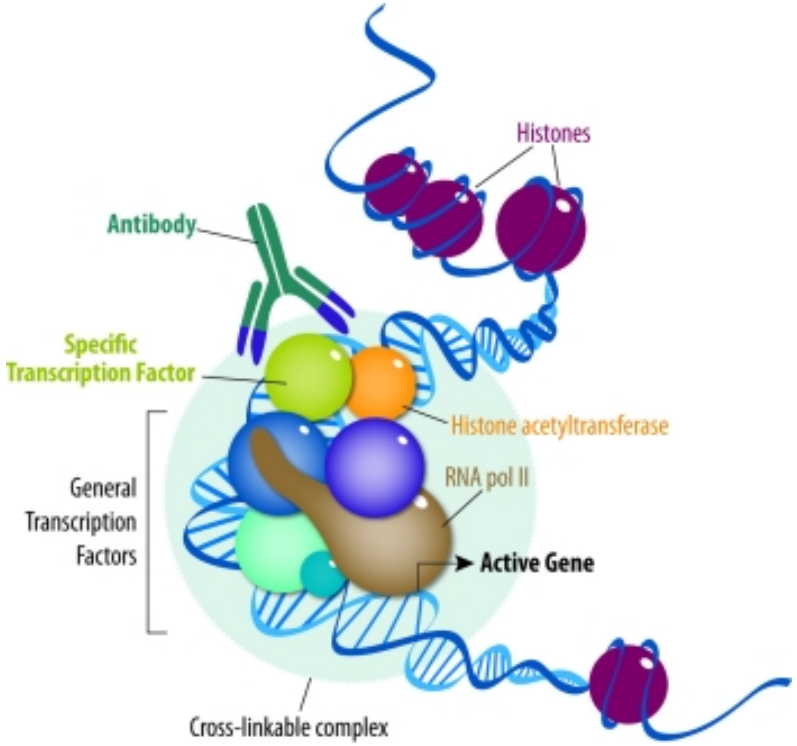
•Detection by PCR (qPCR) is very sensitive.

Weaknesses

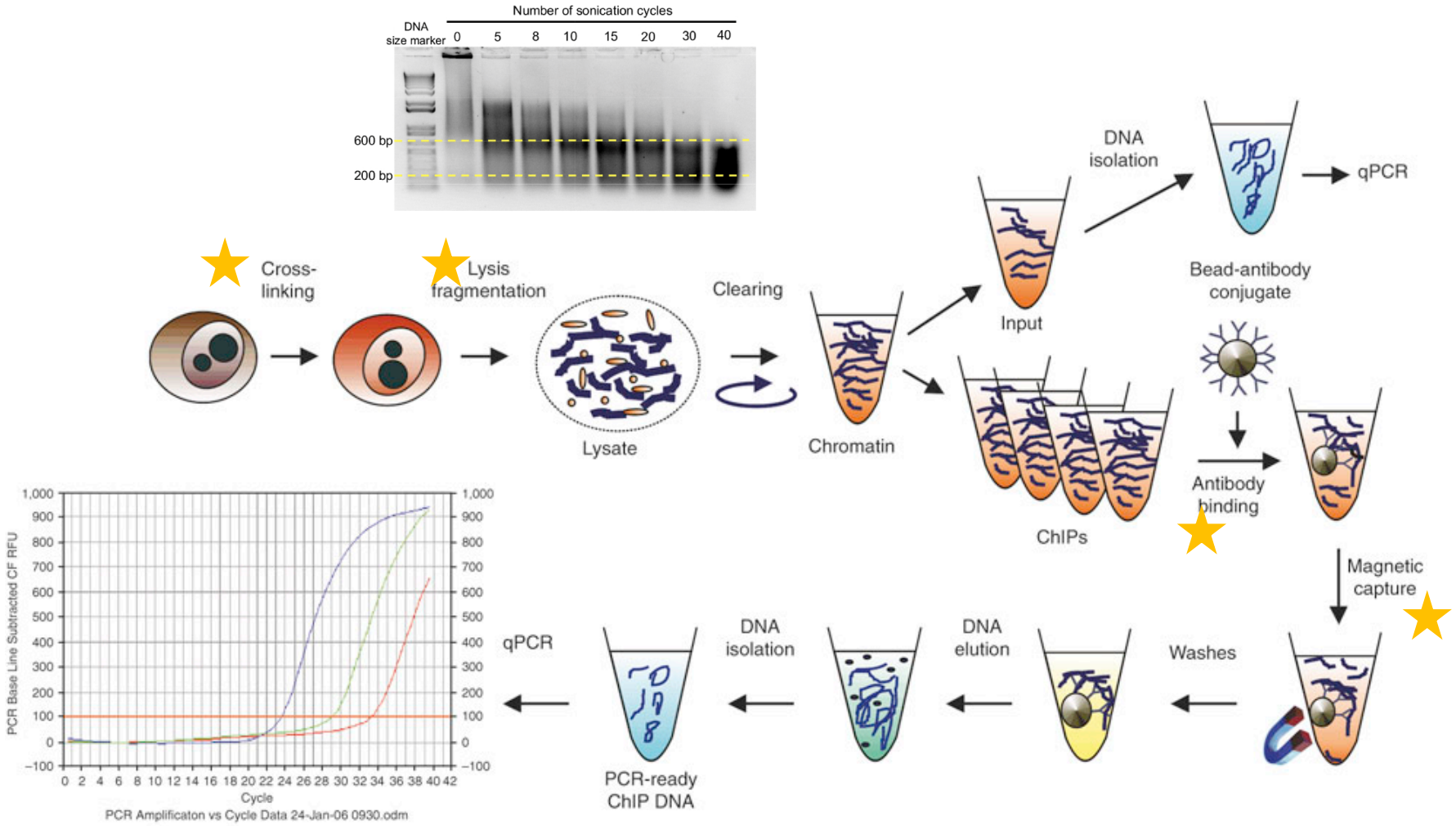
→ Requires long training and optimization steps

→ Requires very good antibodies (CHIP-grade)

→ does not exactly tell you where on DNA protein is binding.



ChIP Steps & Optimization



ChIP Controls

ChIP controls

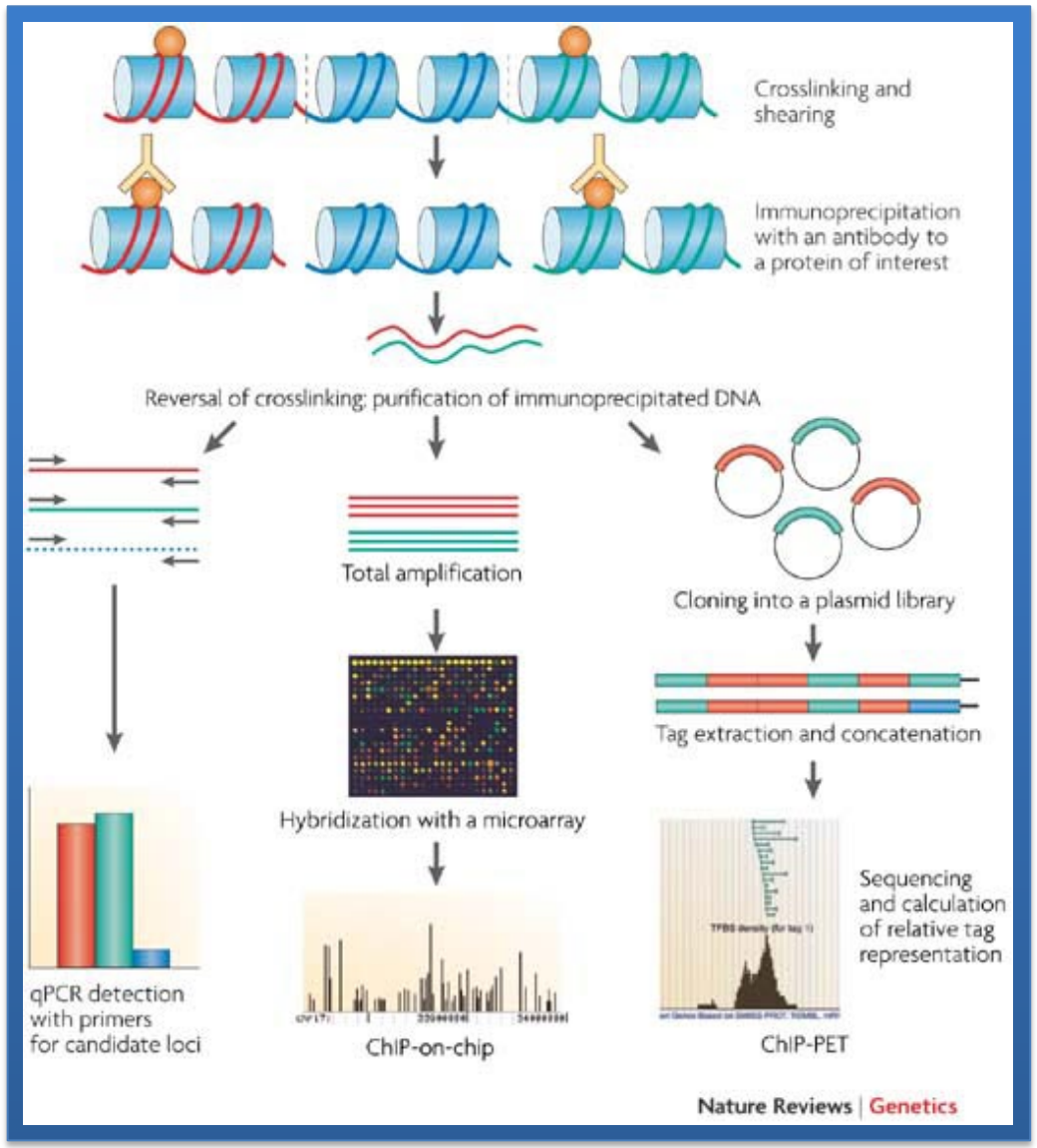
- PolII or histone marks antibodies can be useful if you are unsure about your antibody (positive control for ChIP technique)
- Controls for genes that have previously been shown to be bound by factor if interest (positive control for antibody)
- Controls for unrelated genomic regions that should not bind factor of interest (negative control)
- Normal IgG or pre-immune IgG (negative control for IP)

PCR controls

- Negative PCR controls as usual
- Serial dilutions of input material to calculate reference curve
- Dissociation curve to validate primers

- BREAK

Global Screening Methods



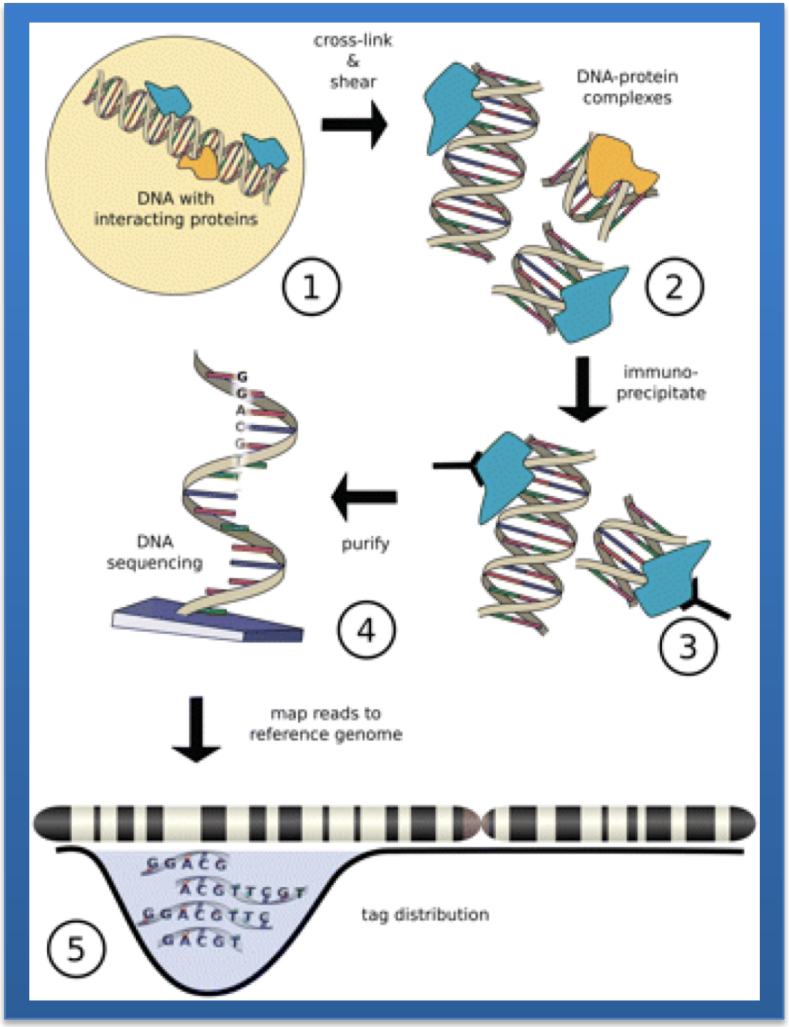
1. ChIP on Chip

Perform regular ChIP then hybridize DNA to to a tiled genomic array, or promoter array.

Advantages: global approach

Disadvantages: expensive, cross-hybridization issues, coverage is limited by the tiling array (oligos are pre-selected).

Global Screening Methods

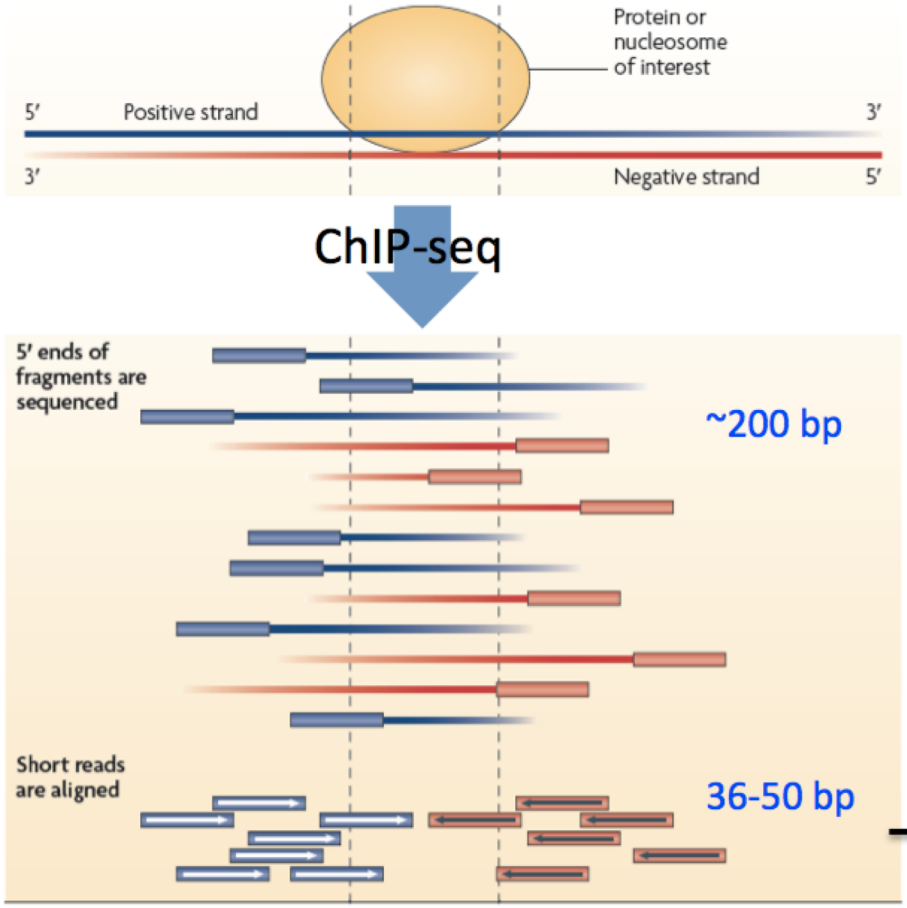


2. ChIP-seq

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

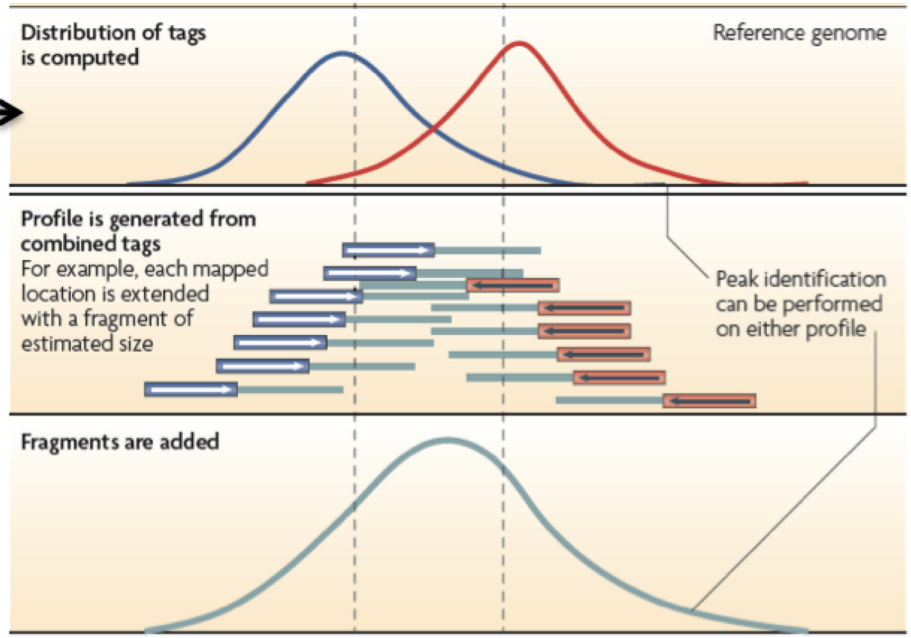
Advantages: little material required, higher resolution, fully open end approach, spatial resolution, less artifacts due to PCR amplification, possibility to multiplex

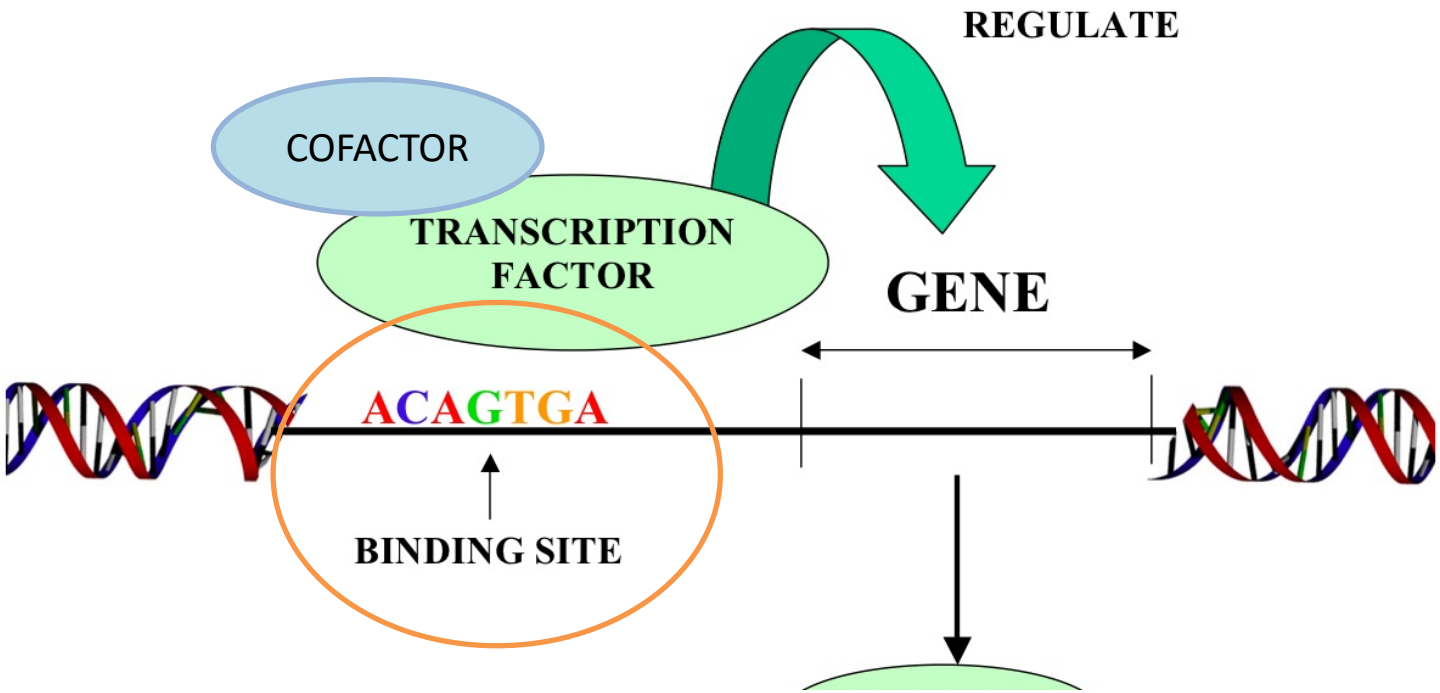
Disadvantages: expensive (particularly if controls included), need large computer storage capacity, requires complex bioinformatics analysis.



Control sample: "Input" or "IgG"

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

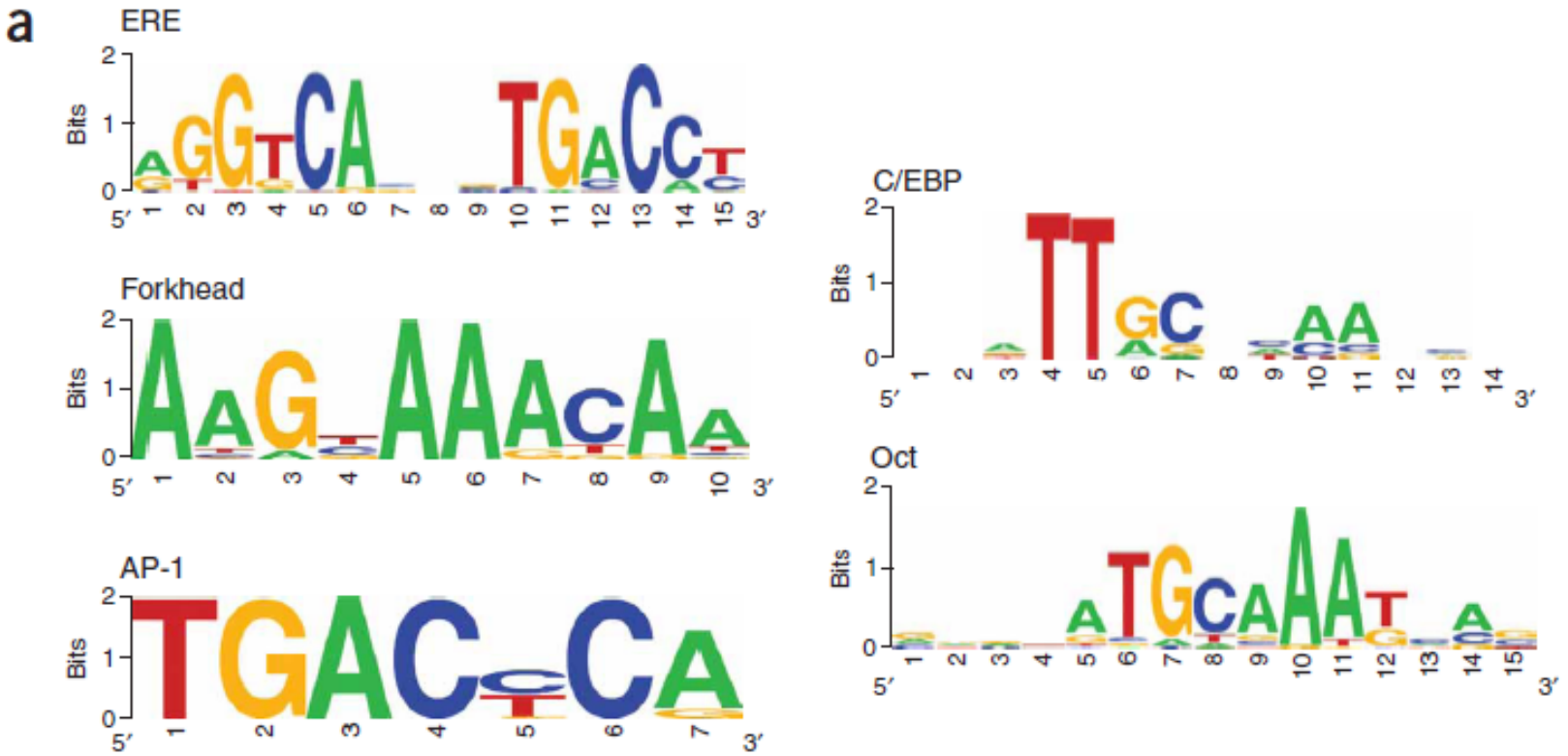



REGULATE
COFACTOR
TRANSCRIPTION FACTOR
GENE
ACAGTGA
BINDING SITE

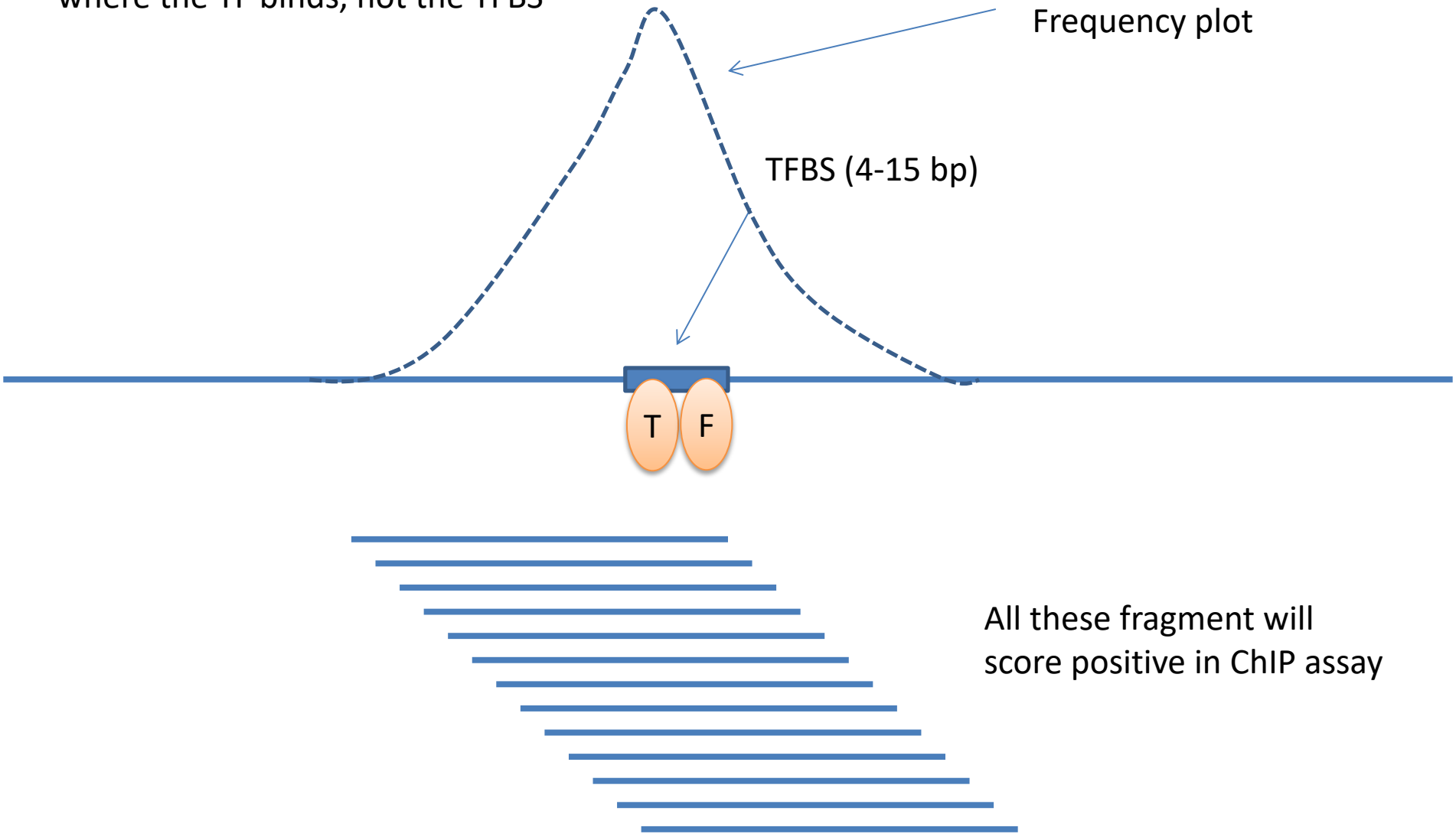
Motif Enrichment Analysis

In a defined interval around binding peaks(i.e +/-500bp), algorithms exist to find unbiased overrepresented motifs, or known motifs based on **positional weight matrices**.

Examples:



ChIp-Seq analysis identifies a region where the TF binds, not the TFBS



All these fragment will score positive in ChIP assay

DAP-seq – Add slide

Identification of TFs responsible for cofactor recruitment through motif analysis

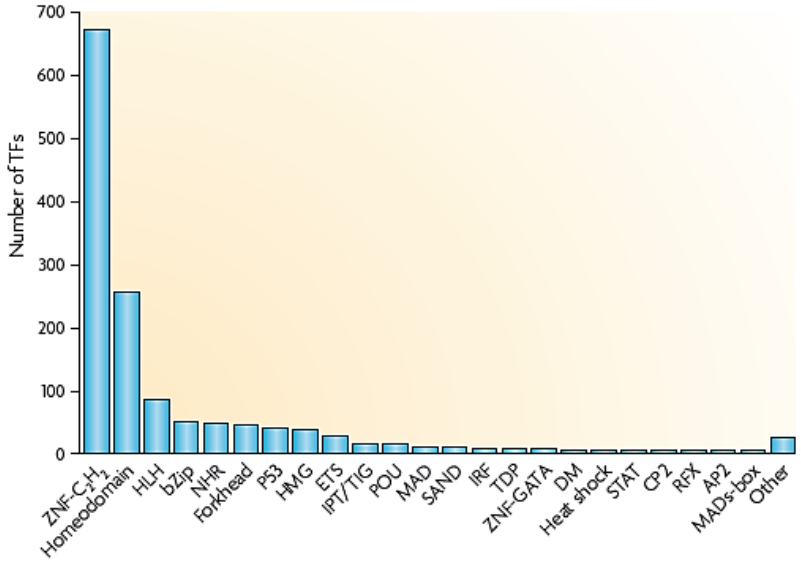


Figure 2 | **Transcription factors classified by DNA-binding domain.** Transcription factors (TFs) were classified into families according to their DNA-binding domain composition. InterPro parent–child relationships between DNA-binding domains were used as the basis for TF family definition (Supplementary information S1 (PDF)). TFs with multiple DNA-binding domains were classified in each of their respective families. Families with less than five members were classified as 'other'.

TRANSFAC© Family	TFBS Designation	TFBS	TF*	Gene Symbol‡
ETS-Type				
	PEA3		PEA3	
	ELF1		ELF-1	
	ETS2		ETS2	
	ETS1		c-Ets-1	
	c-ETS 1-68§			
	GABP			
			GABP-α‖	
			GABP-β‖	
	TEL2		Tel-2	
	NERF		NERF-1a	
Rel/Ankyrin				
	NFKB			
			RelA‖	
			p50‖	
	NFKB p65		RelA	
	NFKB p50		p50	
Interferon-Regulating Factors				
	ISRE		ISGF-3	
Runt-Homology Domain				
	AML1		AML1	

Motif analysis is just the beginning… need to validate TF binding by regular ChIP, and to study functional effect by reporter assays, siRNA studies

ChIP-seq Data Analysis

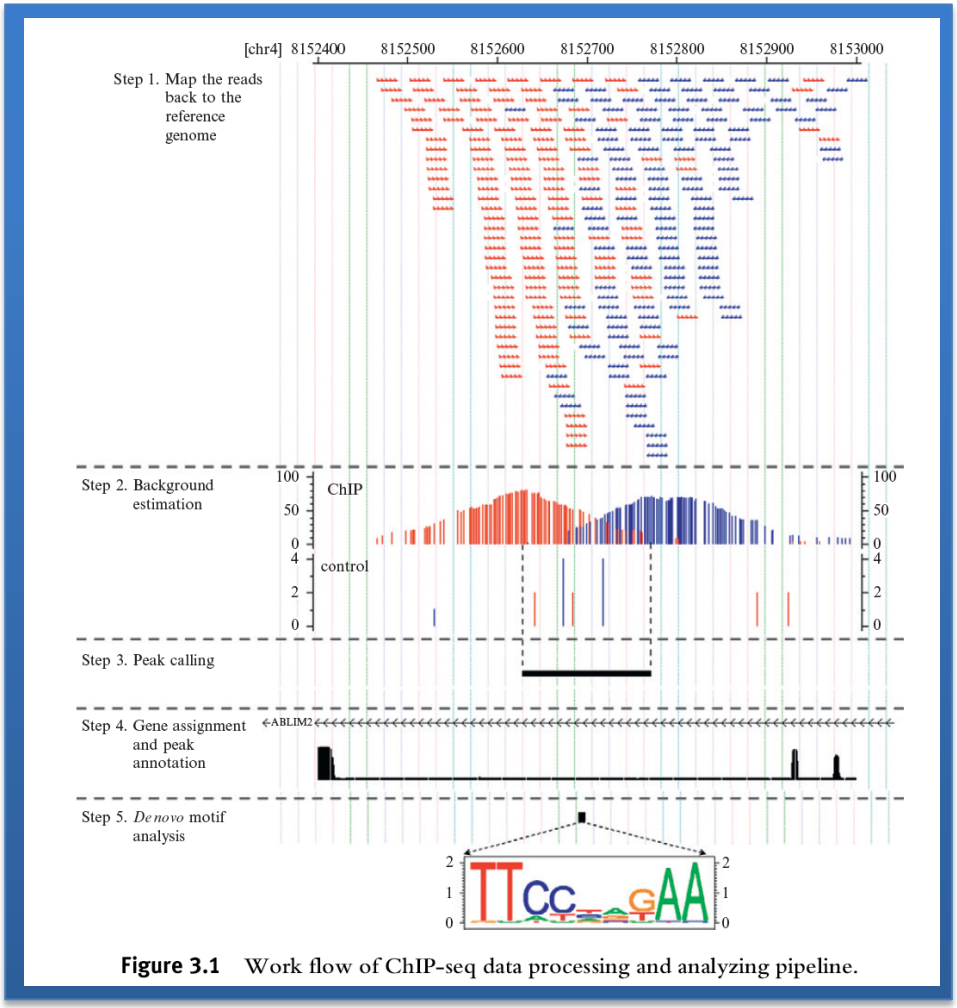
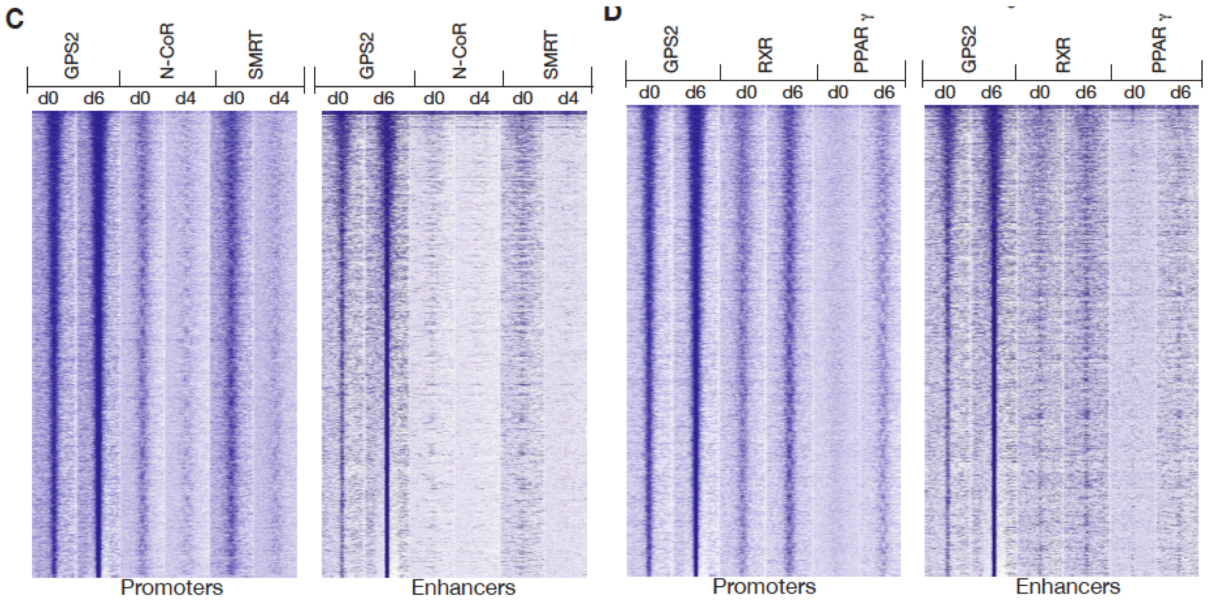


Figure 3.1 Work flow of ChIP-seq data processing and analyzing pipeline.

Integration with other datasets

ChIP-seq Data Secondary Analysis



Overlapping with available NCoR/SMRT and PPARg/RXR
ChIP-seq data (Lefterova et al., 2008; Mikkelsen et al., 2010; Raghav et al., 2012)

ChIP-seq Data Secondary Analysis

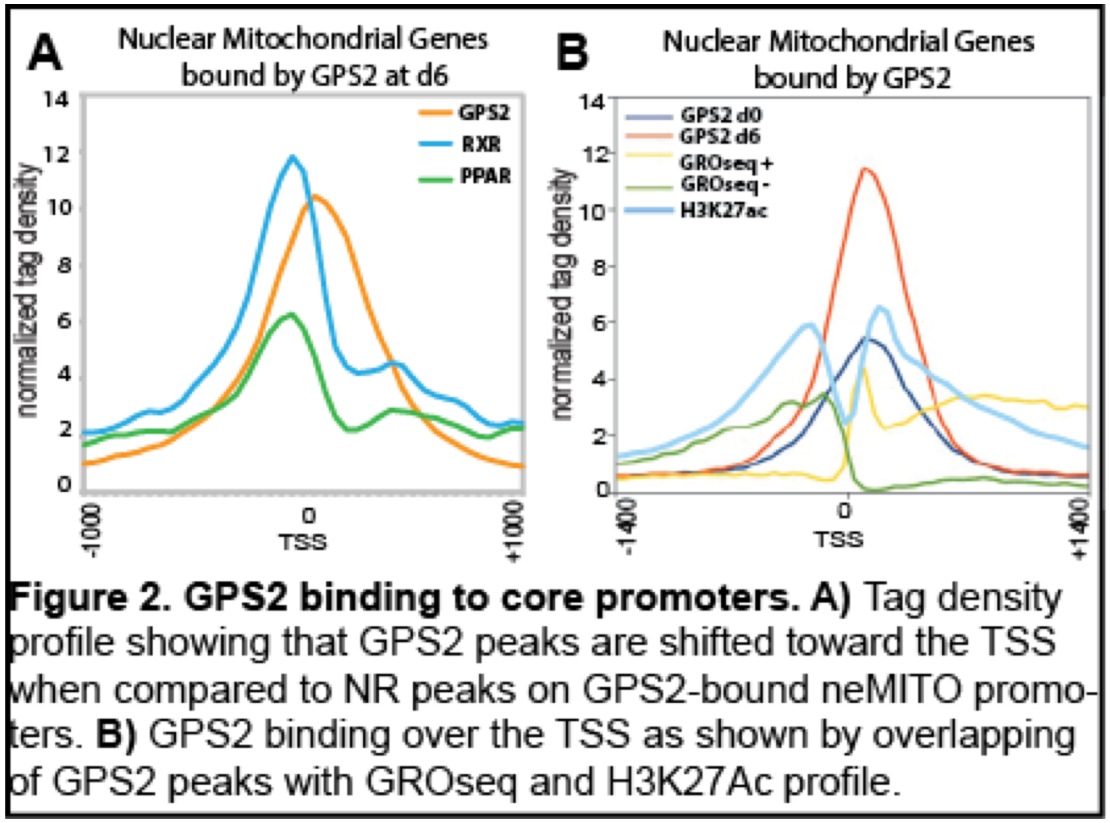
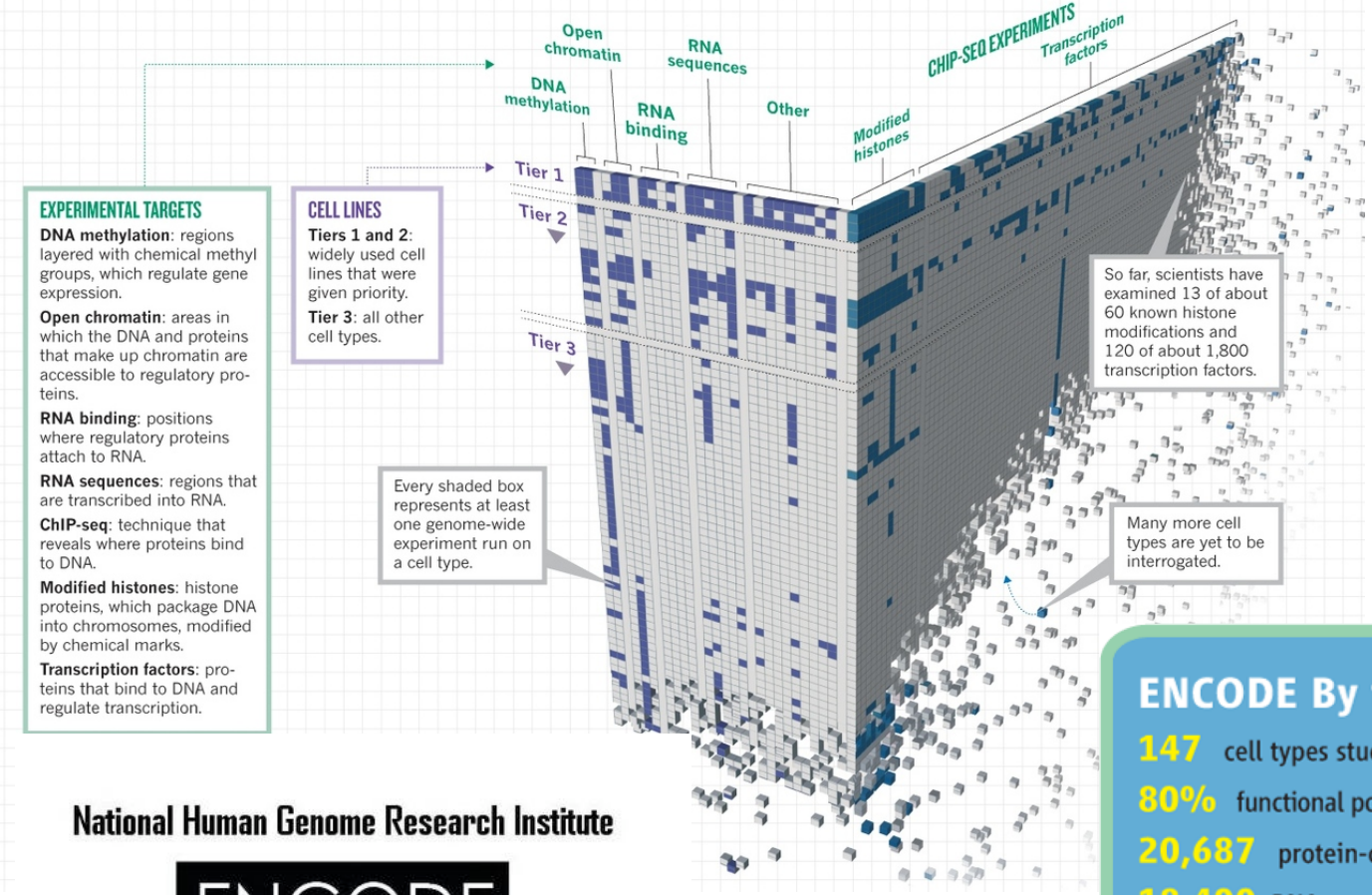


Figure 2. GPS2 binding to core promoters. A) Tag density profile showing that GPS2 peaks are shifted toward the TSS when compared to NR peaks on GPS2-bound neMITO promoters. **B)** GPS2 binding over the TSS as shown by overlapping of GPS2 peaks with GROseq and H3K27Ac profile.

Is there a lot of data available for this type of analysis?

MAKING A GENOME MANUAL

Scientists in the Encyclopedia of DNA Elements Consortium have applied 24 experiment types (across) to more than 150 cell lines (down) to assign functions to as many DNA regions as possible — but the project is still far from complete.



EXPERIMENTAL TARGETS

DNA methylation: regions layered with chemical methyl groups, which regulate gene expression.

Open chromatin: areas in which the DNA and proteins that make up chromatin are accessible to regulatory proteins.

RNA binding: positions where regulatory proteins attach to RNA.

RNA sequences: regions that are transcribed into RNA.

ChIP-seq: technique that reveals where proteins bind to DNA.

Modified histones: histone proteins, which package DNA into chromosomes, modified by chemical marks.

Transcription factors: proteins that bind to DNA and regulate transcription.

CELL LINES

Tiers 1 and 2: widely used cell lines that were given priority.

Tier 3: all other cell types.

National Human Genome Research Institute

ENCODE

ENCODE By the Numbers

147 cell types studied

80% functional portion of human genome

20,687 protein-coding genes

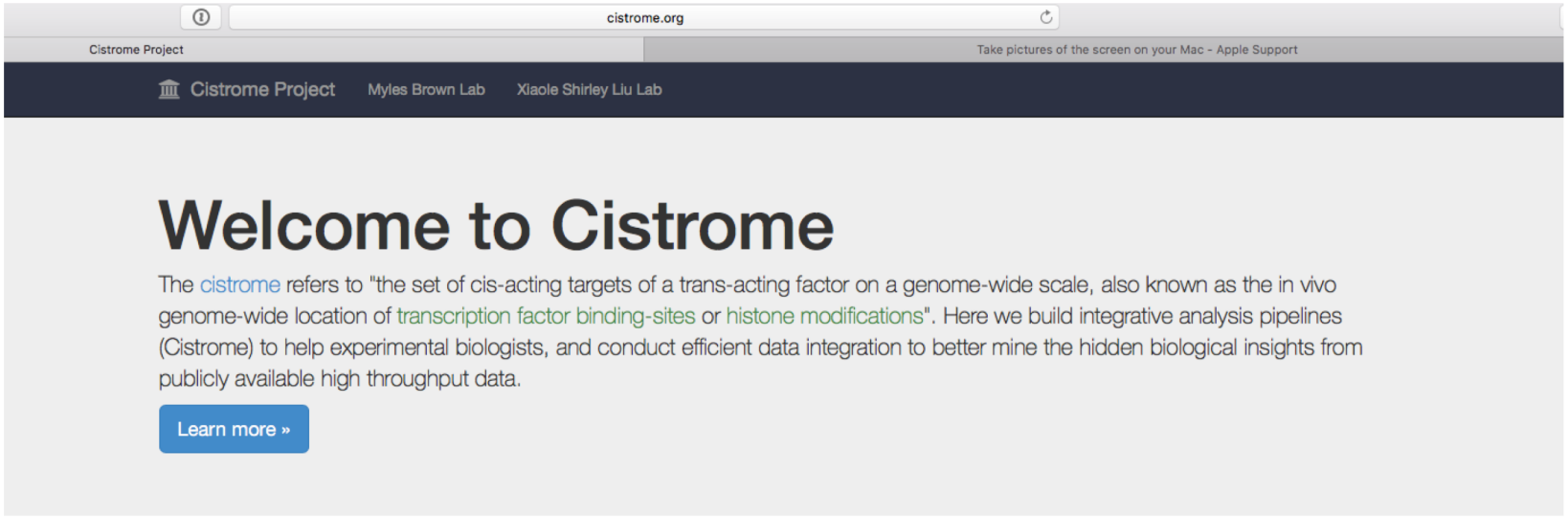
18,400 RNA genes

1640 data sets

30 papers published this week

442 researchers

$288 million funding for pilot, technology, model organism, and current project



Cistrome Analysis Pipeline

An integrative and reproducible bioinformatics data analysis platform based on *Galaxy* open source framework. Besides standard Galaxy functions, Cistrome has 29 ChIP-chip- and ChIP-seq-specific tools in three major categories, from preliminary peak calling and correlation analyses to downstream genome feature association, gene expression analyses, and motif discovery.

Visit site »

Cistrome Data Browser (Beta version)

A new portal to browser public ChIP-seq and DNase-seq datasets. Besides providing a comprehensive knowledgebase of all of the publicly available ChIP-Seq and DNase-Seq data in mouse and human, it also provides functions to analysis and visualize these datasets.

Visit site »

Cistrome Cancer (Beta Version)

A comprehensive resource for predicted transcription factor (TF) targets and enhancer profiles in cancers. The prediction was from integrative analysis of TCGA expression profiles and public ChIP-seq profiles.

Visit site »

CRISPR-DO

This application focus on the whole genome sgRNA design in human and mouse, with accessing both the efficiency and the specificity score. It also have the epigenome browser as a visualization tool for users to identify each of the sgRNA with genome features overlapping like DHS, SNP.

Visit site »

Sequence Scan for CRISPR

A new sequence model for predicting sgRNA efficiency for CRISPR knockout or CRISPRi/a by systematically assessing the DNA sequence features that contribute to single guide RNA (sgRNA) efficiency in CRISPR-based screens.

Visit site »

Binding and Expression Target Analysis

Binding and Expression Target Analysis (BETA) is a software package that integrates ChIP-seq of transcription factors or chromatin regulators with differential gene expression data to infer direct target genes

Visit site »

NCBI

Gene Expression Omnibus

NCBI will be testing https on public web servers from 1:00-4:00 PM EDT (17:00-20:00 UTC) on Wednesday, October 19. You may experience problems with NCBI services during that time. Please plan accordingly.
Read more.

Series GSE57779 Query DataSets for GSE57779

Status	Public on Jul 14, 2014
Title	Promoter-specific recruitment of PPARG in adipocytes depends on GPS2-dependent stabilization of histone demethylase KDM4A/JMJD2
Organism	Mus musculus
Experiment type	Genome binding/occupancy profiling by high throughput sequencing
Summary	This SuperSeries is composed of the SubSeries listed below.
Overall design	Refer to individual Series
Citation(s)	Cardamone MD, Tanasa B, Chan M, Cederquist CT et al. GPS2/KDM4A pioneering activity regulates promoter-specific recruitment of PPARγ. *Cell Rep* 2014 Jul 10;8(1):163-76. PMID: 24953653

Submission date	May 19, 2014
Last update date	Oct 11, 2016
Contact name	Valentina Perissi
E-mail	vperissi@acs.bu.edu
Phone	617 638 4115
Fax	617 638 5339
Organization name	Boston University School of Medicine
Department	Department of Biochemistry
Street address	72 E. Concord St
City	Boston
State/province	MA
ZIP/Postal code	02118
Country	USA

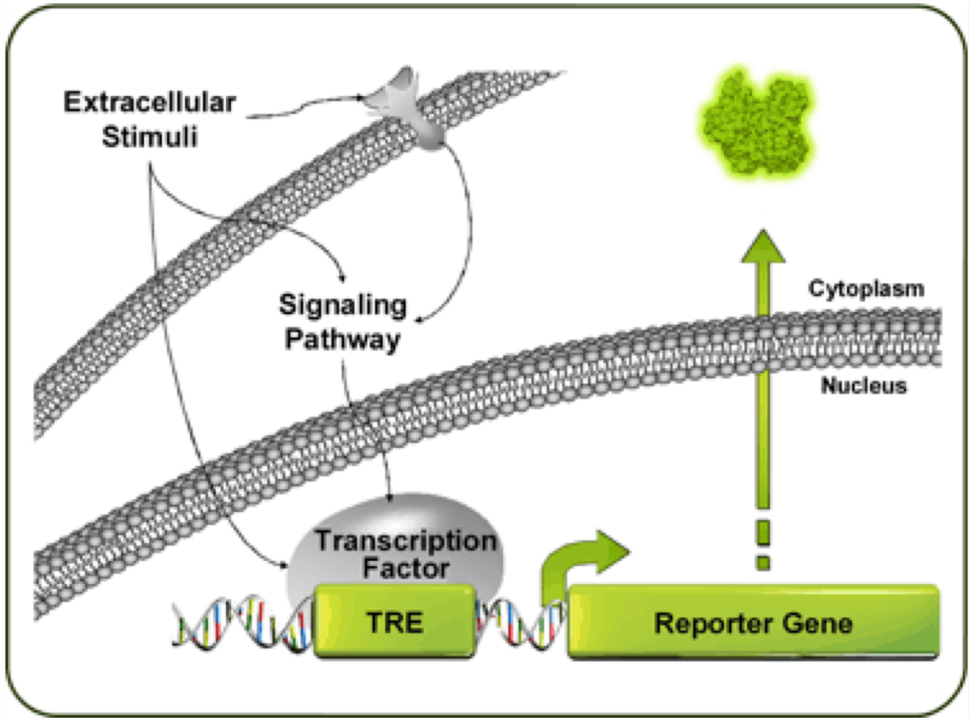
Platforms (2)	GPL9250 Illumina Genome Analyzer II (Mus musculus)
	GPL13112 Illumina HiSeq 2000 (Mus musculus)
Samples (4)	GSM1388416 GPS2_day0
Less...	GSM1388417 GPS2_day6
	GSM1388418 KDM4A_siCTL
	GSM1388419 KDM4A_siGPS2

Different approaches to study Protein-DNA interactions

- Physical Interaction
  - Electrophoretic Mobility Shift Assay (EMSA)
  - DNA Footprinting
  - Chromatin Immunoprecipitation (ChIP)

- Functional readout of Protein-DNA Interactions
  - Reporter Assays
  - *In vitro* transcription

What is a Reporter Assay?



Reporter Assays

Strengths
High throughput
Can measure function of mutations in promoters
Large dynamic range

Many reporters possible
• GFP
• b-galactosidase
• CAT (chloramphenical acetyl transferase)
• Luciferase (firefly, renilla)

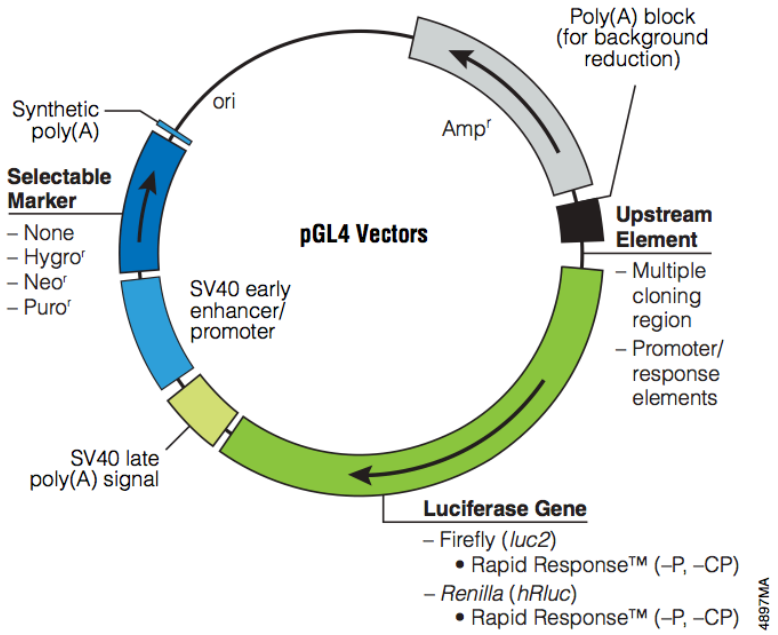
Can be used for in vivo/in cell monitoring.

Weaknesses
• uses exogenous DNA, not chromatin
• Gene dosage artifacts are possible
• relies on transfection, not easy for all cells

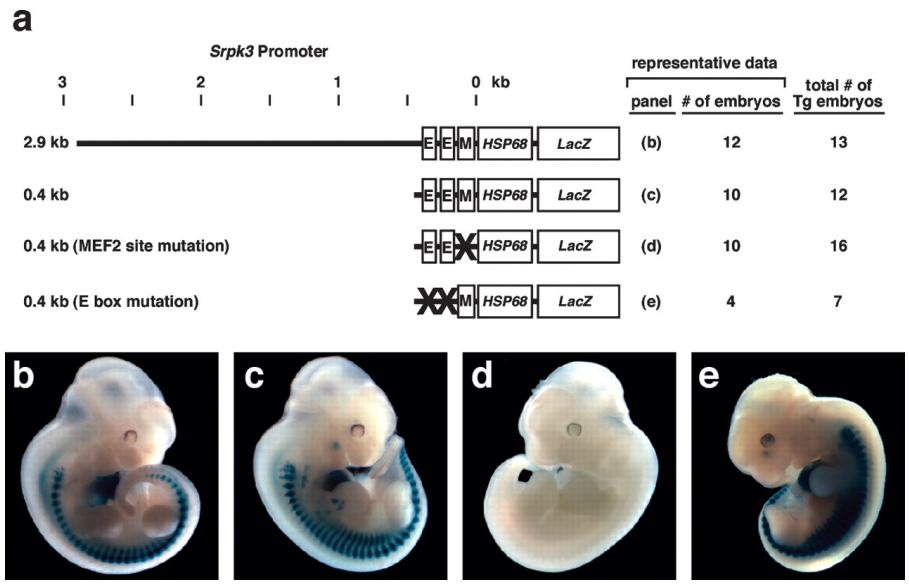
Luciferase Reporters—pGL Family

The pGL4 Vector family includes:

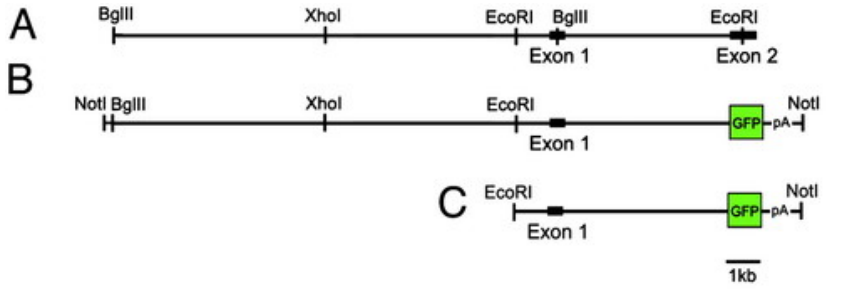
- Basic vectors with no promoter that contain a multiple cloning region for cloning a promoter of choice
- Vectors containing a minimal promoter
- Vectors containing response elements and a minimal promoter
- Promoter-containing vectors that can be used as expression controls or as co-reporter vectors

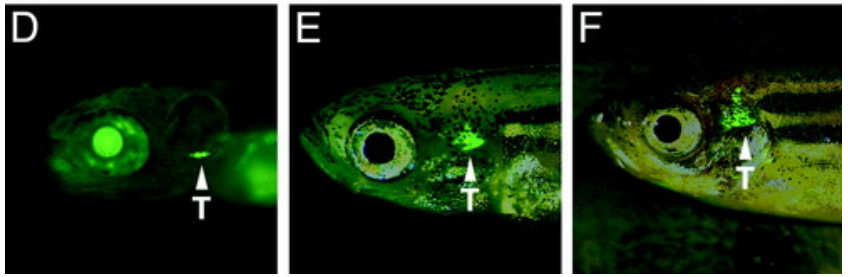


In vivo studies of regulatory regions: LacZ and GFP reporter systems



Nakagawa et al., G&D 2005



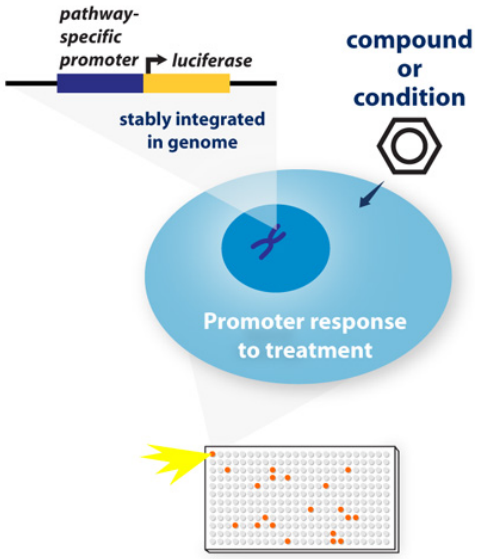


Langenau et al.PNAS 2004

What else can be done with Reporter Assays? Variations on Basic Experiment

- Overexpress exogenous transcription factors
- Overexpress dominant negative transcription factors (variants of normal factor that functions to inhibit the WT allele)

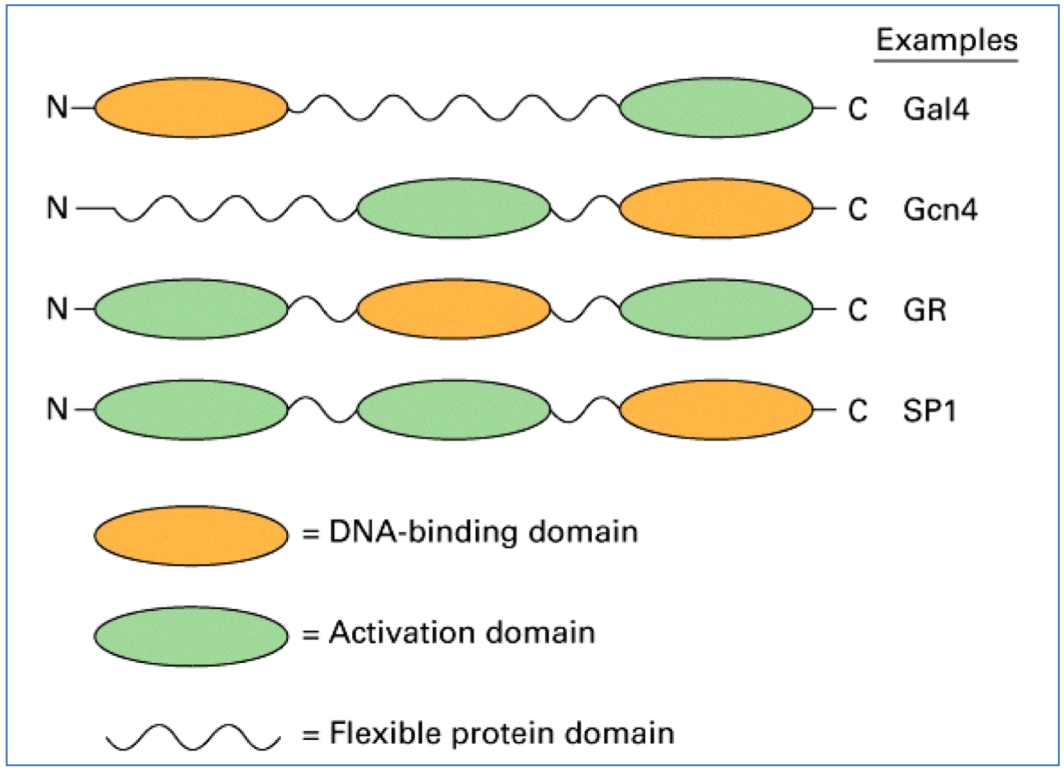
  i.e. Mutant lacking the transcriptional activation domain, the DNA binding….
- High throughput screenings

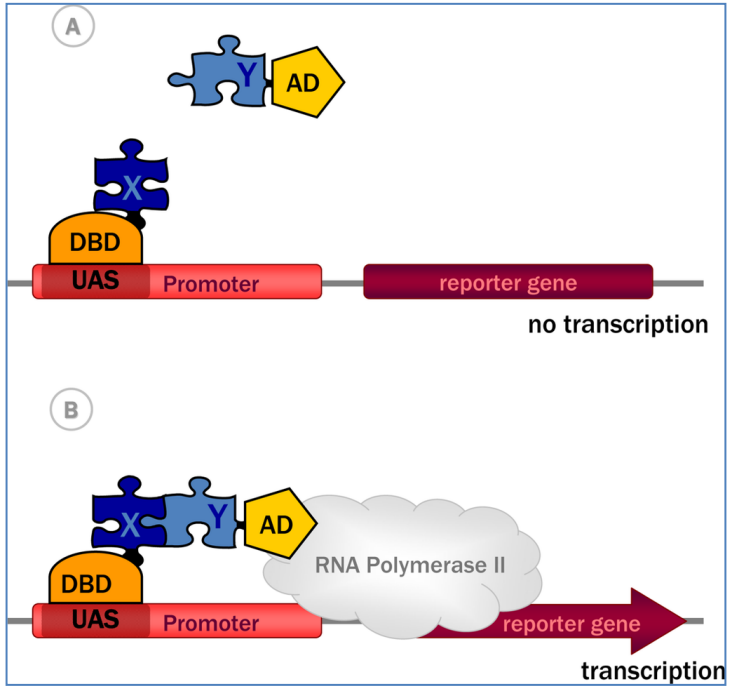


- Artificial transcription factors (Gal4 DNA binding domain, VP16 activation domain) to test corepressors and coactivators

Reporter Assays to measure protein-protein interaction

The structure of transcription factors is modular and modules (domains) are experimentally interchangeable





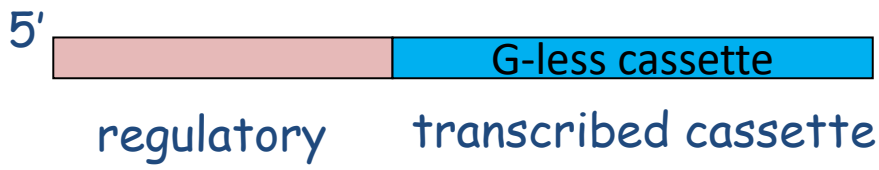
Bruckner et al., *Int. J. Mol. Sci.* **2009**

Different approaches to study Protein-DNA interactions

- Physical Interaction
  - Electrophoretic Mobility Shift Assay (EMSA)
  - DNA Footprinting
  - Chromatin Immunoprecipitation (ChIP)

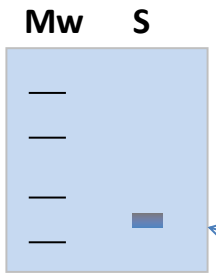
- Functional readout of Protein-DNA Interactions
  - Reporter Assays
  - *In vitro* transcription

In vitro transcription assay



Add ATP, CTP, UTP + Nuclear Extracts
(or purified components of PIC + Pol II)

purify



DNA analysed on high-resolution gels

In vitro transcribed RNA