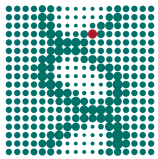
cis regulatory elements:

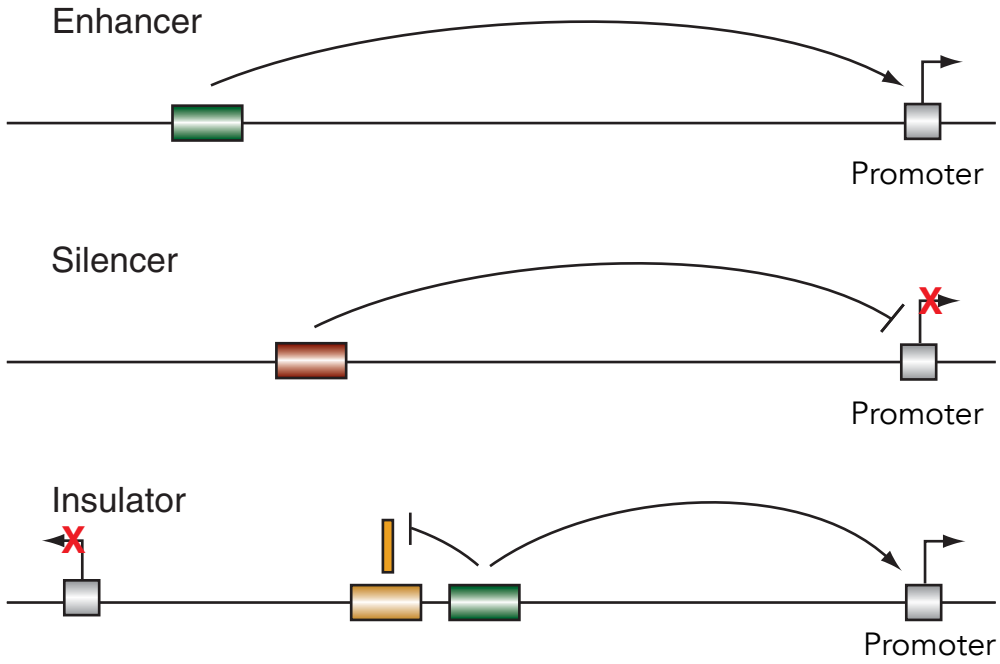
How to find them: Past, presence & future

Sebastiaan Meijsing
Transcriptional regulation group
meijsing@molgen.mpg.de



Max Planck Institute
for Molecular Genetics

Cis-regulatory elements: Different flavors



Maston *et al.*, Annu. Rev. Genom. Human Gent. (2006)

Today we mostly focus on cis regulatory elements that act as enhancers

Outline

- Techniques to find cis regulatory elements
  (& transcription factor binding sites)

Function:

- Linking binding and regulation

Identification of cis regulatory elements:

Approach1: - Conservation analysis (evolution)

Approach 2: - Biophysical approaches (e.g. binding)

Approach 3: - Chromatin-based methods

Approach 4: Functional approaches (reporter genes & mutagenesis)

Clues from conservation analysis

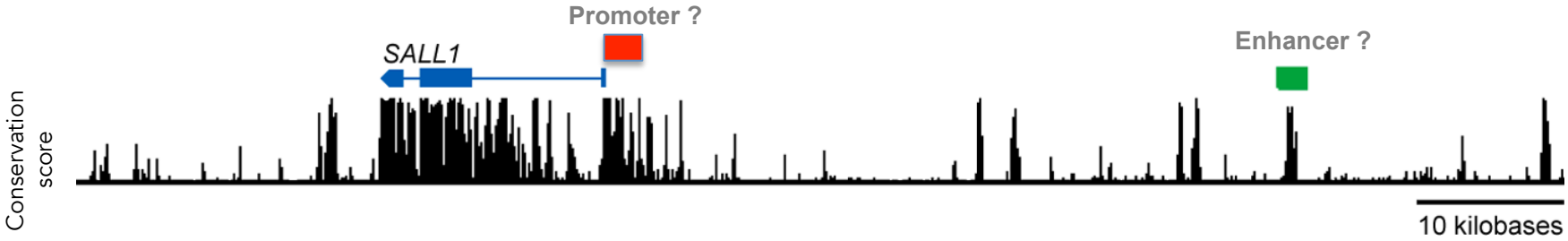
Conservation analysis can be used to identify functional elements in the genome.

↓

Typically larger stretches of sequence conserved ("phylogenetic footprint")

↓

Cis-regulatory modules (enhanceasomes)



Petersen *et al.*, Plos One (2009)

The good: Genome-wide approach (cheap…)

The bad: False positives / false negatives /overlapping functional elements

Identification of cis regulatory elements:

Approach1: - Conservation analysis (evolution)

Approach 2: - Biophysical approaches (e.g. binding)

Approach 3: - Chromatin-based methods

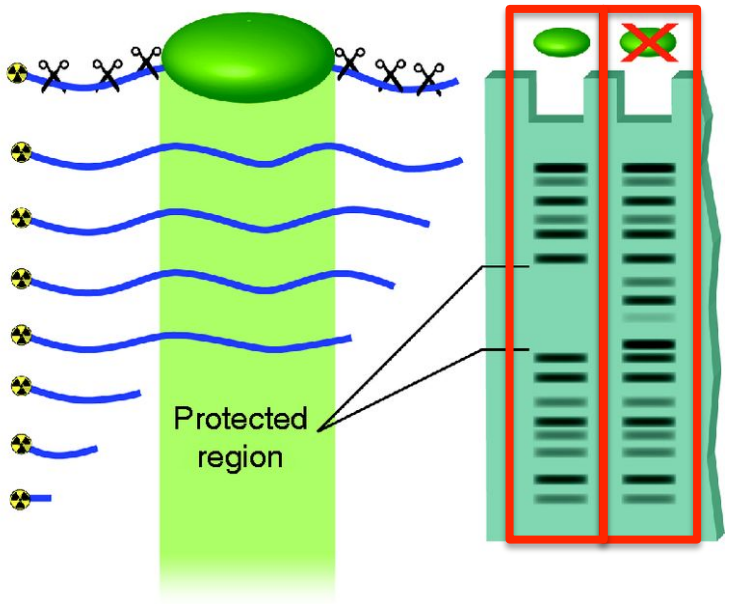
Approach 4: Functional approaches (reporter genes & mutagenesis)

Biophysical approaches: DNAse footprinting-I

DNAse footprinting (classic version):

What it tells you:
- Which Regions of a DNA-fragments are bound by protein(s) (using nuclear extracts)
- Where on the DNA your protein of interest binds (using purified proteins)

How it works:



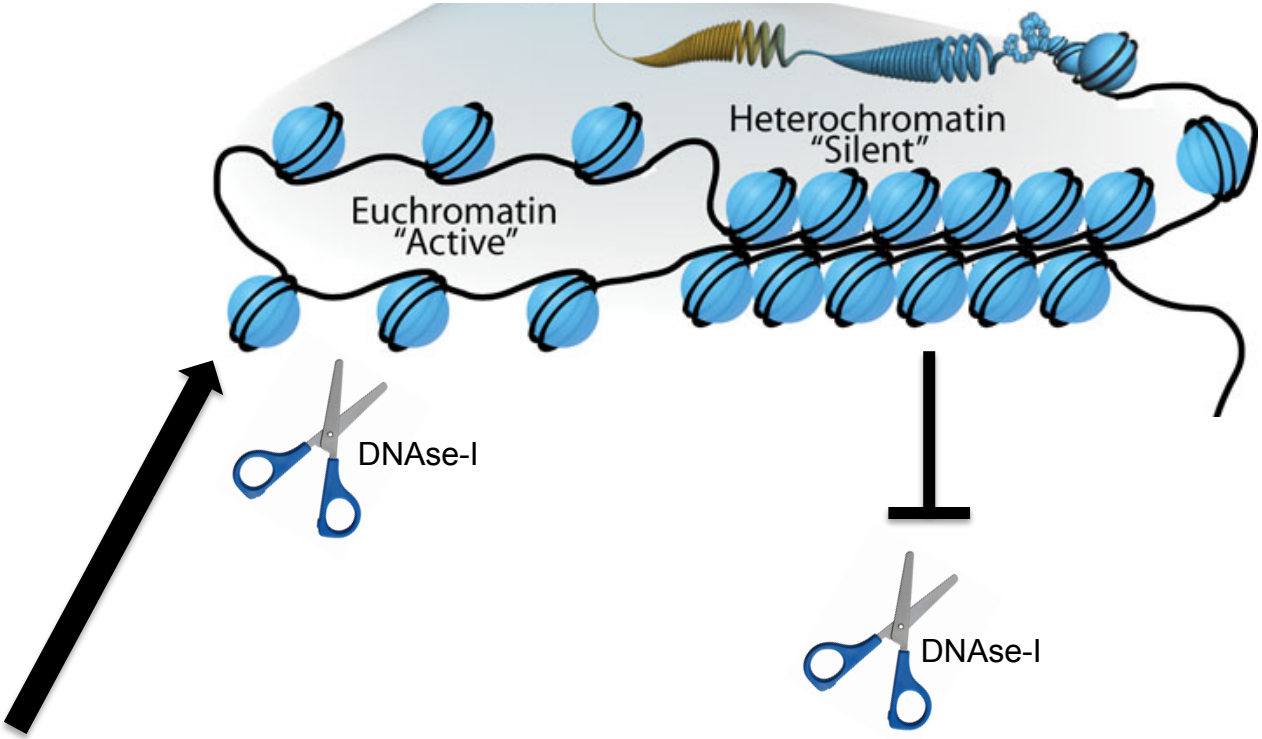
Vinckevicius and Chakravarti. *J. Mol. Endocrinol.* (2012)

The good: high resolution information for area of interest

The bad: In vitro (can bind vs actually bound in vivo? / low throughput)

Biophysical approaches: DNAse footprinting-II

Mapping DNAse-I hypersensitive regions genome-wide



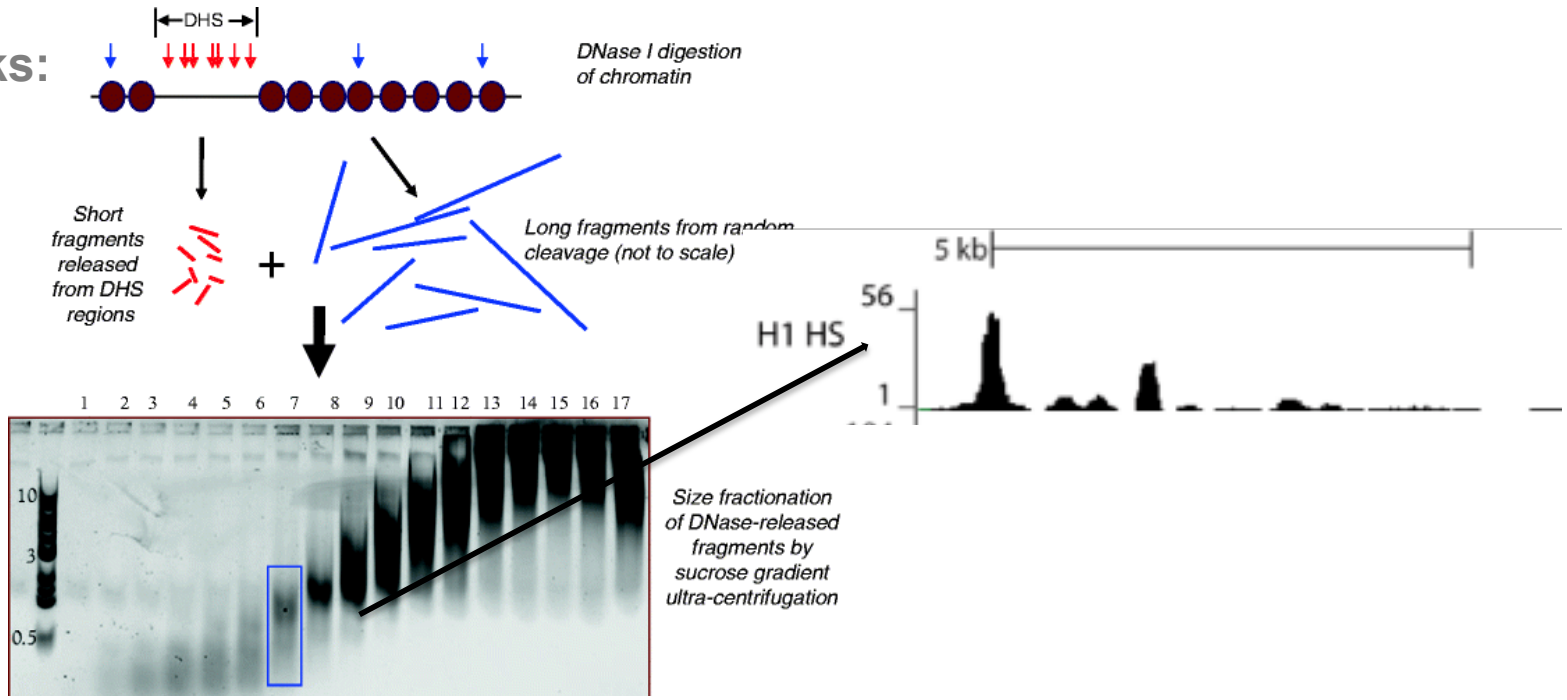
Active enhancers / promoters typically localize to DNAseI hypersensitive regions ("Open chromatin")

Biophysical approaches: DNAse footprinting-II

Mapping DNAse-I hypersensitive regions genome-wide

What it tells you: - Which Regions of the genome can be cut (implying that they are "open")
- Where enhancers & promoters are located in the genome

How it works:



Ling and Waxman. *Methods in Molecular biology.* (2013)

The good: Genome-wide approach / maps enhancers & promoters

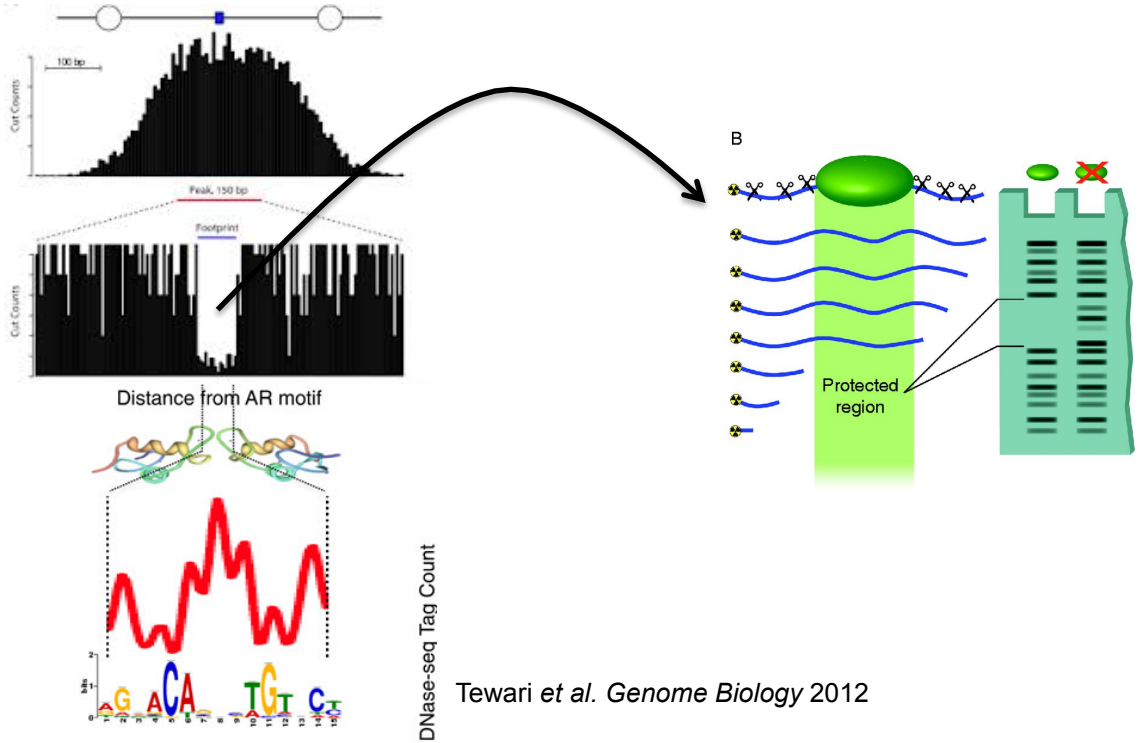
The bad: Biases in cutting preference DNAse-I / Mapping /sequencing

Biophysical approaches: DNAse footprinting-III

Mapping DNAse-I sensitivety at high resolution

What it tells you: - Which Regions of the genome can be cut (DNAse-I hypersensitive sites)
- Within such regions: Footprints of bound proteins

How it works:



Tewari *et al. Genome Biology* 2012

The good: Genome-wide approach

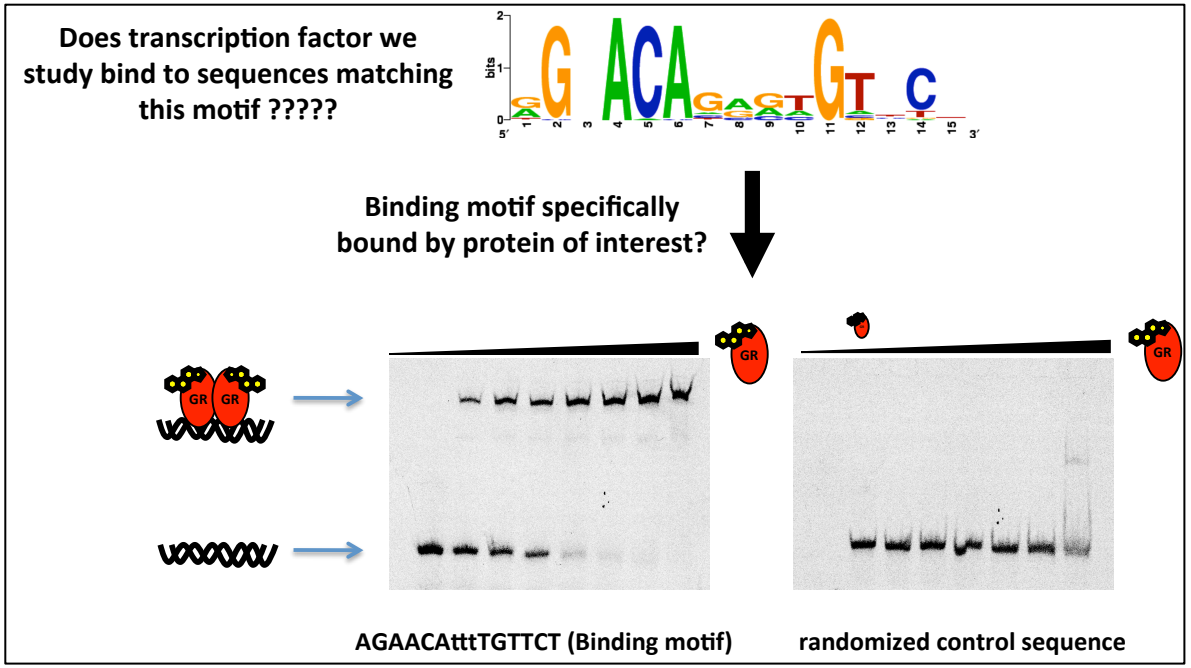
The bad: expensive / cause of protection unclear / Biases DNAse-I

Biophysical approaches: EMSAs

Electrophoretic Mobility Shift Assays (Gelshifts)

What it tells you: - Affinity of protein of interest for a particular sequence

How it works:



The good: Quantitative data / high resolution

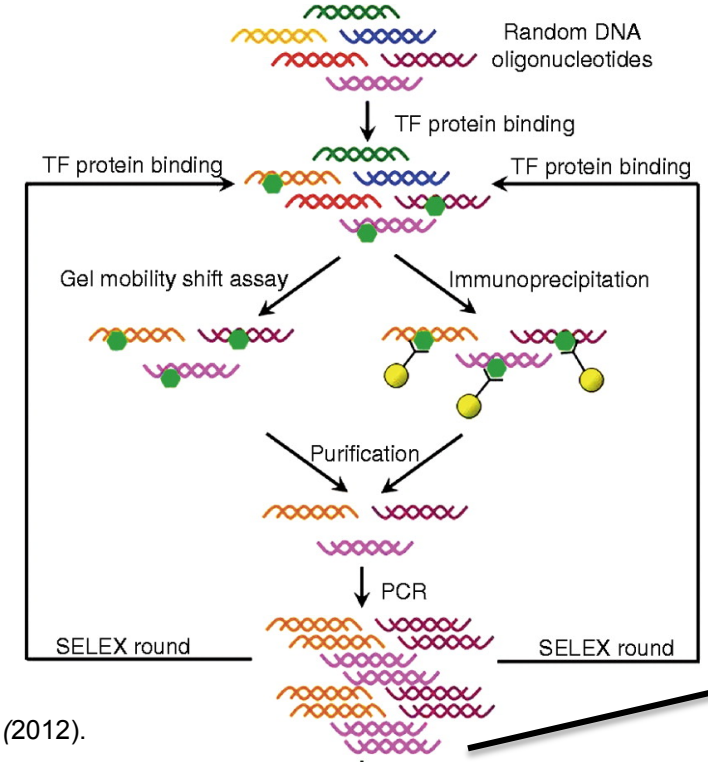
The bad: in vitro (can bind vs binds in vivo) / low throughput

Biophysical approaches: SELEX

SELEX (Systematic Evolution of Ligands by Exponential enrichment)

What it tells you: - To which DNA sequences a protein of interest can bind

How it works:



Next generation sequencing

Wang *et al. J. Endocrinol. (*2012).

The good: No prior knowledge of recognition sequence needed

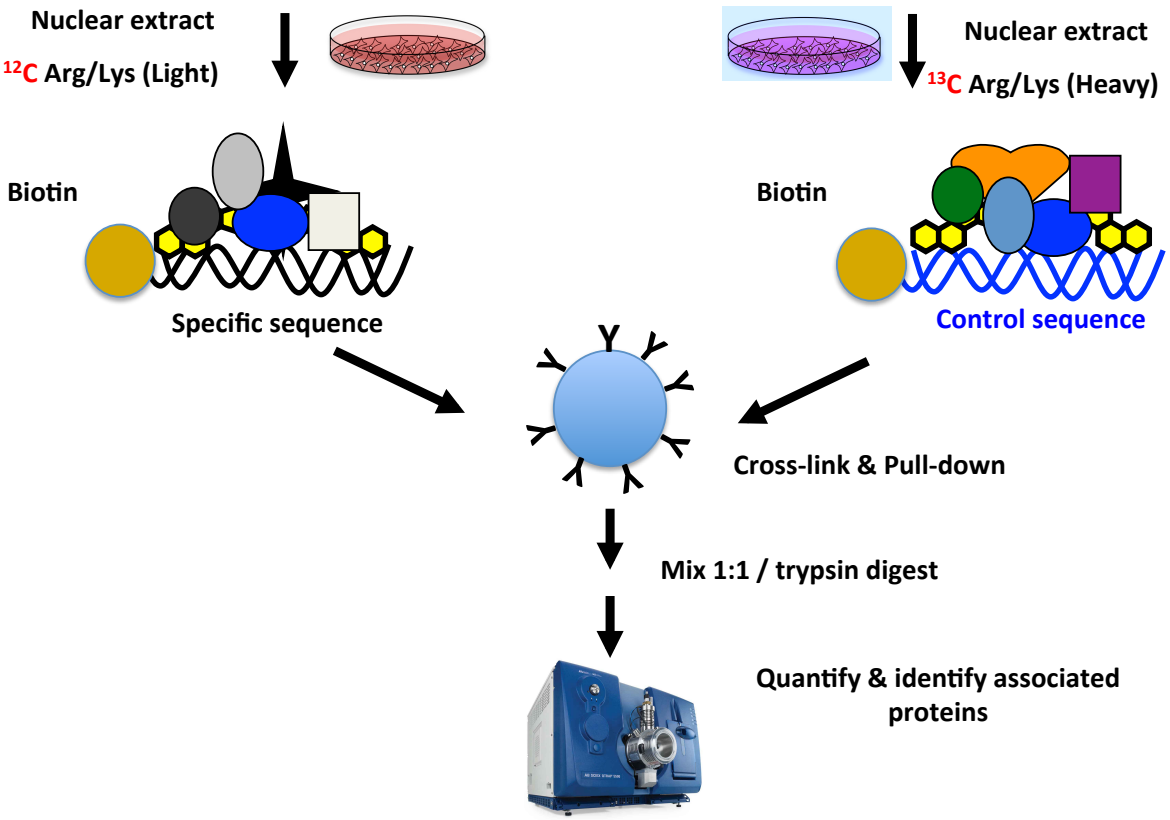
The bad: In vitro technique / bias towards high-affinity sites…

Biophysical approaches: DNA-pull down + mass-spec

DNA pull-down assays & mass-spectrometry

What it tells you: - Which proteins bind to a DNA sequences of interest

How it works:

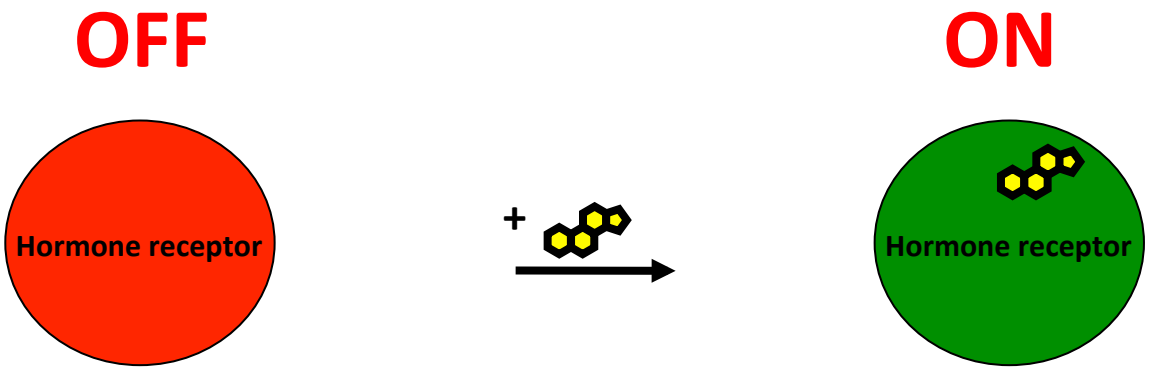


The good: No prior knowledge needed

The bad: In vitro assay

Hormone receptors:

A nice model system to study Transcription



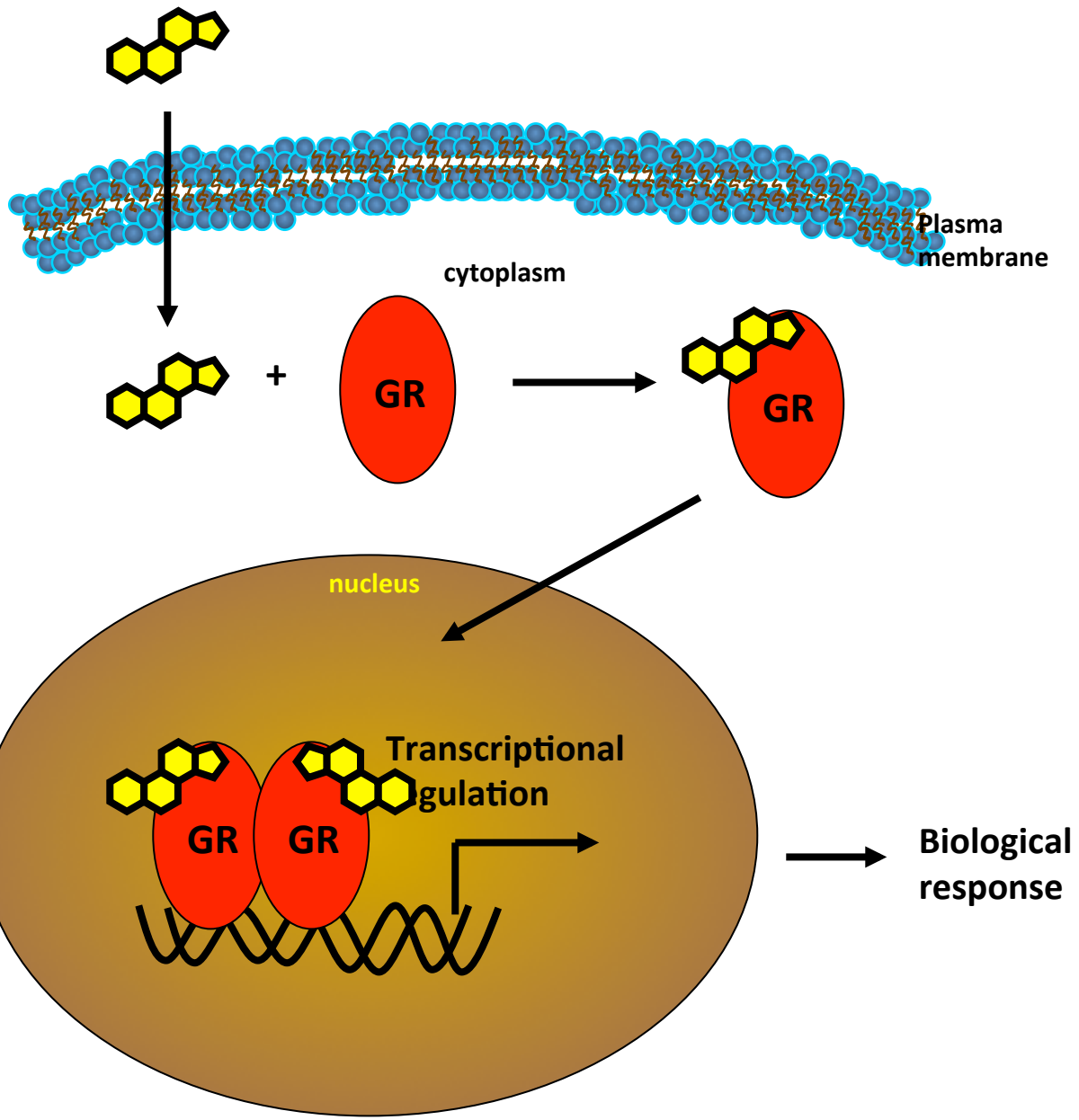
Glucocorticoid receptor signaling: Transcription

Steroid hormone receptor
(estrogen-/androgen receptor)

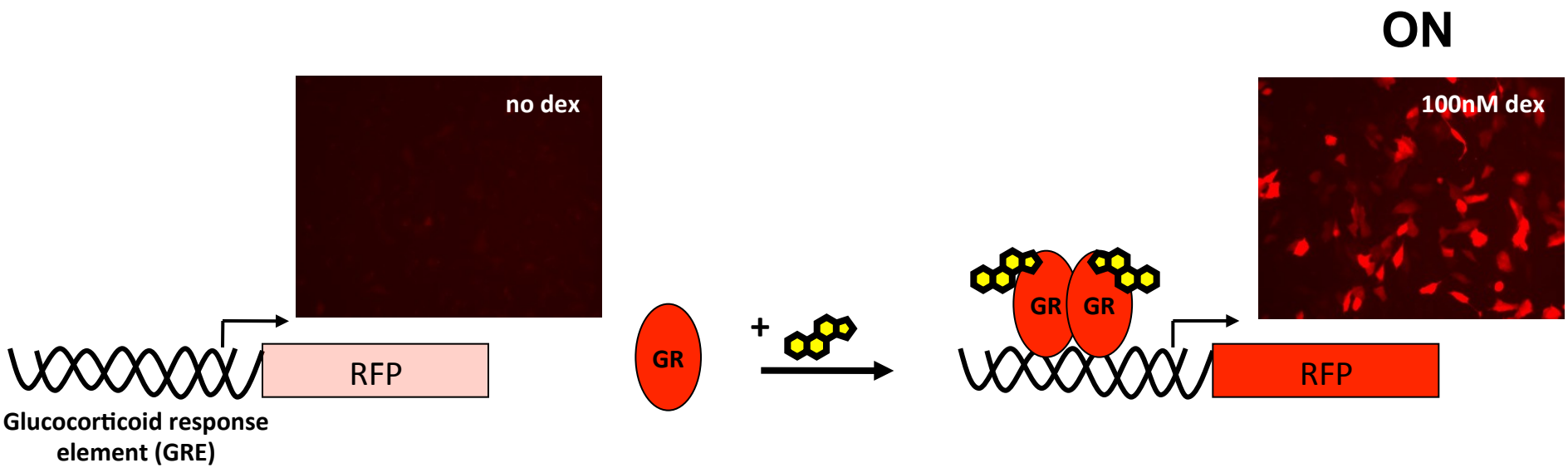
Activity controlled by hormone (cortisol/ dexamethasone)

Regulates expression of target genes

Biological response

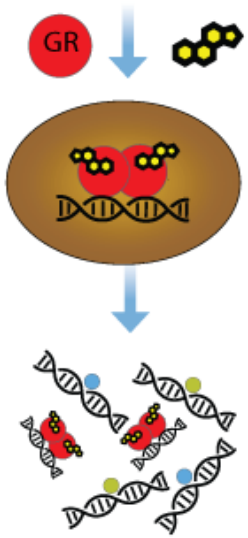


Good model system : Ligand as on/off switch



•Gerber *et al.*, PNAS 2009

Chromatin Immunoprecipitation (ChIP)

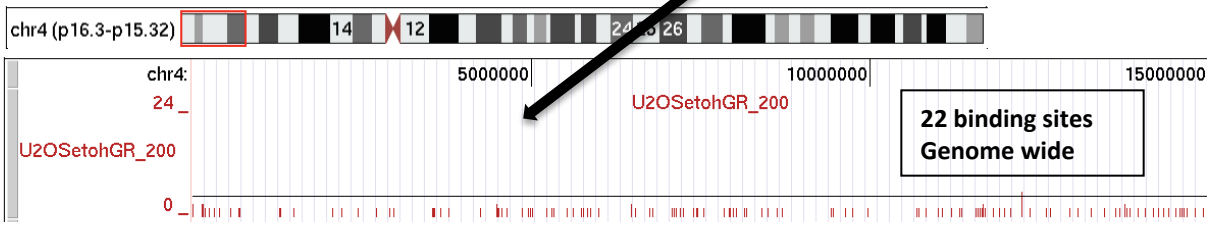


Treat cells with hormone

Crosslink and Fractionate Chromatin

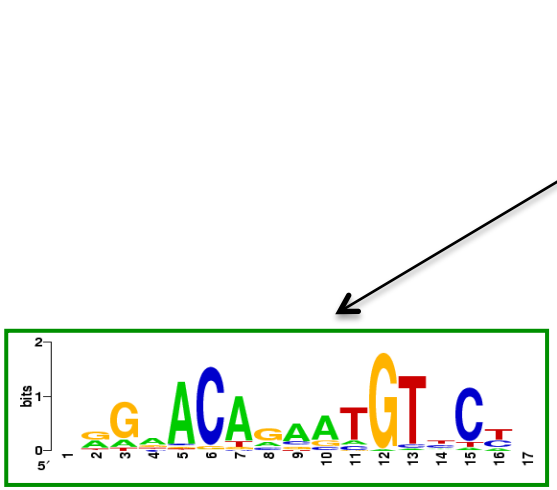
How it works:

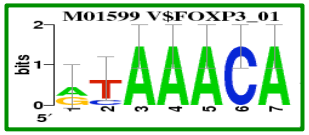
Control (no antibody / IgG / RNAi/knockout of TF)



Information from: Chromatin Immunoprecipitation (ChIP)

Sequence motifs enriched at bound regions





Lots of other enriched motifs found

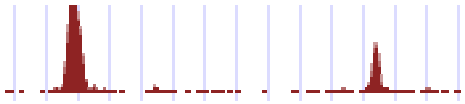
The good: Genome-wide approach

The bad: Biases e.g. in shearing efficiency of DNA / requires good Antibody

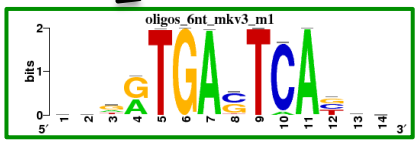
Resolution (200-300bp windows identified)

Information from: Chromatin Immunoprecipitation (ChIP)

What does this mean?



Lots of other enriched motifs found



?

?

?

GR

GR

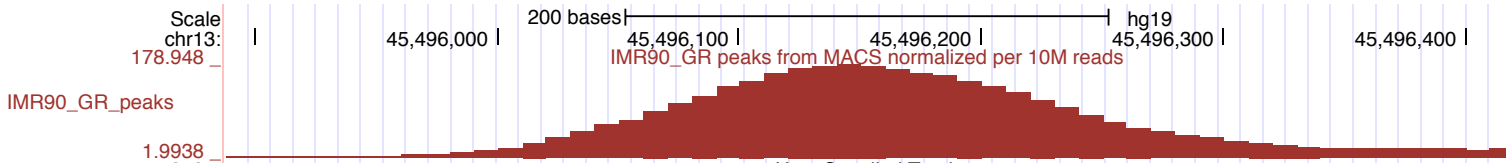
FoxA

GR

GR

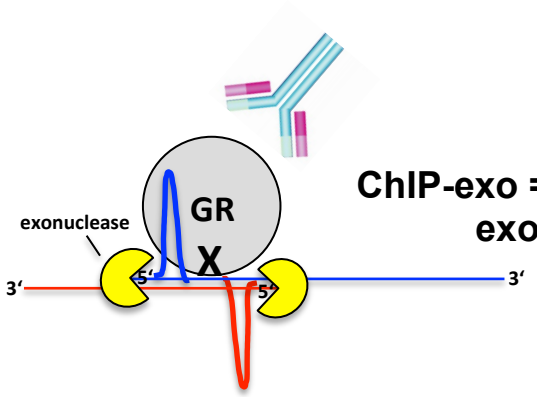
FoxA..

Improving resolution: ChIP-exo



Issue: **ChIP-seq:** **Resolution (200-300bp peaks):**

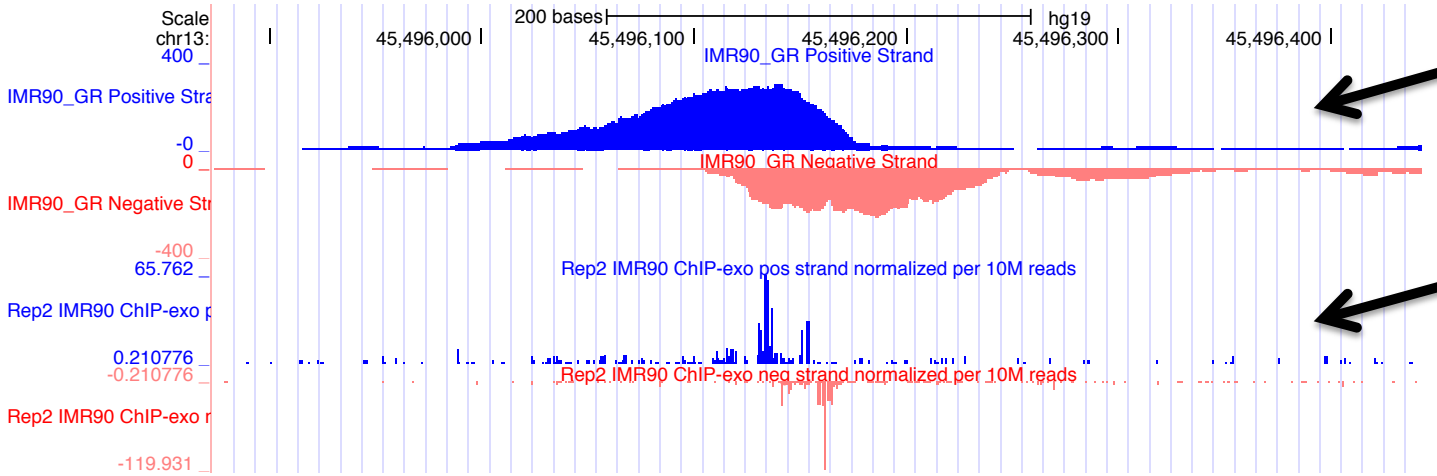
Solution?: **Improve Resolution using ChIP-exo (footprints)**



ChIP-exo = ChIP + lambda exonuclease

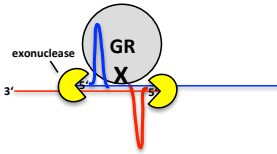
Rhee and Pugh. *Cell* (2011).
Mymryk and Archer. *NAR* (1994).

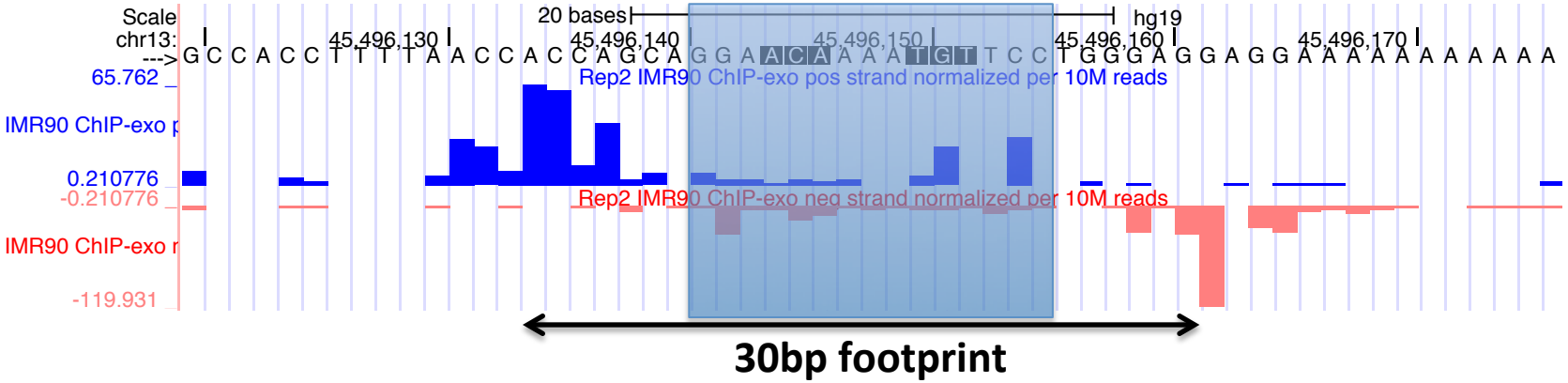
Improved resolution (footprint)



ChIP-seq

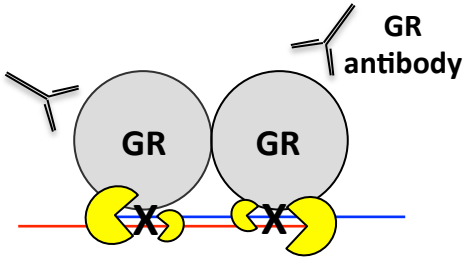
ChIP-exo
(Peconic)



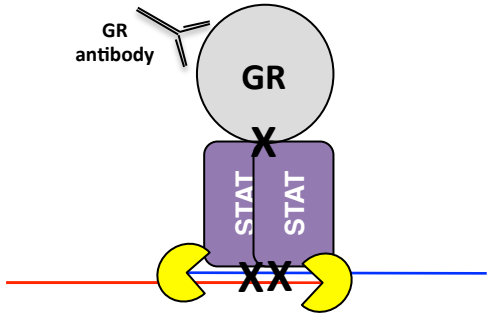


30bp footprint

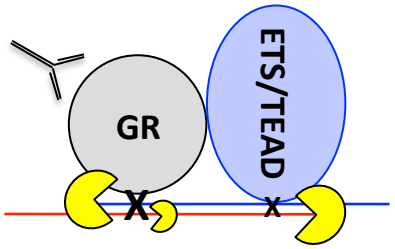
Footprints for other sequences ??



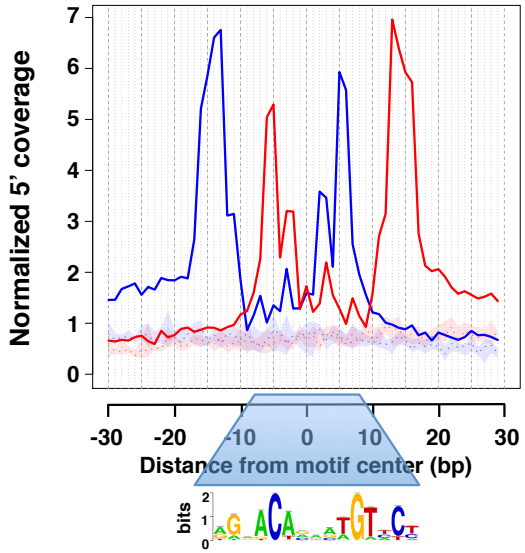
Dimeric GR binding

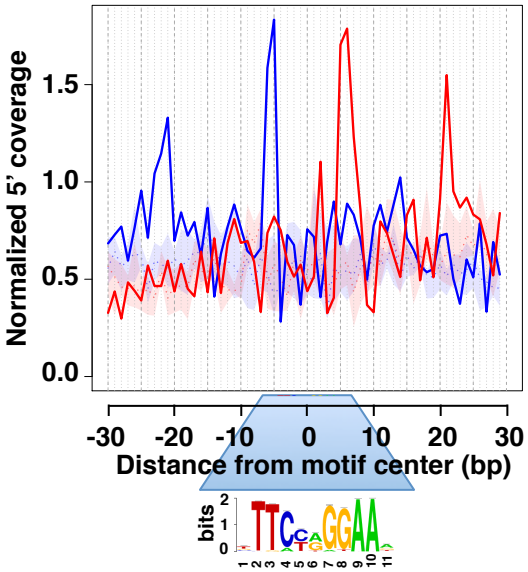


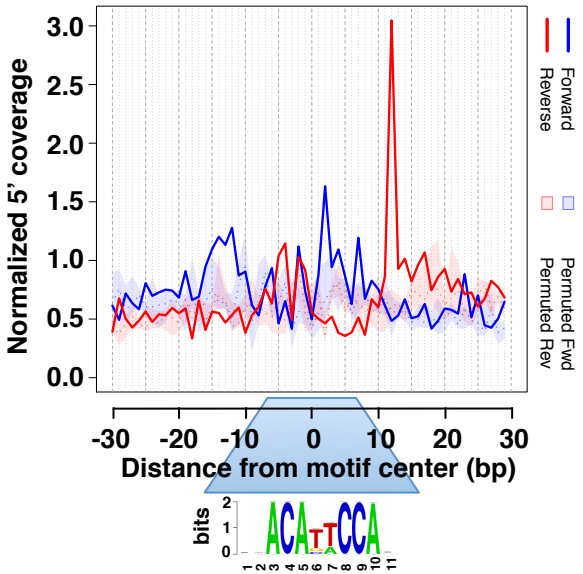
Tethered GR binding



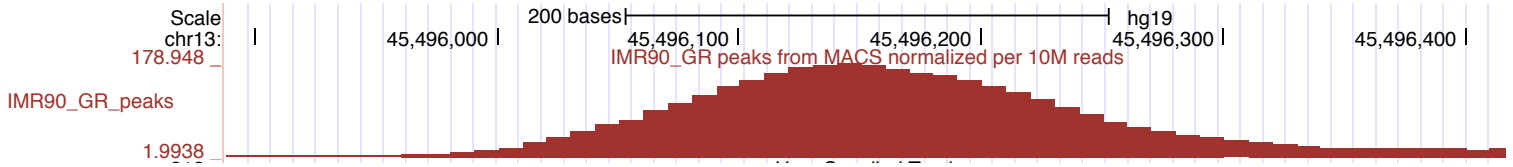
Composite / monomeric GR binding





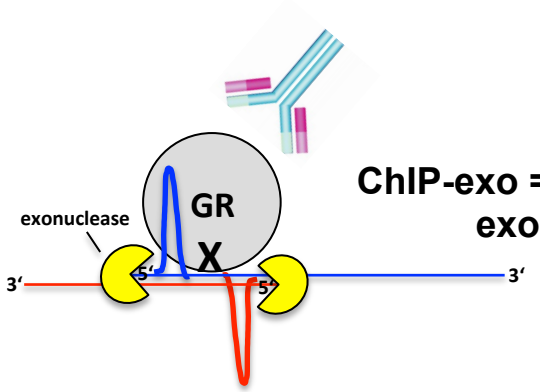


Improving resolution: ChIP-exo



Issue: **ChIP-seq:** **Resolution (200-300bp peaks):**

Solution?: **Improve Resolution using ChIP-exo (footprints)**



ChIP-exo = ChIP + lambda exonuclease

Rhee and Pugh. *Cell* (2011).
Mymryk and Archer. *NAR* (1994).

The good: Improved resolution

The bad: Only appears to work for subset of peaks (immature technology)????

Identification of cis regulatory elements:

Approach1: - Conservation analysis (evolution)

Approach 2: - Biophysical approaches (binding)

Approach 3: - Chromatin-based methods

Approach 4: Functional approaches (reporter genes & mutagenesis)

Chromatin & genomic elements

Chromatin / histone-modification based identification of cis-regulatory elements

What it tells you: - Candidate cis-regulatory elements

How it works: - ChIP-seq using anti-bodies against post-translationally modified histones

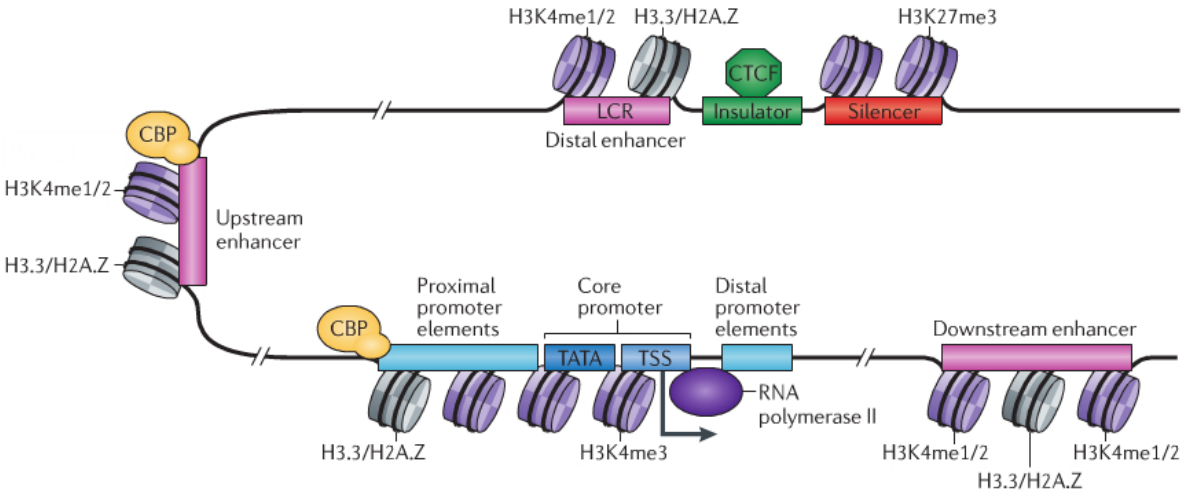


Figure 1 | **Transcriptional regulatory elements in metazoans.** The promoter is typically comprised of proximal, core and downstream elements. Transcription of a gene can be regulated by multiple enhancers that are located distantly and interspersed with silencer and insulator elements, which are bound by regulatory proteins such as CCCTC-binding factor (CTCF). Recent genome-wide data have revealed that many enhancers can be defined by unique chromatin features and the binding of cyclic AMP-responsive element-binding (CREB) protein (CBP). H3K4me1/2, histone H3 mono- or dimethylation at lysine 4; H3K4me3, histone H3 trimethylation at lysine 4; H3K27me3, histone H3 trimethylation at lysine 27; H3.3/H2A.Z, histone variants H3.3 and H2A.Z; LCR, locus control region; TATA, 5′-TATAAAA-3′ core DNA sequences; TSS, transcription start site. Figure is modified, with permission, from REF. 97

Promoter:
- H3K4me3 high
- H3K9ac high

Enhancer:
- H3K4me1 high
- H3K27ac high

Silencer:
- H3K27me3 high

……

Ong and Corses. Nature Reviews Genetics (2011).

Chromatin & genomic elements

Consortia cataloguing histone modifications:

(cataloguing functional regions in the genome)

The good: Genome-wide approach

The bad: Correlation ≠ causation (biased) / specificity of antibodies

Outline

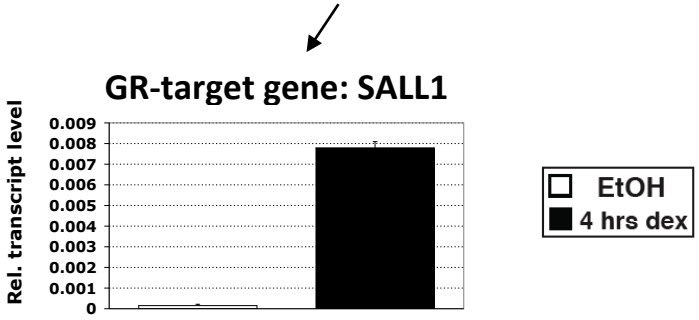
- Techniques to find cis regulatory elements (& transcription factor binding sites)

Function:

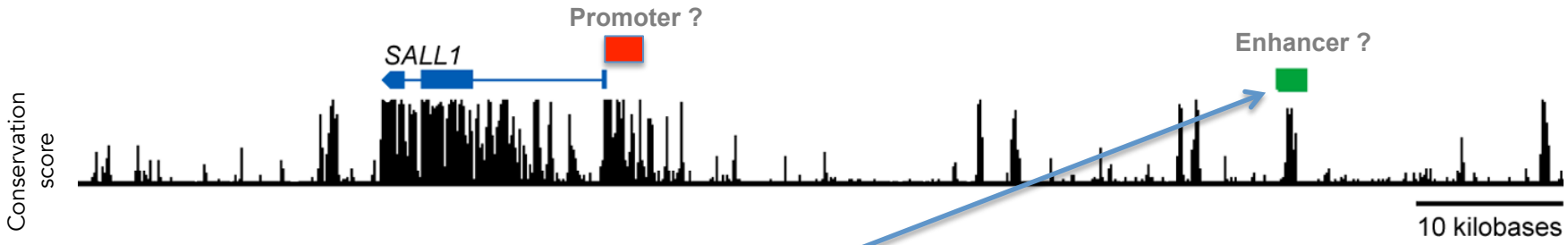
- **Linking binding and regulation**

Functional characterization of cis regulatory elements:

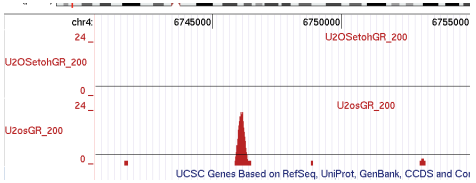
Observation: GR induces the expression of SALL1 gene:



Question: What enhancer allows GR to control the expression of SALL1 gene?



This enhancer ????

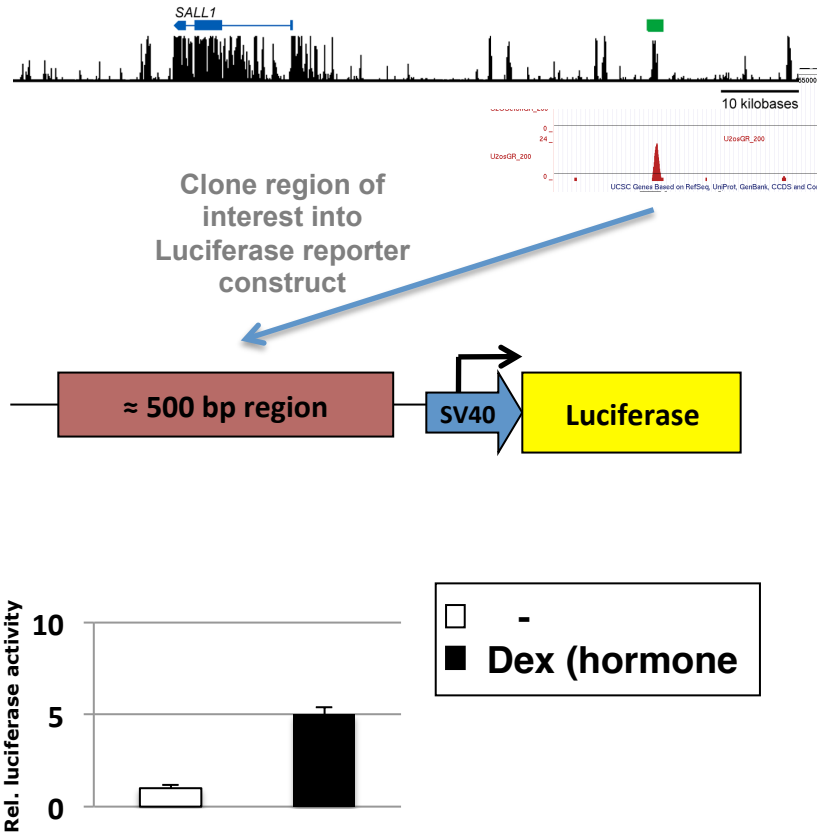


Evidence:
- **Region conserved**
- **DNAse hypersensitive („open“)**
- **High histone H3K4me1 & H3K27ac levels**
- **ChIP shows that TF of interest (GR) binds to this locus**

Functional approaches: Reporter assays

Question: **Can GR activate from this sequence??**

Approach: **Test it in Luciferase reporter assay**

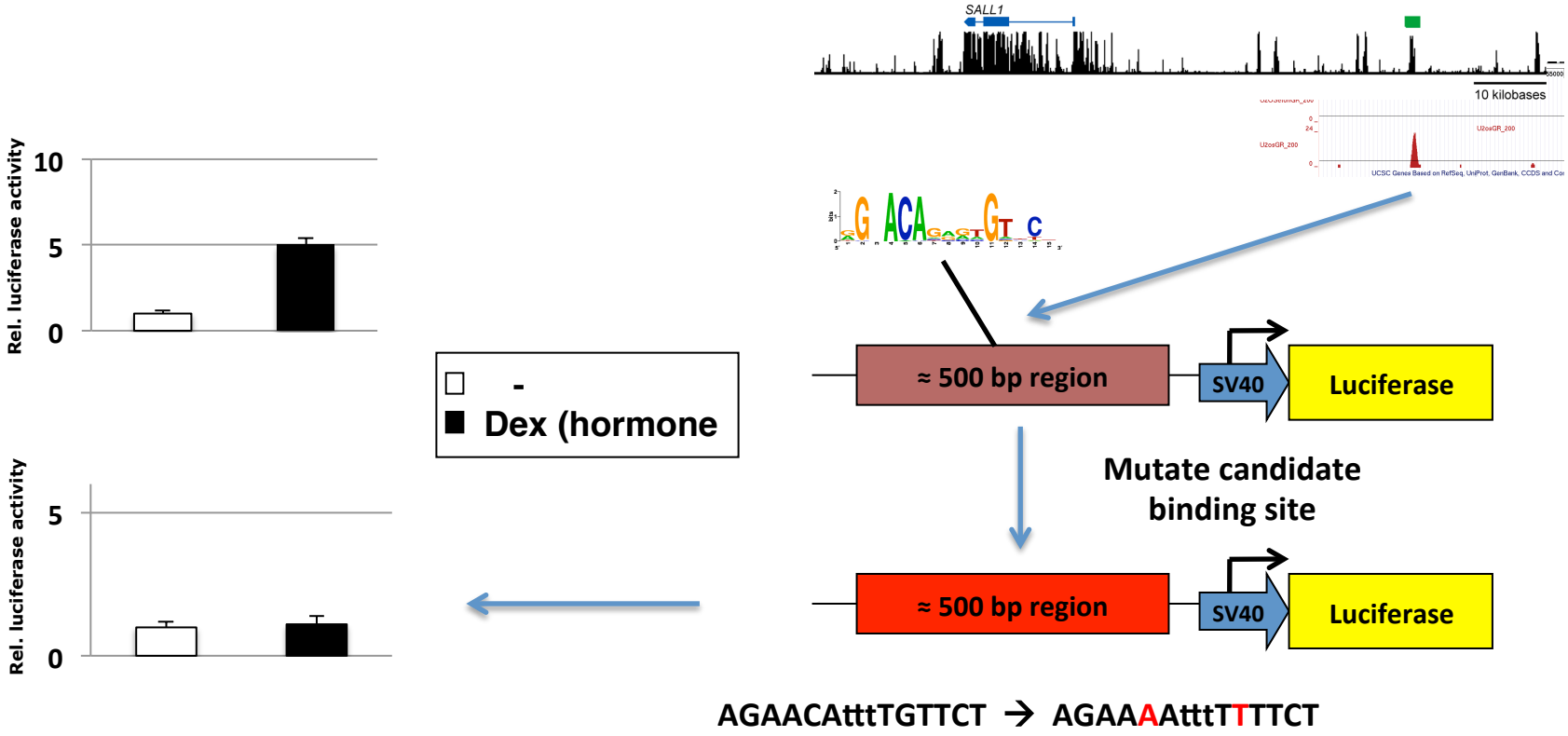


Result: **Region is indeed capable to activate reporter**

Functional approaches: Reporter assays

Question: What sequence(s) allow GR to activate ???

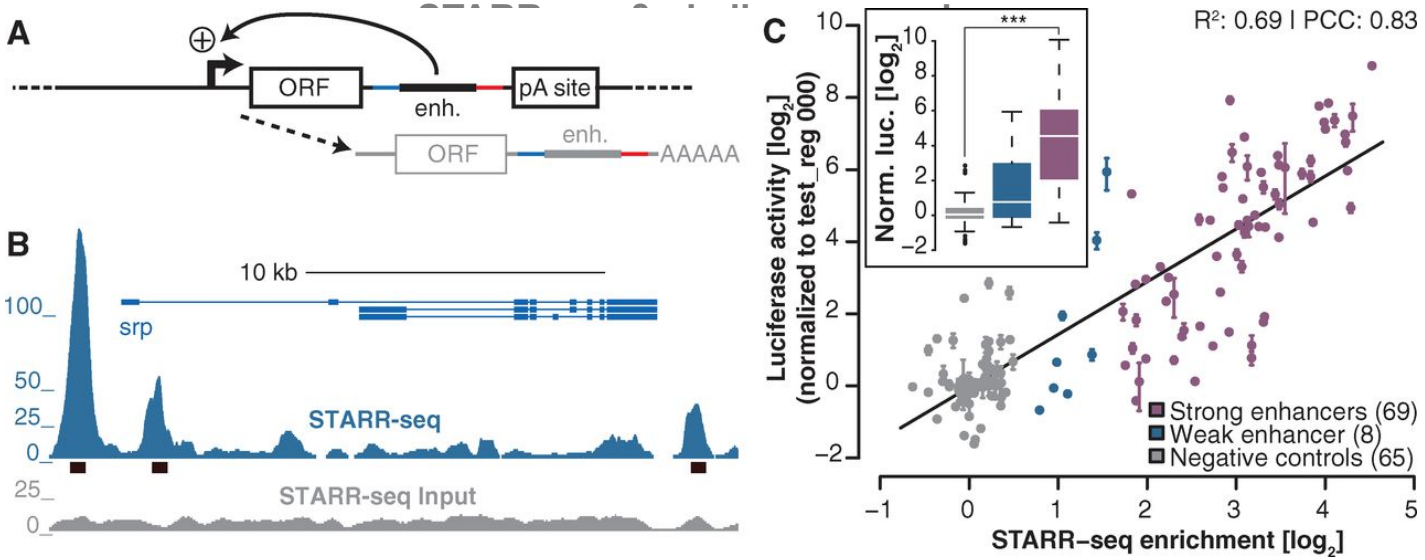
Approach: Test effect of mutations in Luciferase reporter assay



Result: Candidate sequence indeed required for GR-dependent transcriptional activation

Functional approaches: Reporter assays…

Multiplex to test large sets of test sequences…..



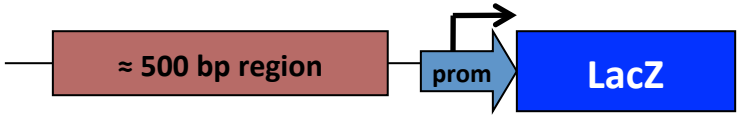
Arnold et al., Science 2013

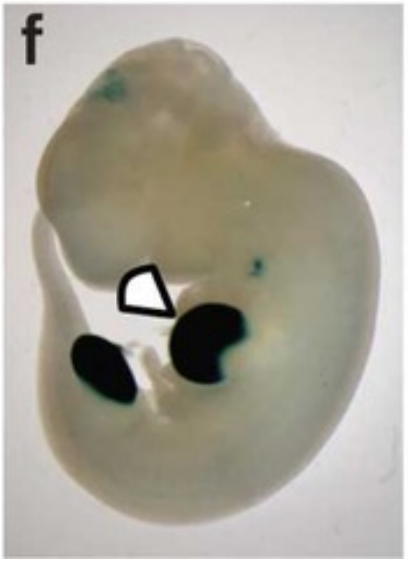
Functional characterization of cis regulatory elements:

Observation: GR induces the expression of SALL1 gene in limbs

Question: Does enhancer allow GR to control the expression of SALL1 gene in limbs?

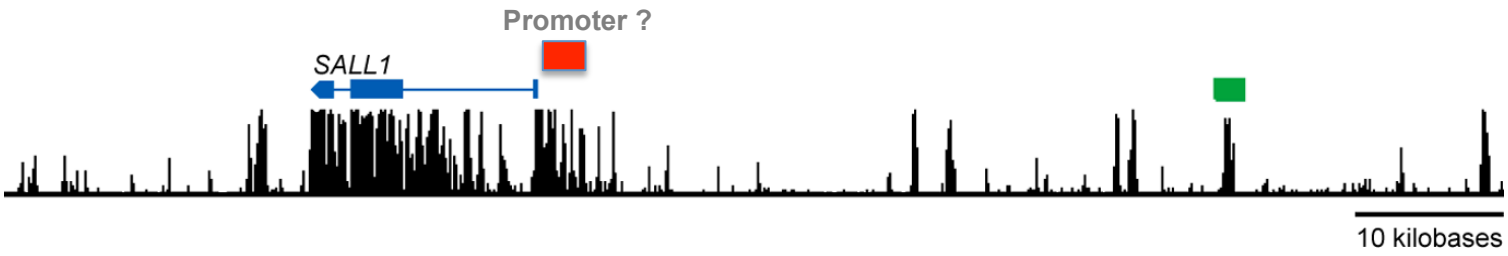
Approach: LacZ reporter with enhancer sequence of interest is injected into mouse zygotes & expression of lacZ monitored.

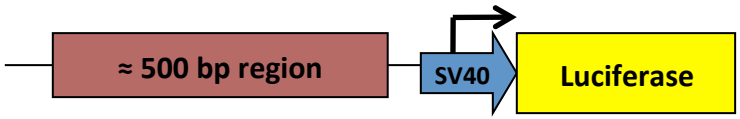




Result: Enhancer can indeed drive expression in limbs (& brain?)

Reporter assays:





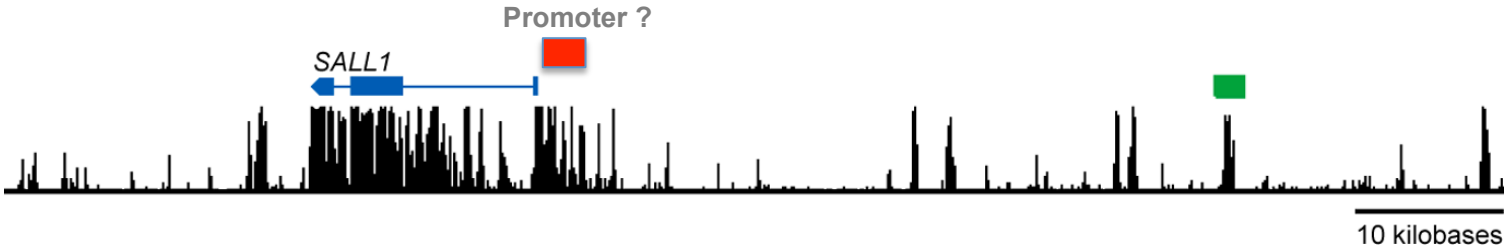
The good: - Fast

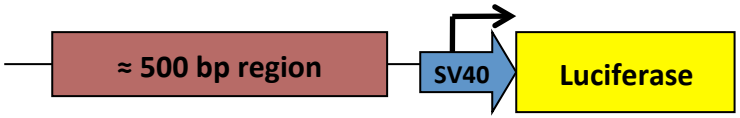
- Easy to manipulate e.g. sequence

- Allow comparison of regulatory sequences in controlled context

The bad: Artificial setting…….

Issues with reporter assays…



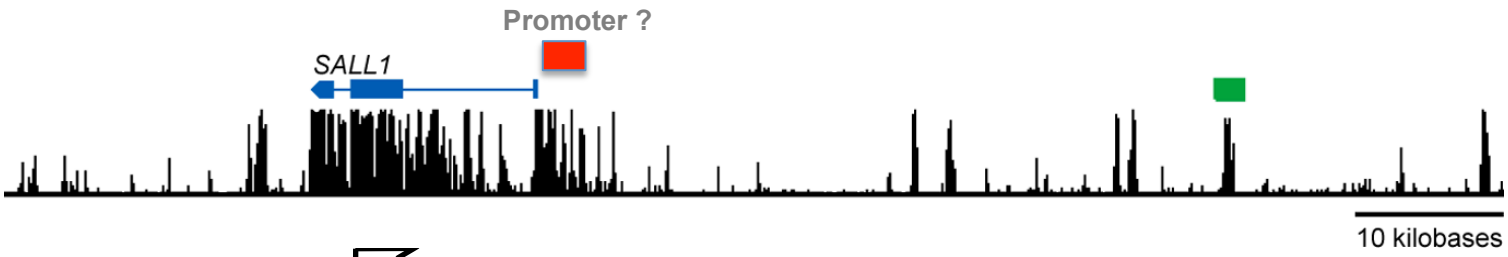


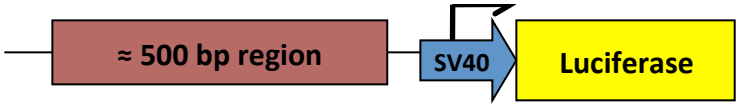
1. Distance to promoter different
2. Chromatin context recapitulated ?
3. Context changes
4. Often different promoter
5. ……

Take home message:

These reporter studies show if a regulatory sequence CAN drive expression for instance in a particular tissue NOT if it actually does……..

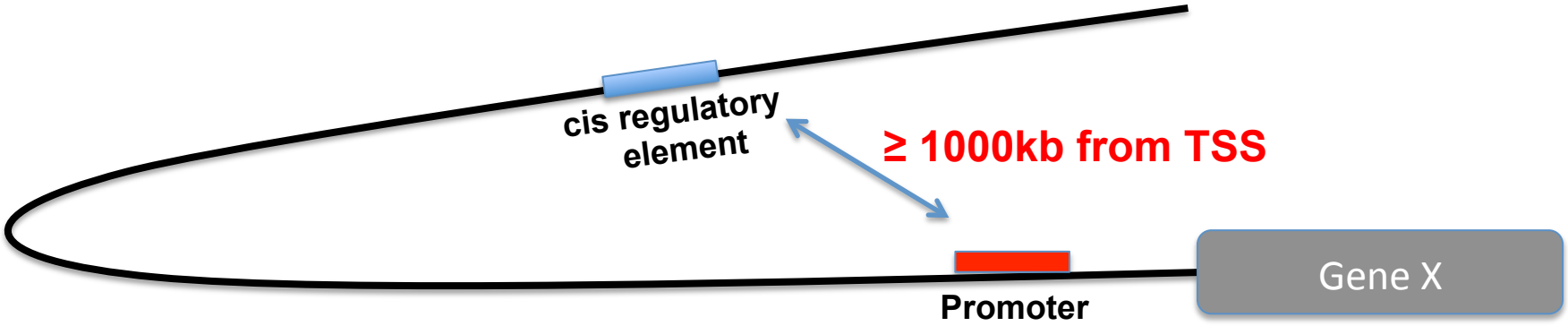
Q: Does enhancer play role in regulating gene ?





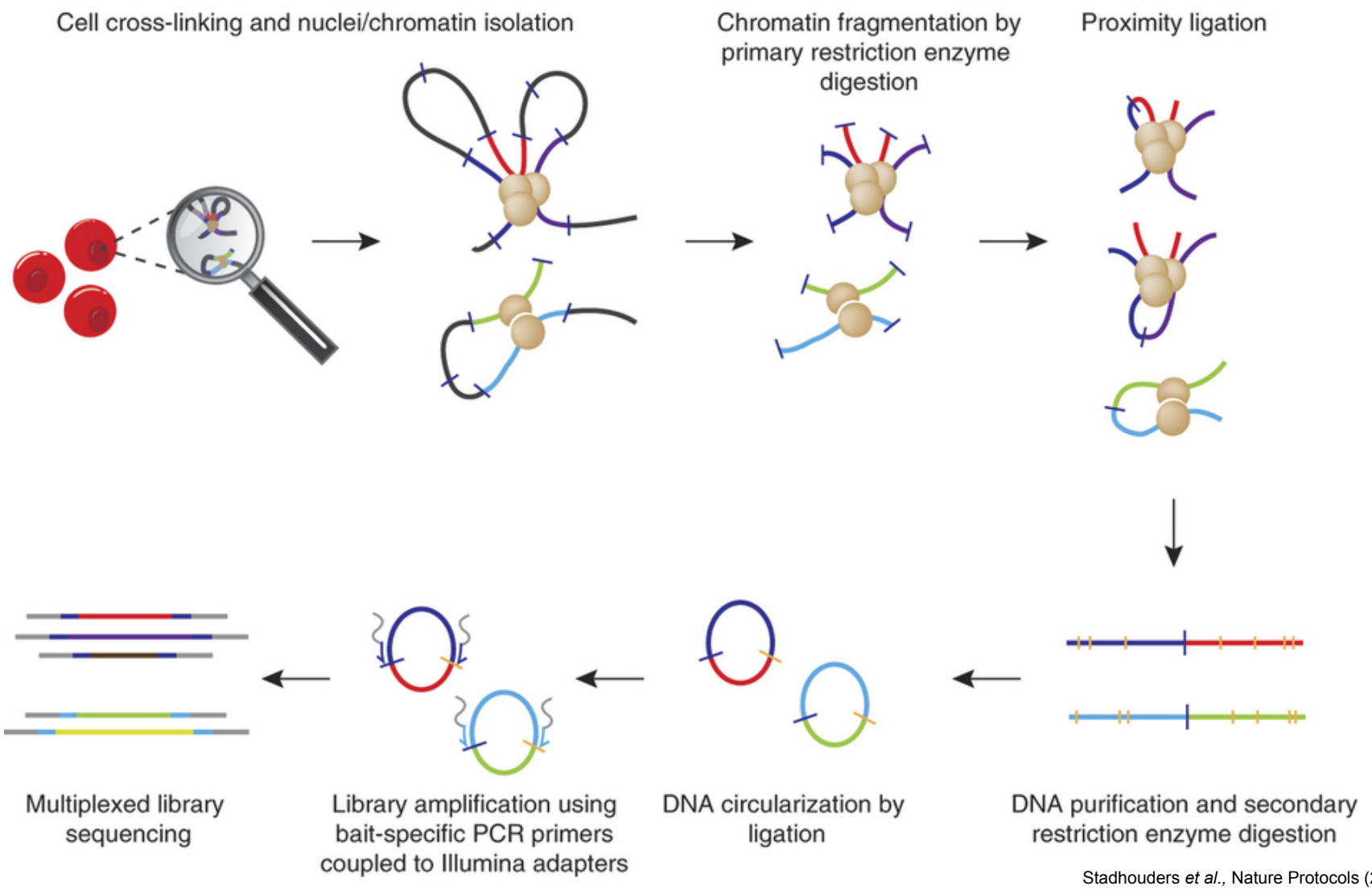
1. **Distance to promoter different….**

Additional information that would make it more likely that it actually does??



Proximity of enhancer to promoter (in 3D space)

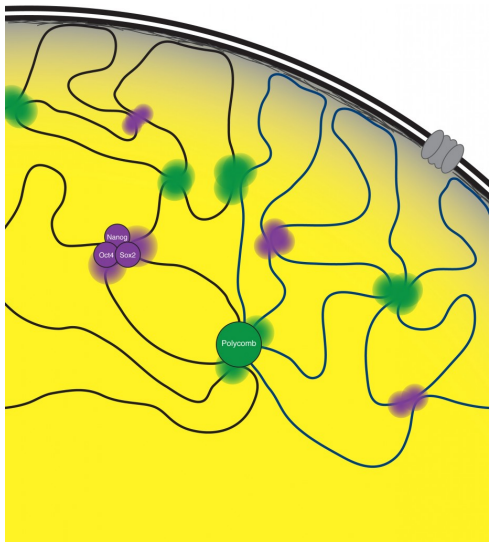
Chromatin Conformation Capture (3C)

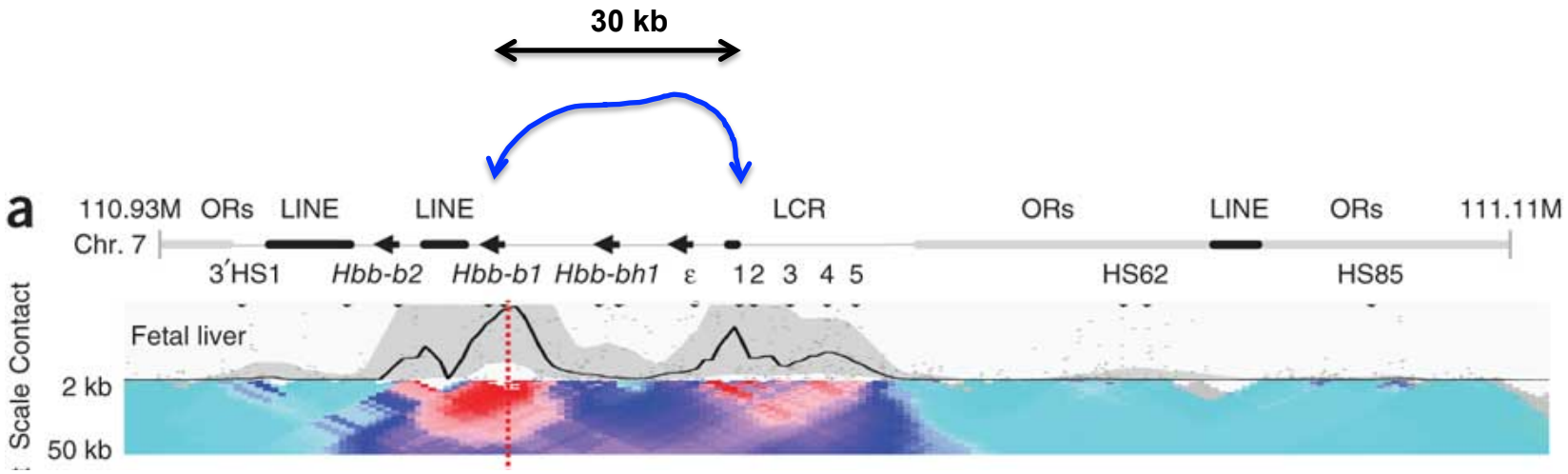


Stadhouders *et al.,* Nature Protocols (2013).

What it tells you: If two sequences are near one another in nucleus

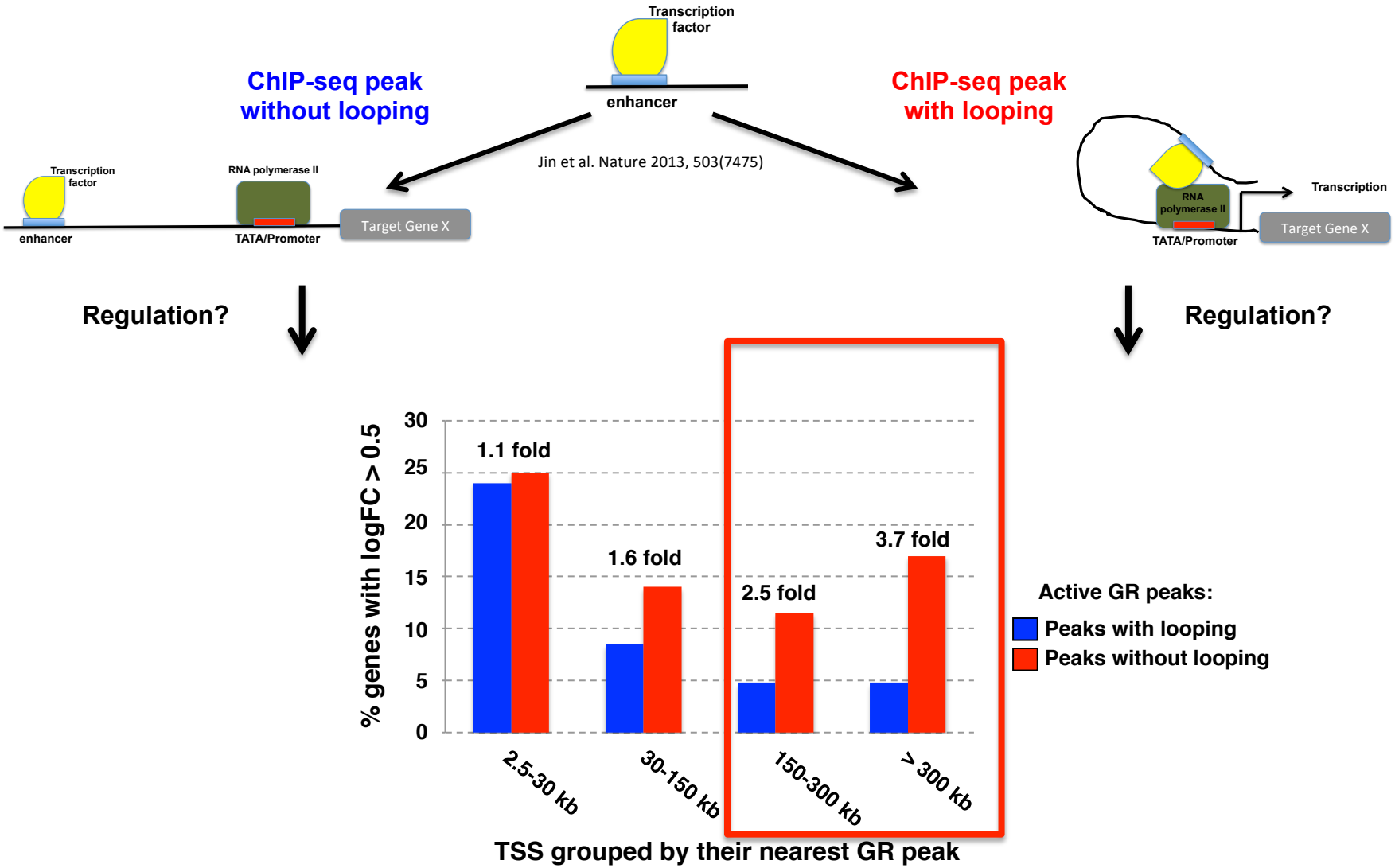
Hi-C & long range interactions





Werken et al. Nature Methods 2012

Finding missing link(s) between binding and regulation…..

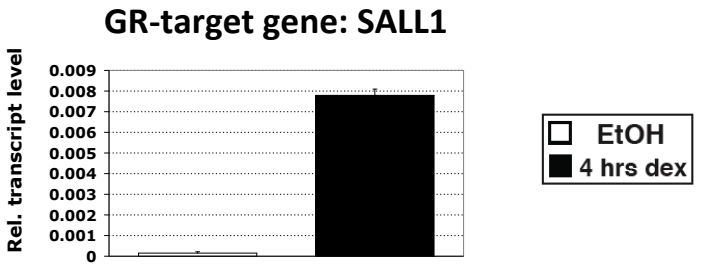


- Improved correlation between binding and regulation

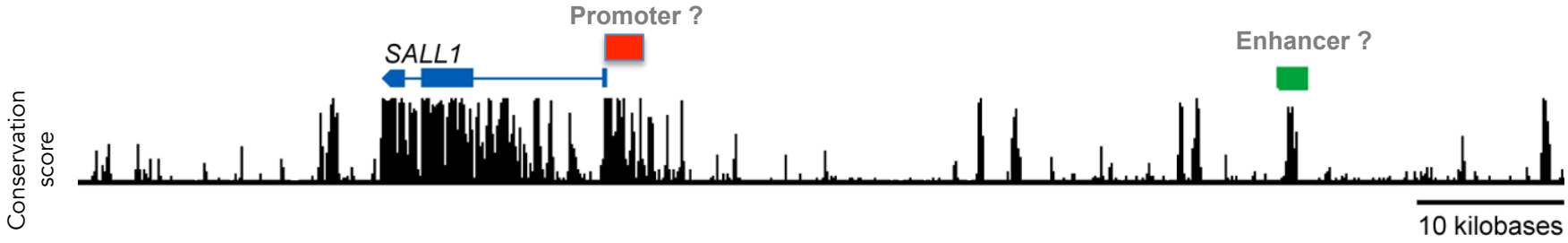
Robert Schöpflin

Functional characterization of cis regulatory elements:

Observation: GR induces the expression of SALL1 gene:

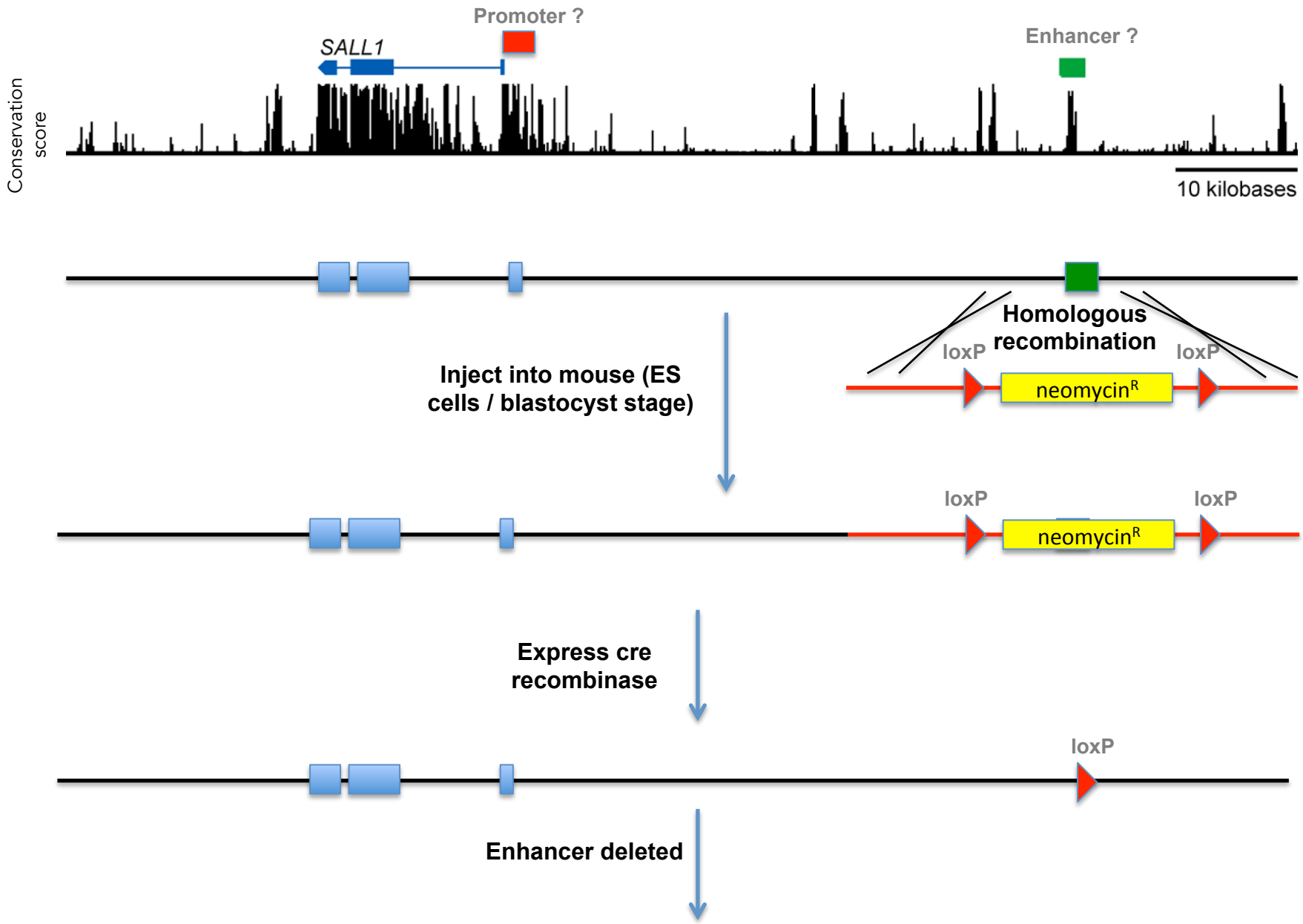


Question: What enhancer allows GR to control the expression of SALL1 gene?

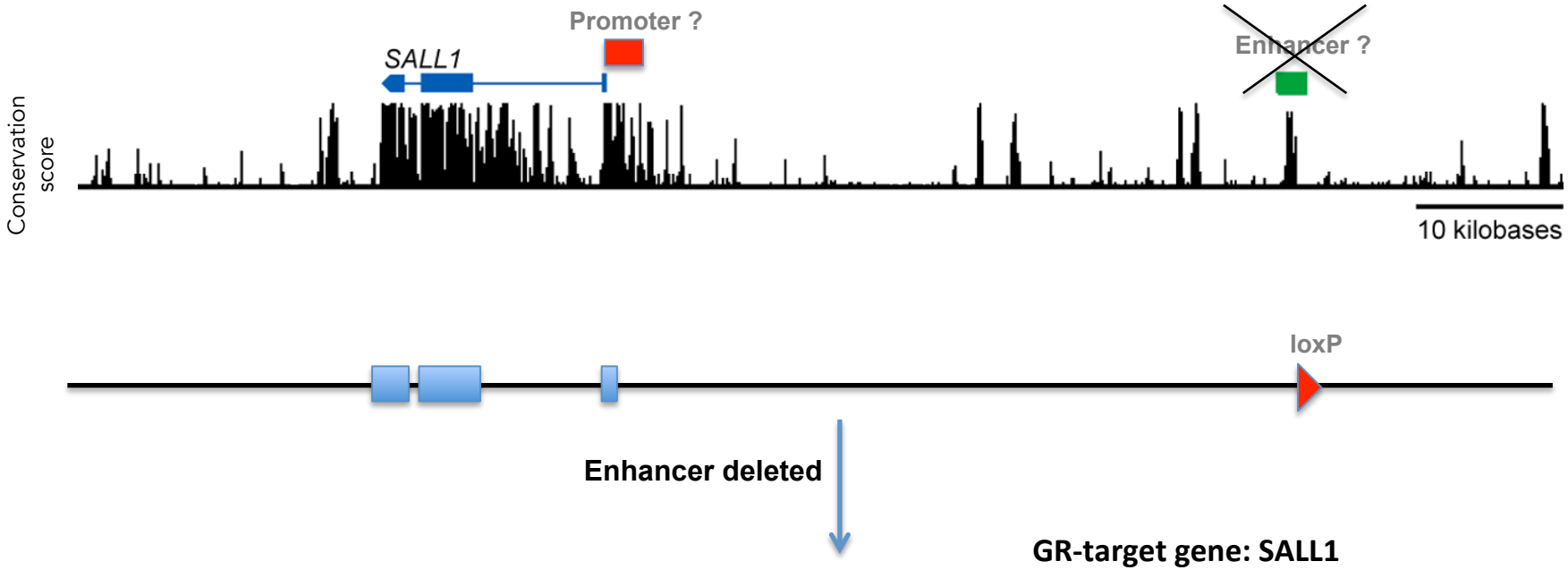


How to test this? Delete enhancer in natural context..... (Mouse transgenics)

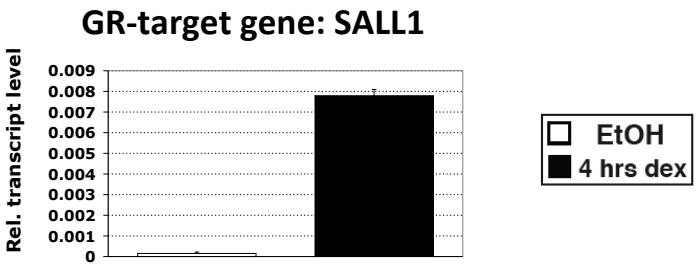
Mouse transgenics



Mouse transgenics

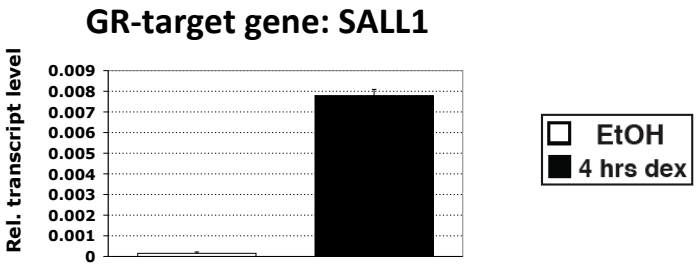


Observation: GR induces the expression of SALL1 gene:

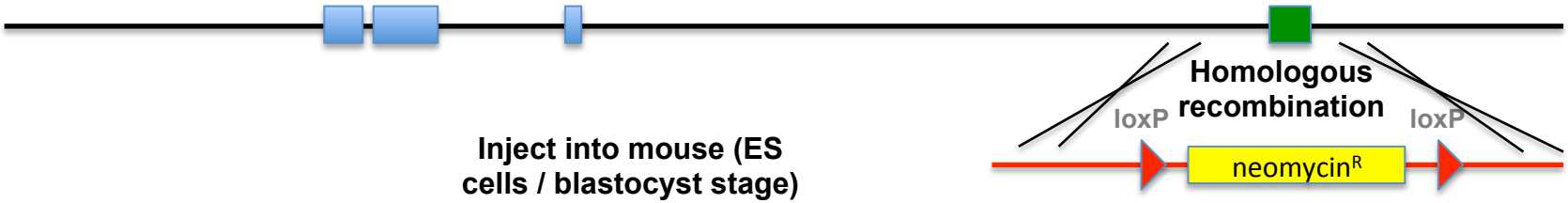


Couple of months (years..) later: test

Lucky: regulation lost
Unlucky Regulation still present
(Hypothesis wrong / redundancy....)



Mouse transgenics



The good: Experiments done in endogenous setting

The bad: Slow, expensive, risky, need to go through mice……..

The future is now….. Genome editing…

Ideally we would be able to change individual base pairs of our interest in any cell type of interest

Therapeutic interest (Future): Repairing disease causing mutations (gene therapy e.g. oncogenes)

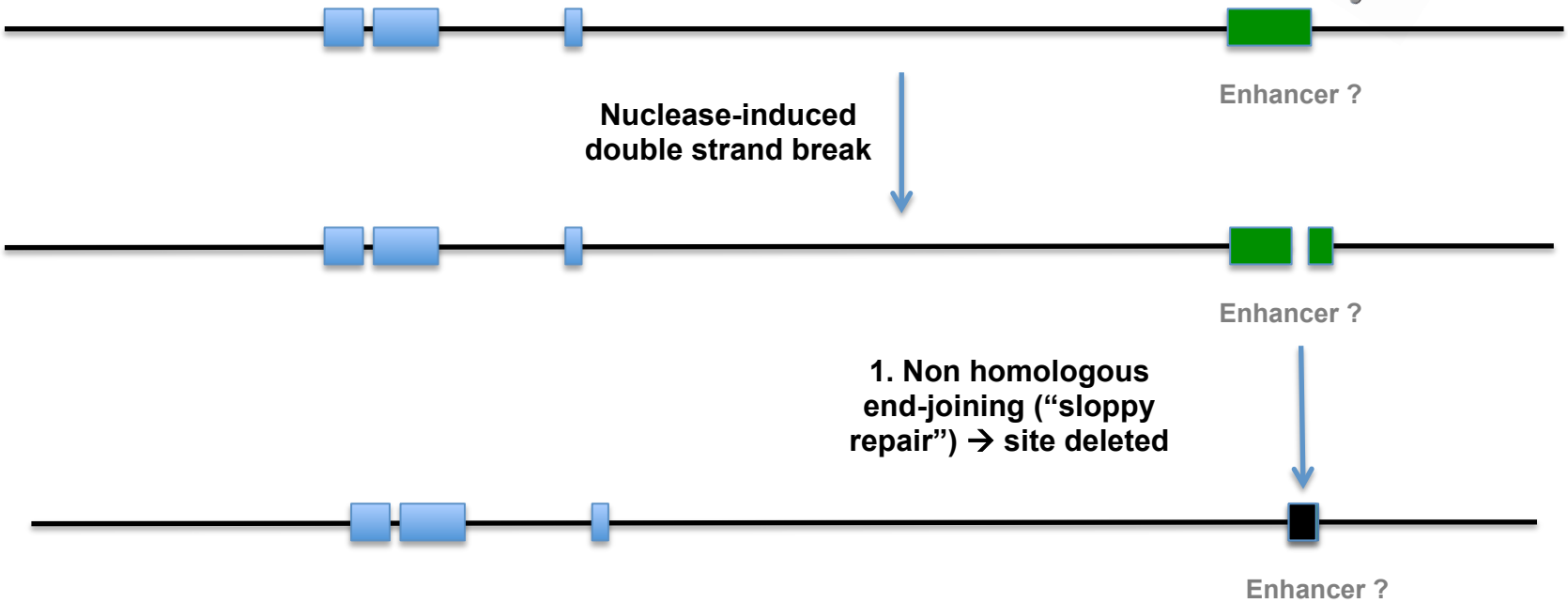
Mechanistic interest: Being able to change putative regulatory sequences to test their function

Genome editing:

The principle:

- **Induce double strand breaks in region of interest**

- **Hijack cell's repair mechanism to either:**
  1) Delete site

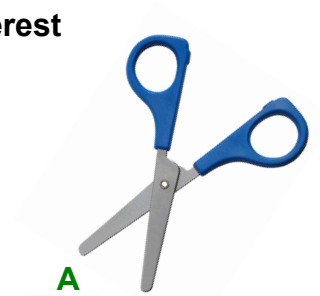


Genome editing:

The principle:

- **Induce double strand breaks in region of interest**

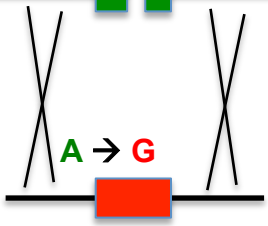
- **Hijack cell's repair mechanism to either:**
  1) **Delete site**
  2) **Mutate site**



Enhancer ?

Nuclease-induced double strand break

1. Homologuous recombination → allow you to change individual bases…



Specific sequence in Enhancer required ?

Genome editing:

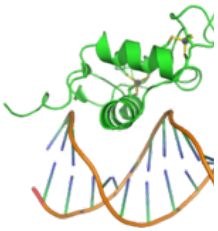
Challenges:

1. **Directing the nuclease to the appropriate location**

- **Achieved by fusing nuclease to zinc finger DNA binding domains that can be engeneered to recognize site of interest**

2. **Specificity (induce break only at site of interest)**

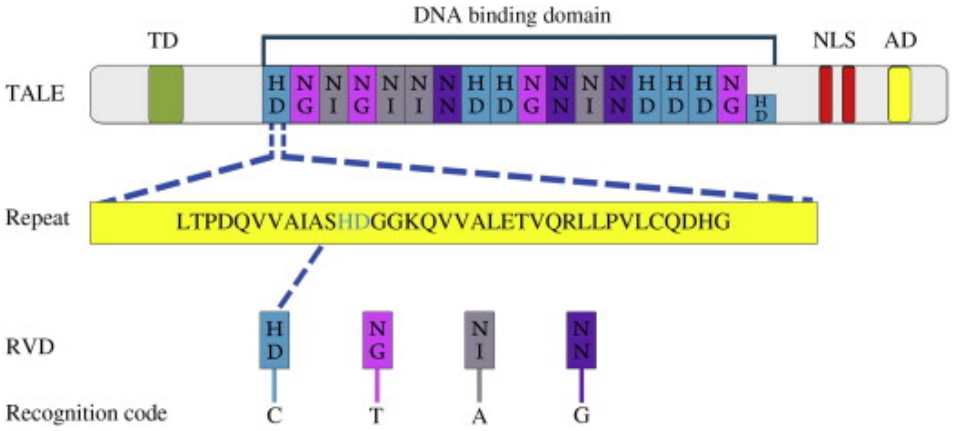
- Achieved by : bring obligatory homodimeric FokI sites together

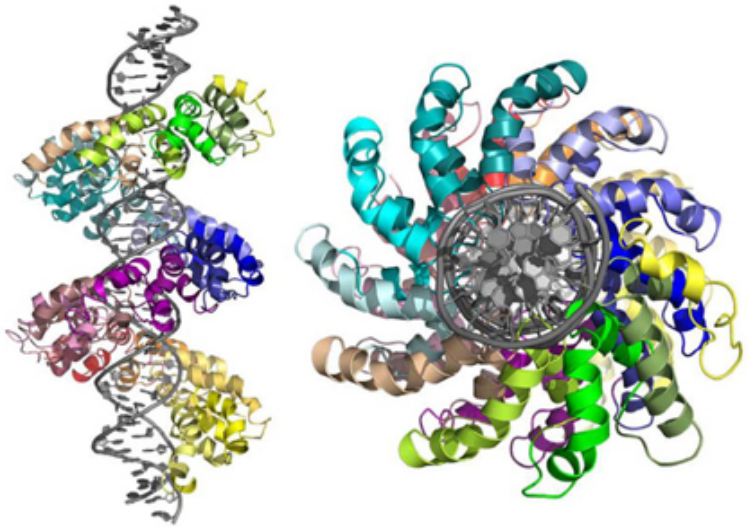


Genome editing:

Recent development 1.

TALENs: Better (modular DNA recognition) but still tricky



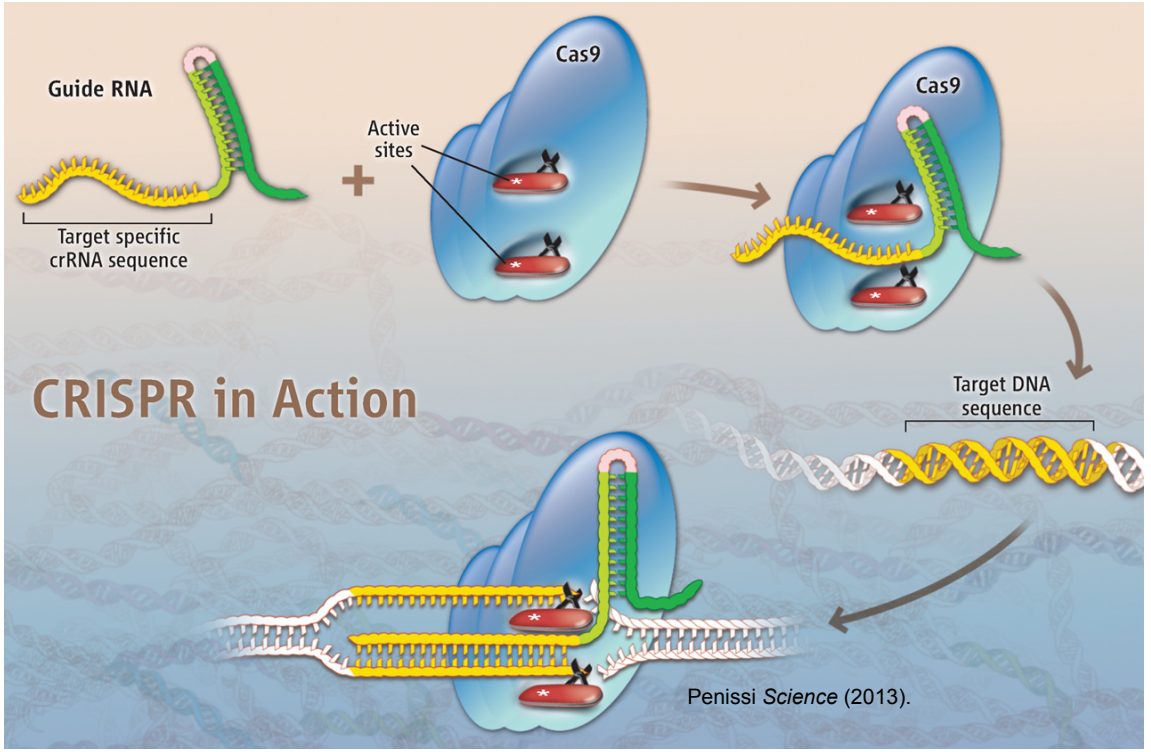


Genome editing:

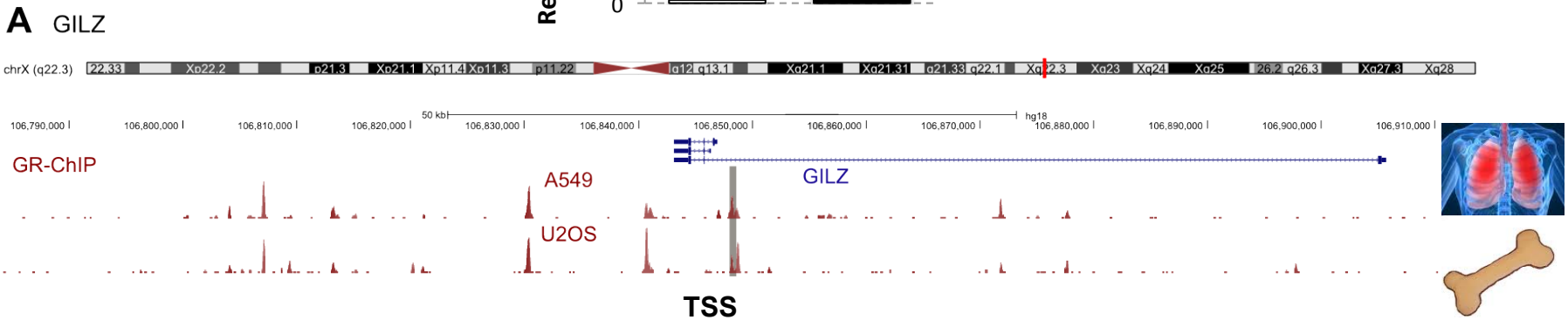
Recent development 2.

CRISPR/CAS9 editing technology (complementary RNA guided)....

(derived from a kind of bacterial adaptive immune system)

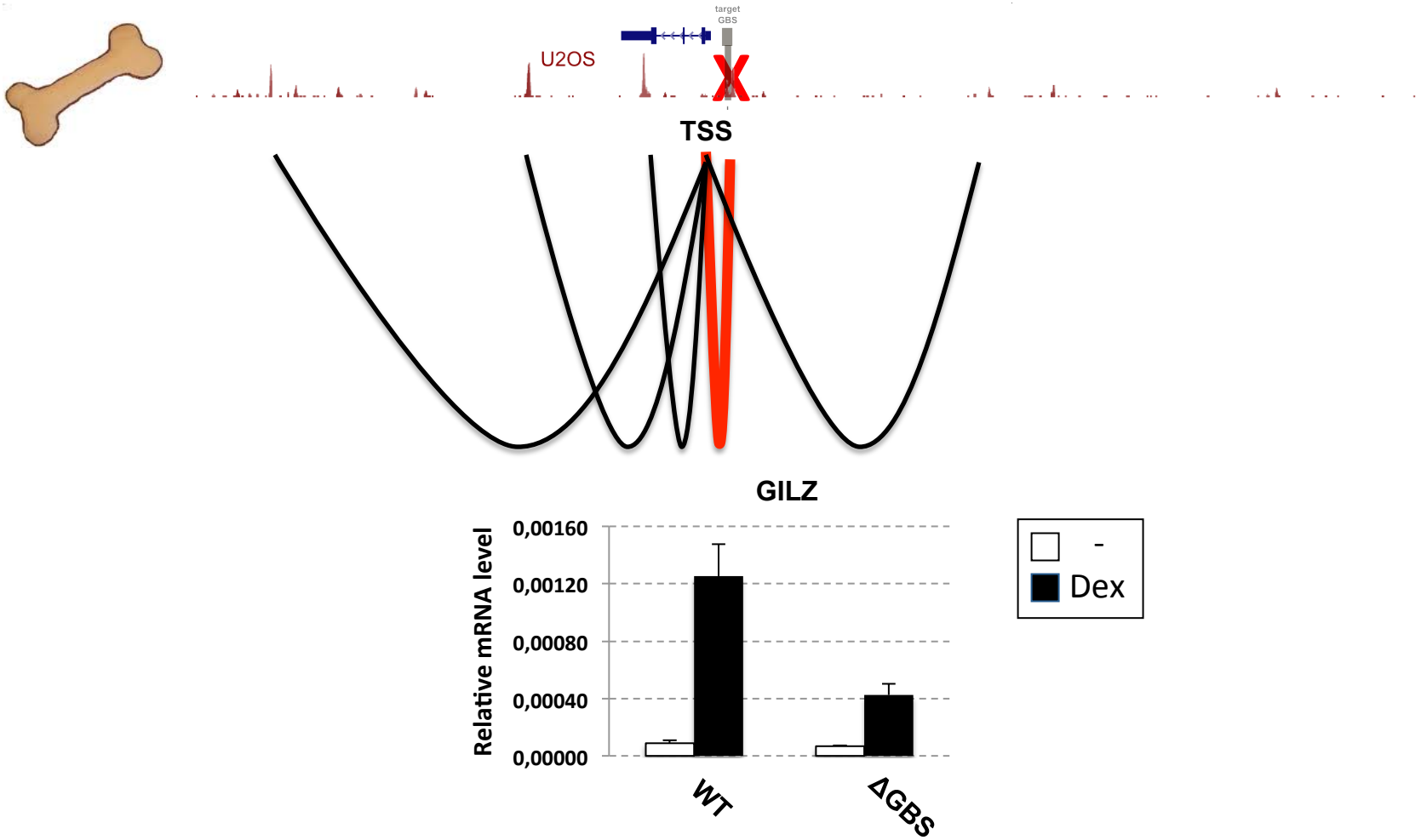


Promoter bashing 2.0 (genomic)



Potentially complex wiring → Study role of individual wires in regulation of gene

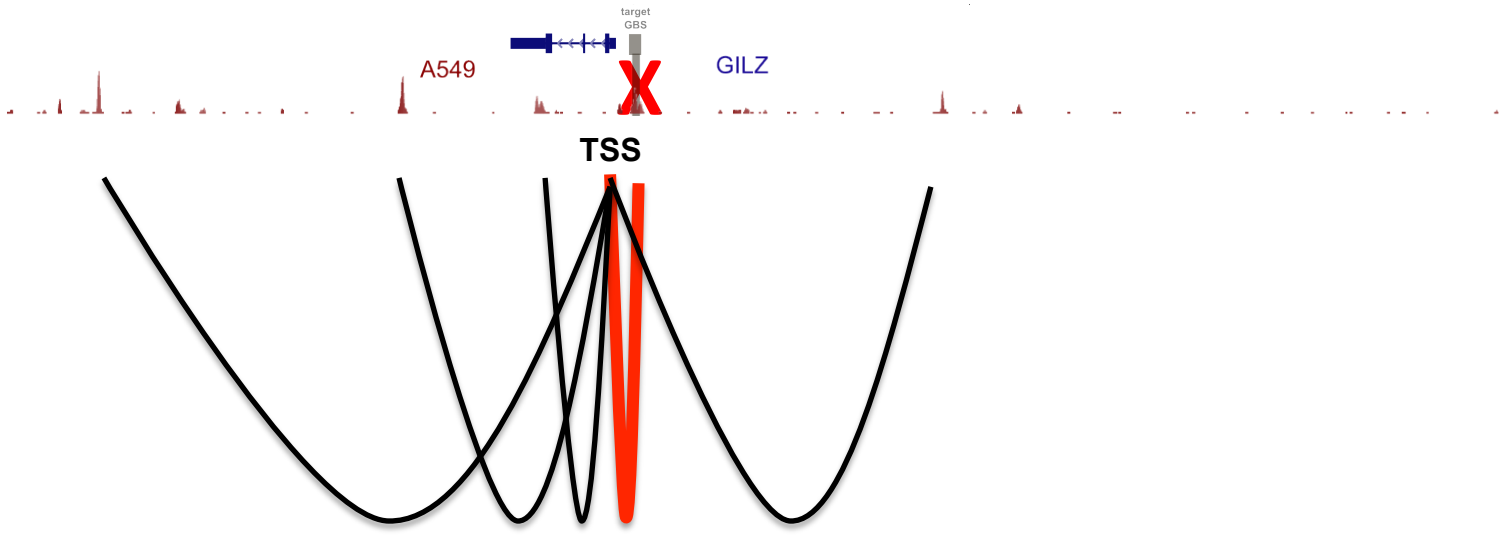
Promoter bashing 2.0:



- GR binding site important for regulation

Maika Rothkegel / Verena Thormann

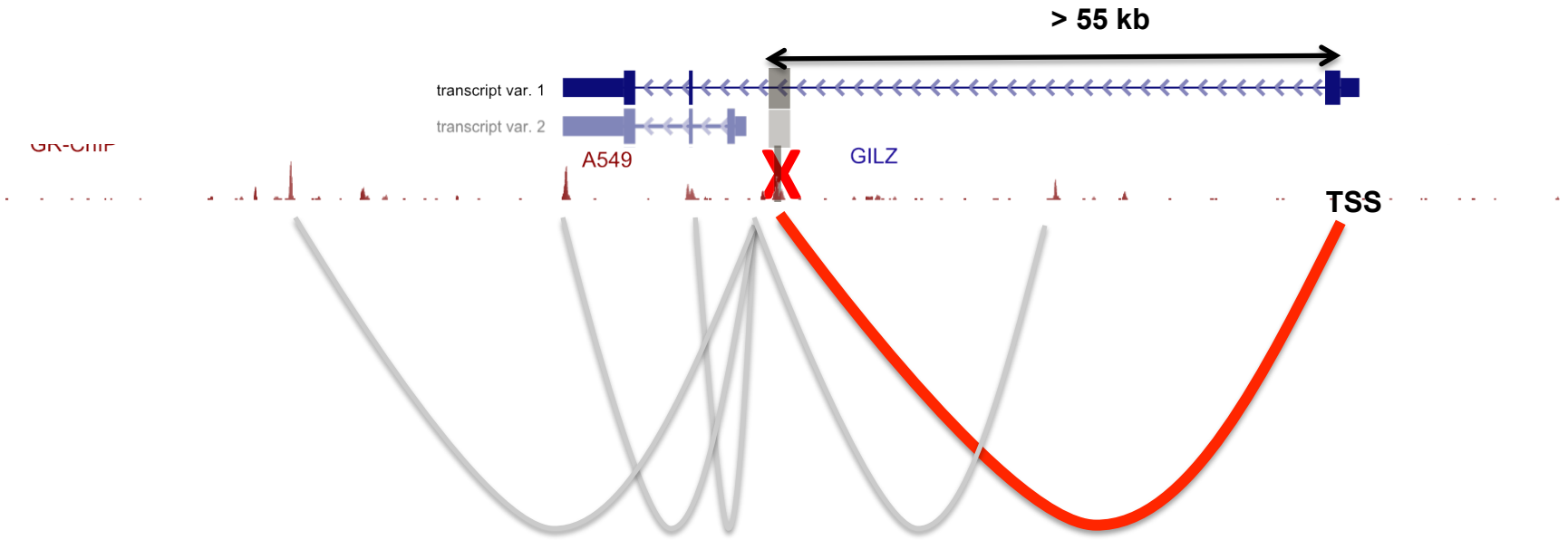
Promoter bashing 2.0:

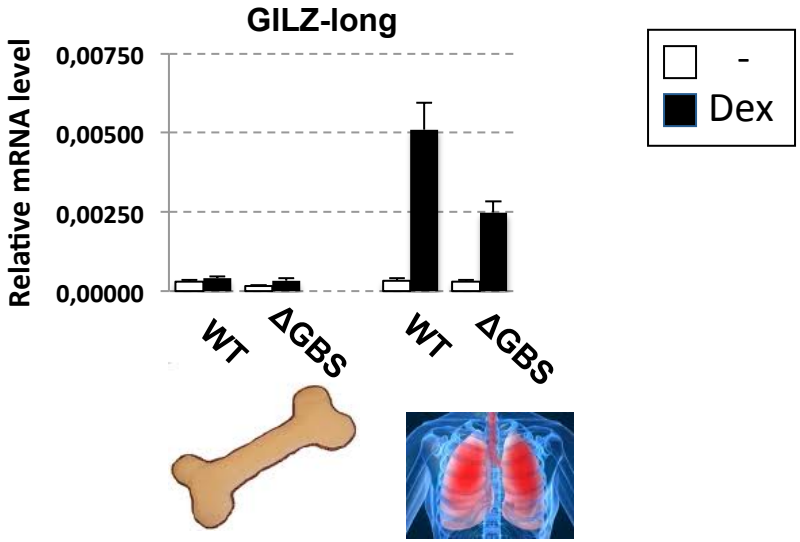


- GR binding site NOT important for regulation

Maika Rothkegel / Verena Thormann

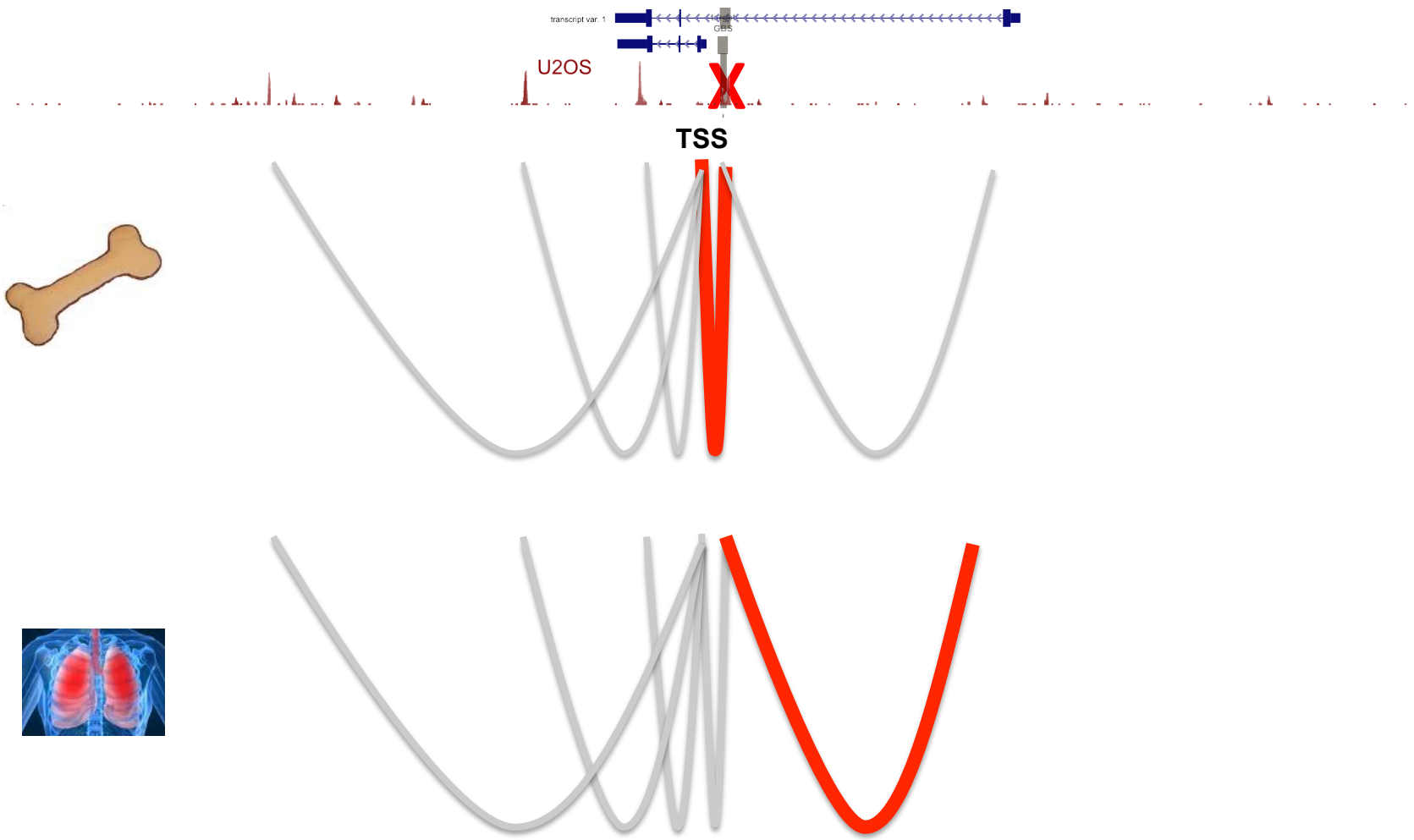
Promoter bashing 2.0:





- Context-dependent role for GR binding site

Cell-type-specific interactions



Maika Rothkegel / Verena Thormann

Genome editing

The good: Takes place in natural context / relatively fast / works for cell lines & whole animals

The bad: Off target effects (specificity??)

Overview

- Many techniques available to identify & characterize cis regulatory elements

- Each with their strengths & weaknesses……

Which one(s) to use depends on question asked, (time available, budget…)

- Future / current challenges:
  - Linking Cis regulatory elements to genes
  - How does integration of multiple signals at individual cis regulatory elements influence transcriptional output?
  - ……..