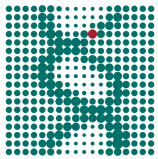
Applying bioinformatical analysis to study transcription…….

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



Max Planck Institute
for Molecular Genetics

Combining wetlab & Bioinformatical approaches to study transcriptional regulation

Correlation ≠ causation

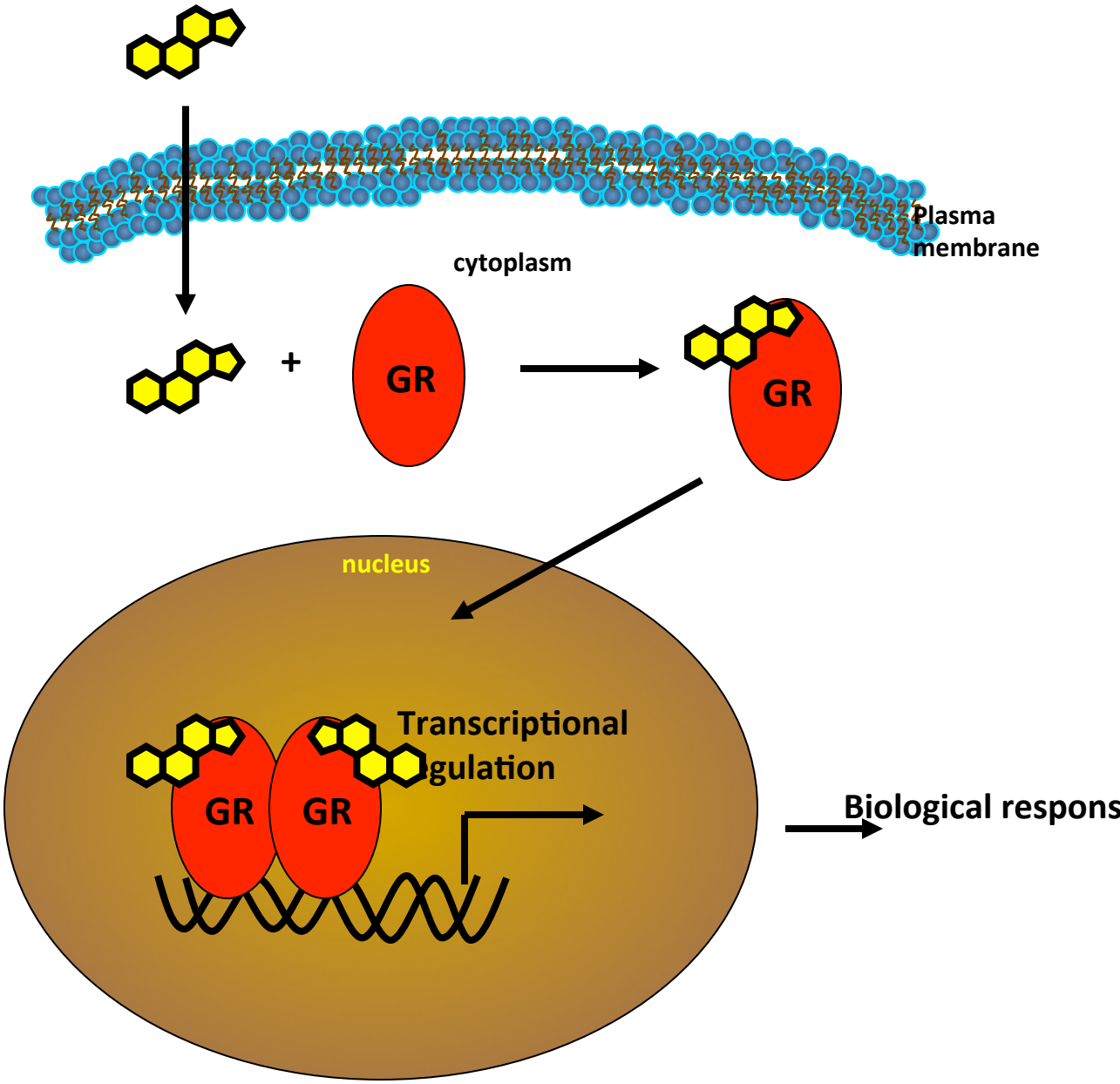
Glucocorticoid Receptor Signaling : Transcriptional regulation

Steroid hormone receptor
(estrogen-/androgen receptor)

Activity controlled by hormone (cortisol/ dexamethasone)

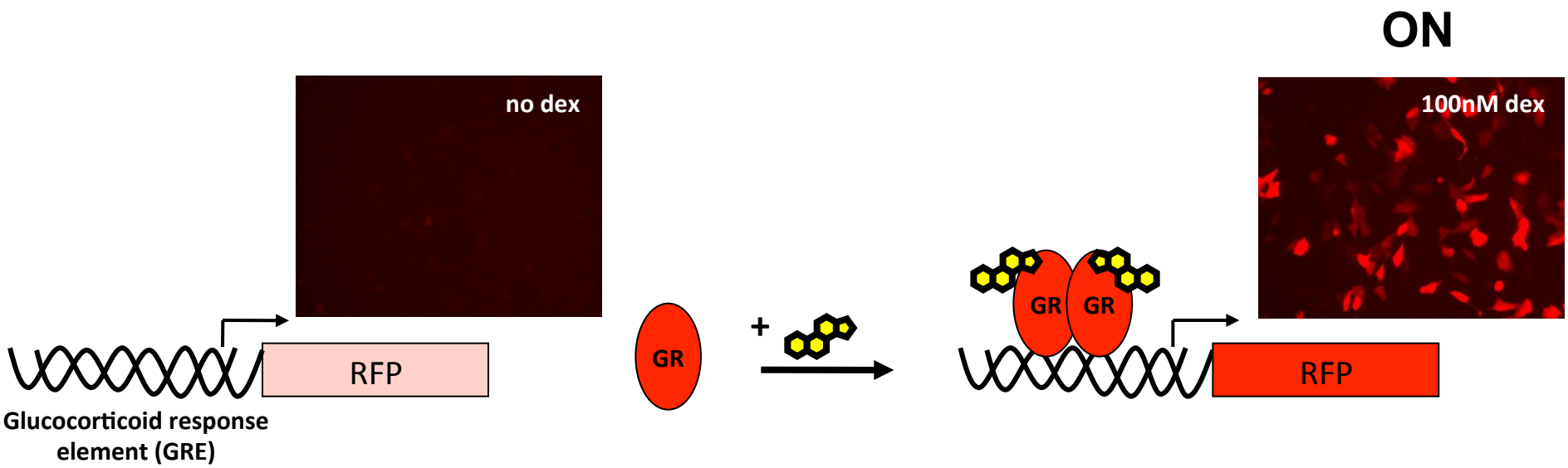
Regulates expression of target genes

Biological response



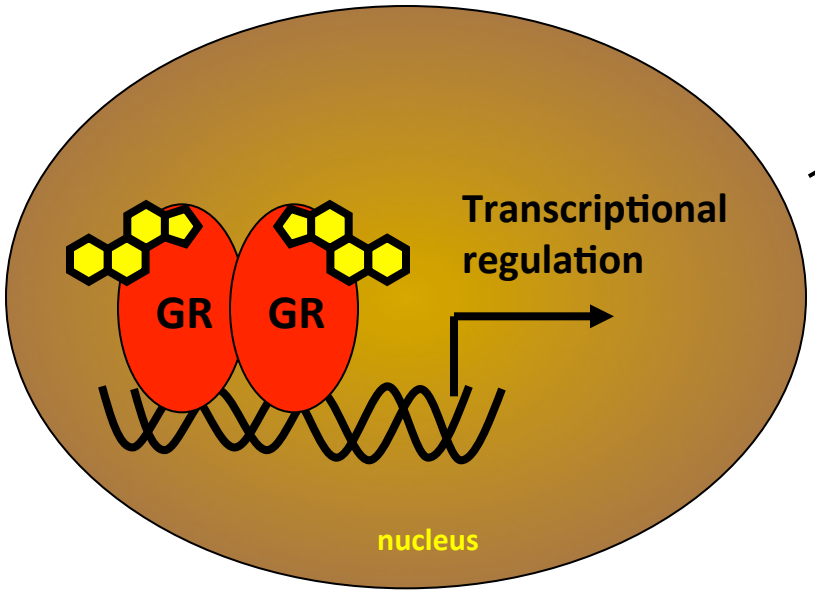
Why use the Glucocorticoid Receptor to study transcription?

- Activity strictly ligand dependent -> switch



•Gerber *et al.*, PNAS 2009

Glucocorticoid Receptor Signaling : Diverse biological responses



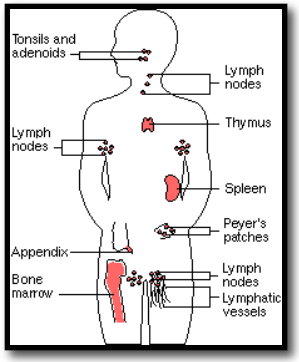
fat specific target genes

Immune spec. target genes

Bone specific target genes

Adipose tissue

Lipolysis, Obesity



Immune Cells

Anti-inflammation

Bone

Reduce bone mass

Examples of integrating experimental and computational approaches:

Tissue specific regulation

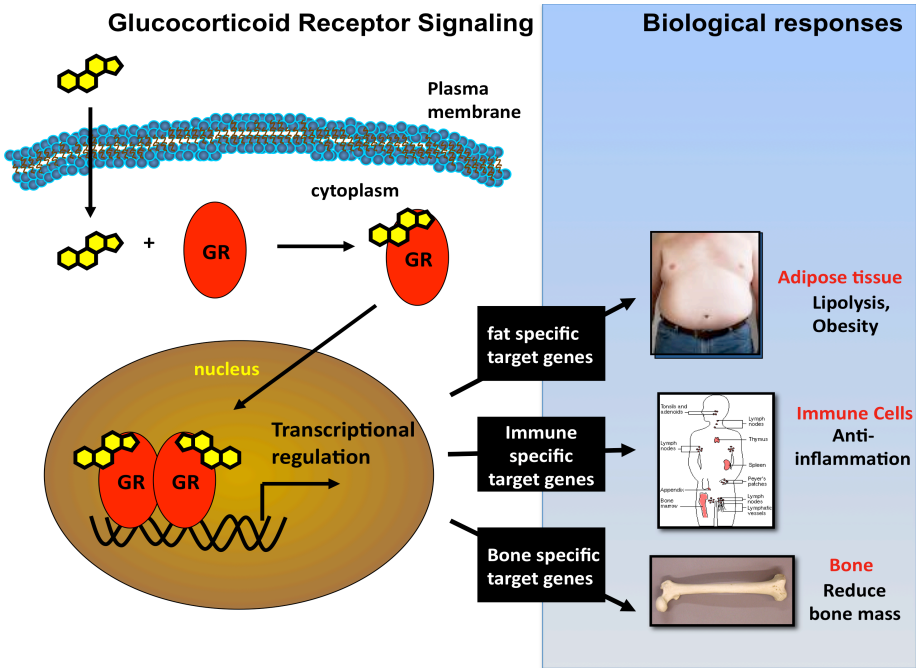
Negative regulatory sequences….?

Getting gene dosage right

Conclusions/Outlook:

Question we're trying to answer:

- Glucocorticoid receptor (GR) expressed throughout the body
- Effects highly tissue specific



Approach: Compare regulated genes across cell lines

U2OS: Bone 

A549: Lung epithelial cells 

NALM-6: preB cells (blood) 

Determine GR target genes cell lines derived from different tissues

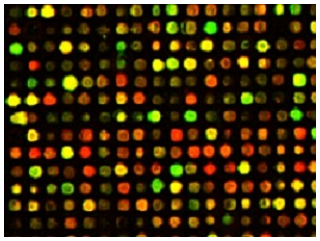
Compare target genes across tissues

Approach:

Experimental part of the experiment.....

Bioinfomatical analysis.....

Treat cells with hormone isolate RNA hybridize wash/scan etc.....



Normalize Quality control data/ statistical analysis/identify cell type (non) specific target genes

Cy5 (red) untreated

Cy3 (green) hormone treated

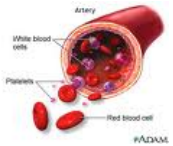
Ratio cy5/cy3 → UP, Down or Unchanged

.....Bioinfomatical analysis.....

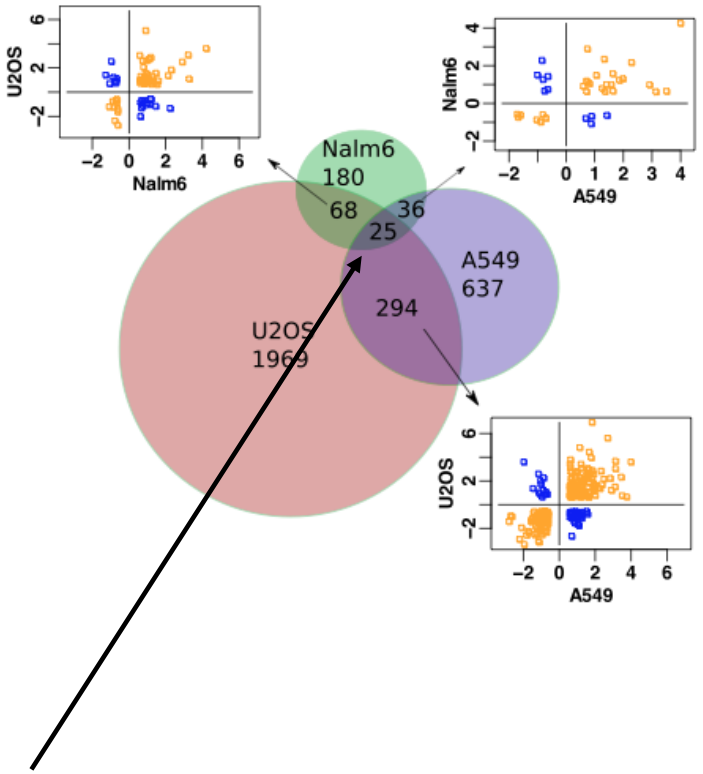
U2OS: Bone

A549: Lung epithelial cells

NALM-6: preB cells (blood)



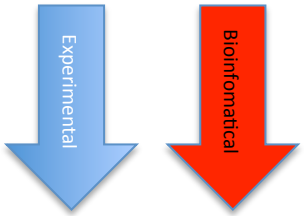
List of genes for each cell type:



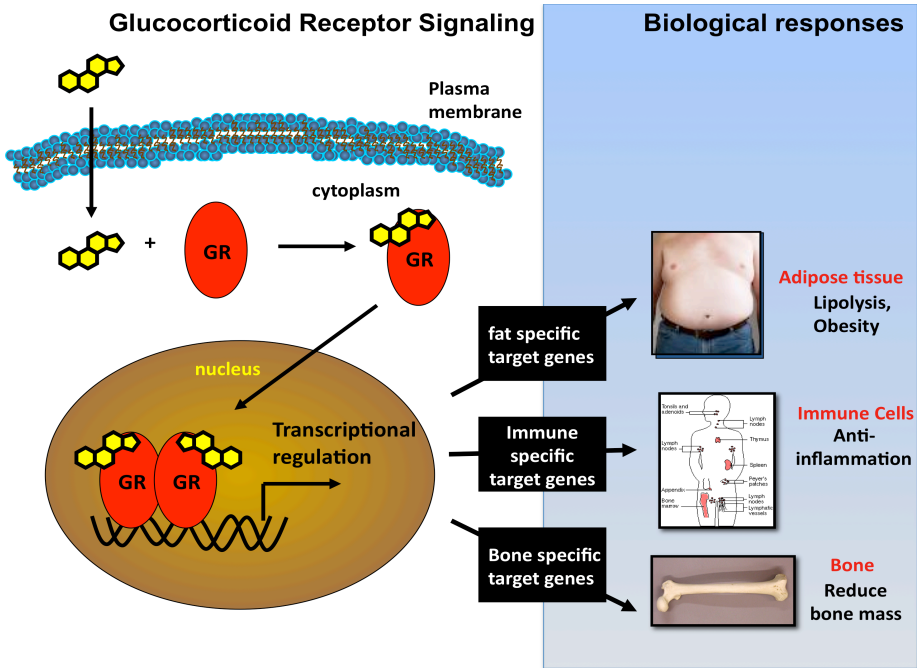
Microarrays show little overlap in transcriptional regulation between 3 cell types

Question we're trying to answer:

- Glucocorticoid receptor (GR) expressed throughout the body
- Effects highly tissue specific

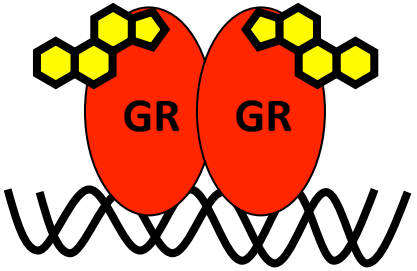


- Because different genes are regulated in different cell types



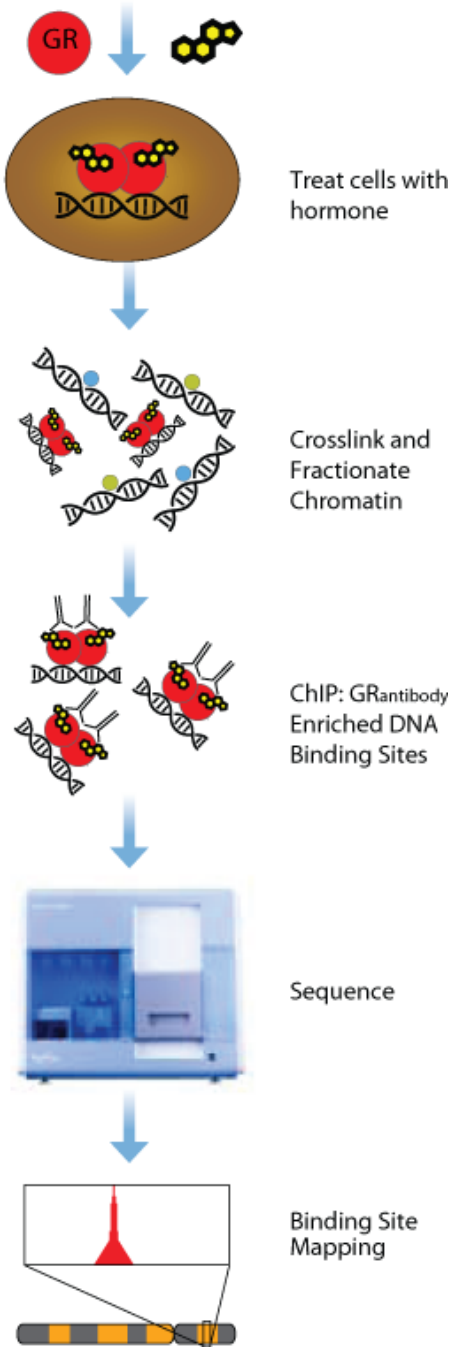
Q: Cause for Cell type/tissue specific transcriptional regulation by GR ?

Genomic DNA Binding?



Determine genome-wide where GR binds by

Chromatin Immunoprecipitation (ChIP) in:



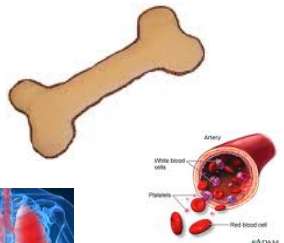
Determine GR binding sites in different cell types

→ **ChIP-seq**

Experimental part of the experiment.....

Bioinfomatical analysis.....

Treat cells with hormone
Perform ChIP etc.....

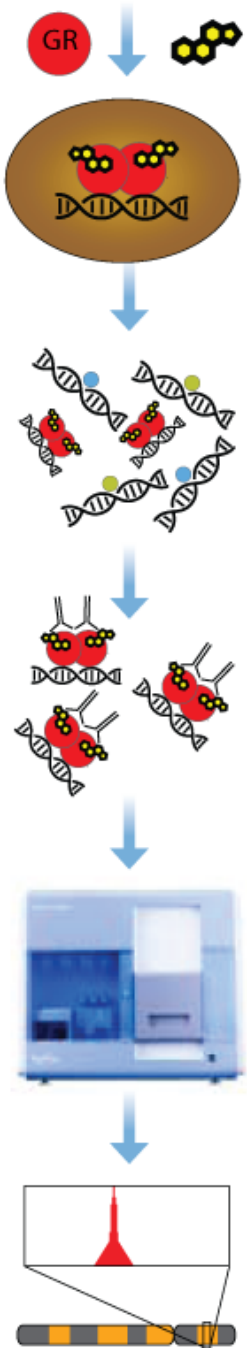


Convert data to sequence
Quality control data/ map to genome/call peaks etc etc.....

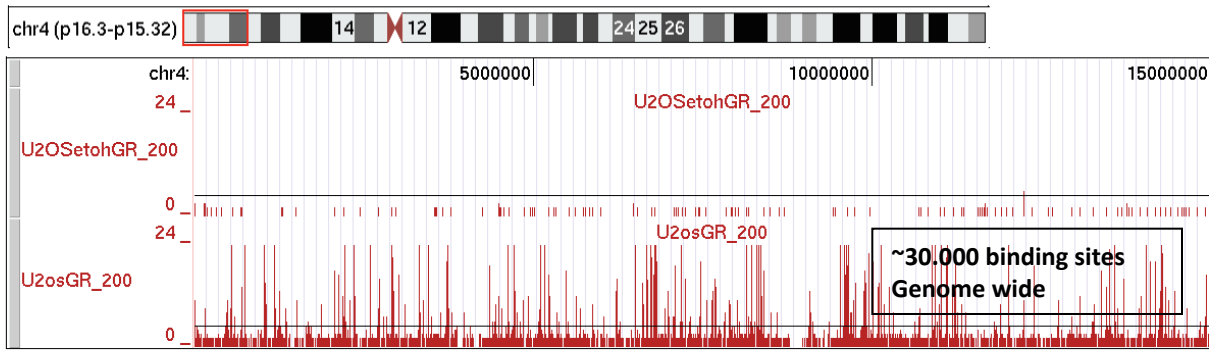
Input: library of immunoprecipitated material

Output: Lots of images that can be converted into millions of short sequence reads

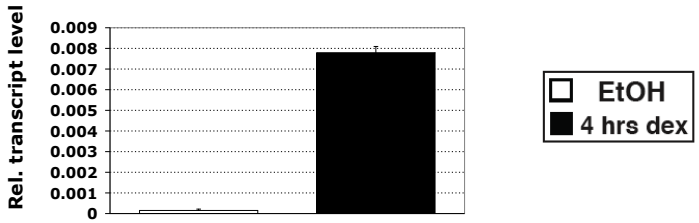
Determine GR binding sites in different cell types → ChIP-seq



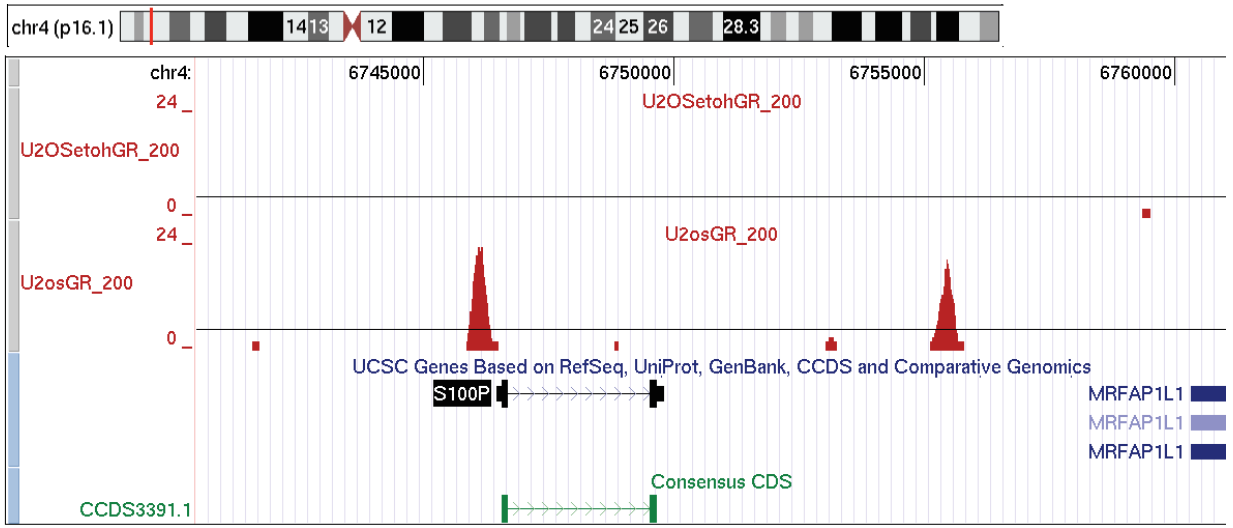
+



GR-target gene: S100p



+



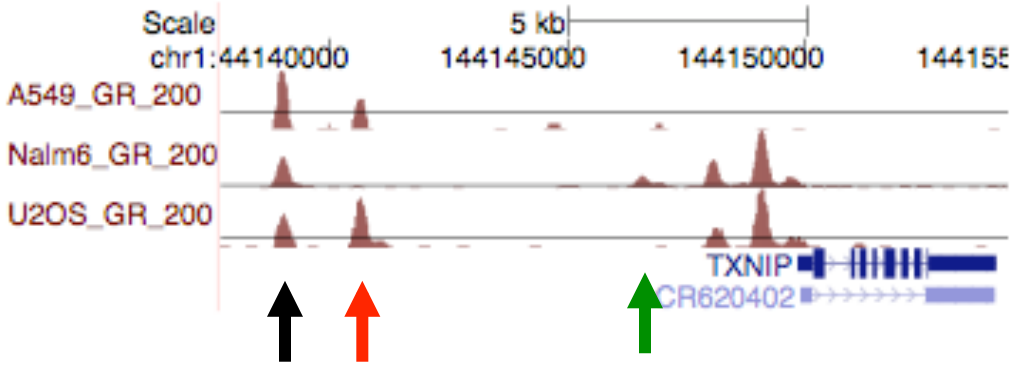
.....Bioinfomatical analysis.....

List of peaks for each cell type:

U2OS: Bone

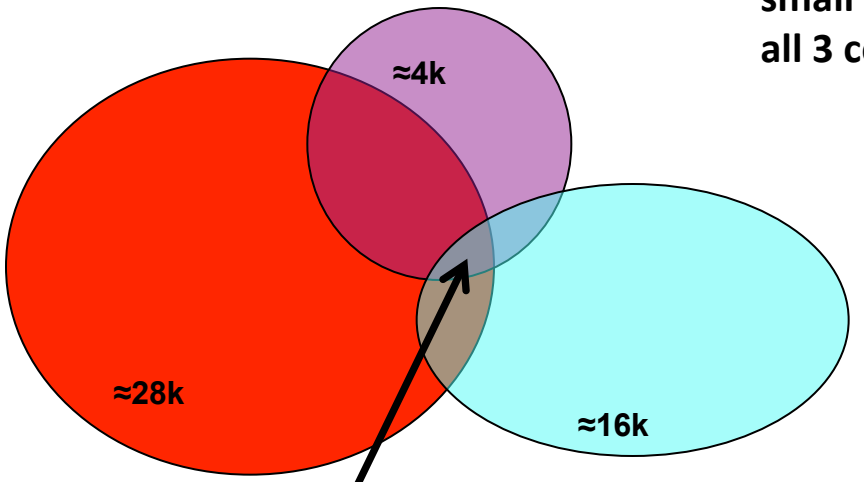
A549: Lung epithelial cells

NALM-6: preB cells (blood)



.....Bioinfomatical analysis.....

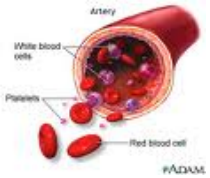
small overlap between
all 3 cell lines



U2OS

A549

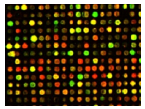
Nalm6



Little overlap of genomic
binding of GR

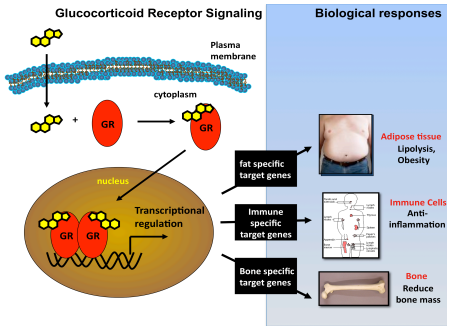
Question we're trying to answer:

- Glucocorticoid receptor (GR) expressed throughout the body
- Effects highly tissue specific



Experimental

Bioinformatical



- Different genes are regulated in different cell types

Experimental

Bioinformatical

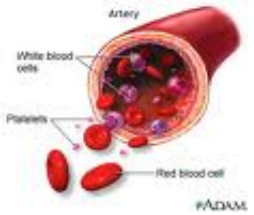
- Cell-type specific binding of transcription factor

Bioinformatical

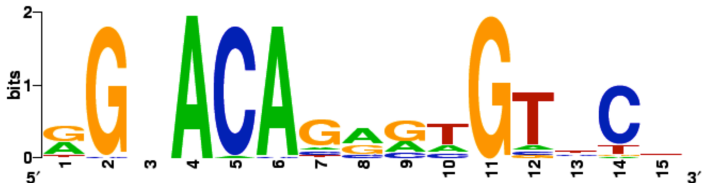
What causes cell-type specific binding?

.....Bioinfomatical analysis.....

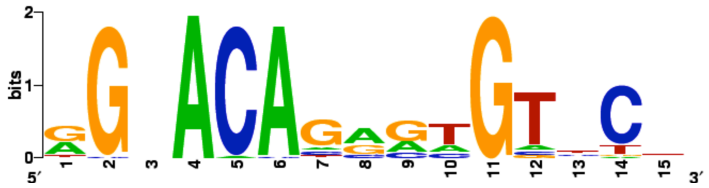
What causes cell-type specific binding?



Compare enriched binding site motifs between cell types (peak-motifs / MEME/AMADEUS….)

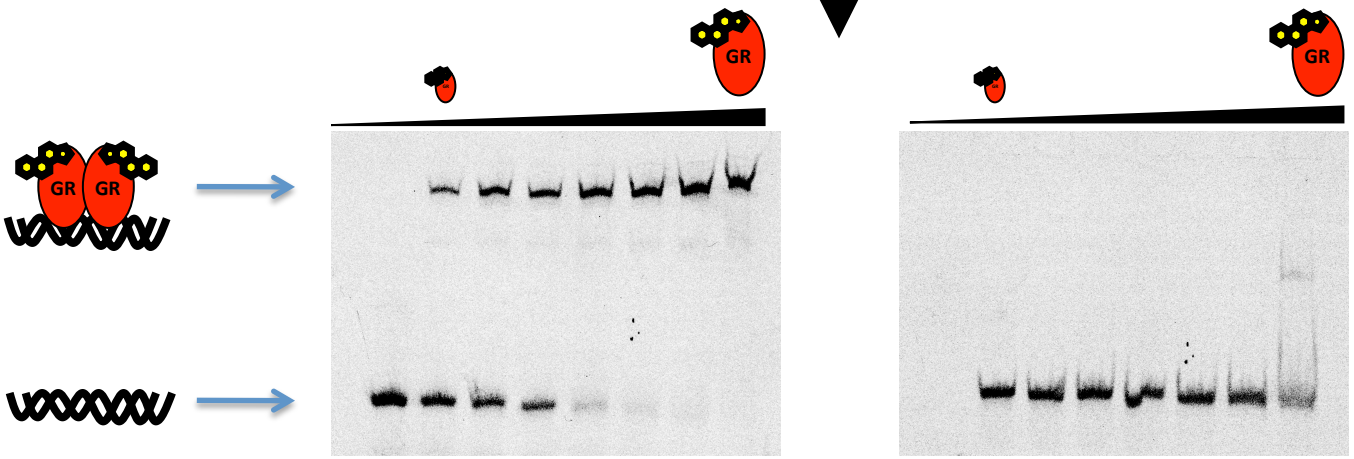


Does transcription factor we study bind to this sequence?????



Binding motif for Glucocorticoid receptor

Gelshifts (EMSA)

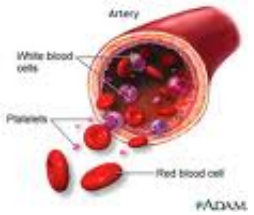


AGAACAtttTGTTCT (Binding motif)

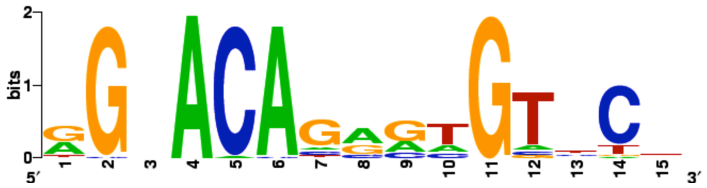
randomized control sequence

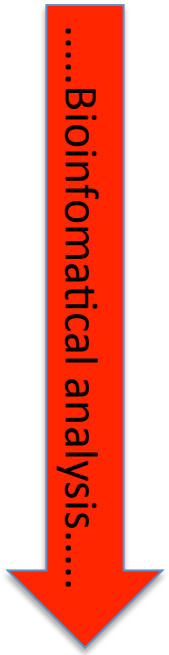
.....Bioinfomatical analysis.....

What causes cell-type specific binding?



Compare enriched binding site motifs between cell types (Peak-motifs/ MEME/AMADEUS….)

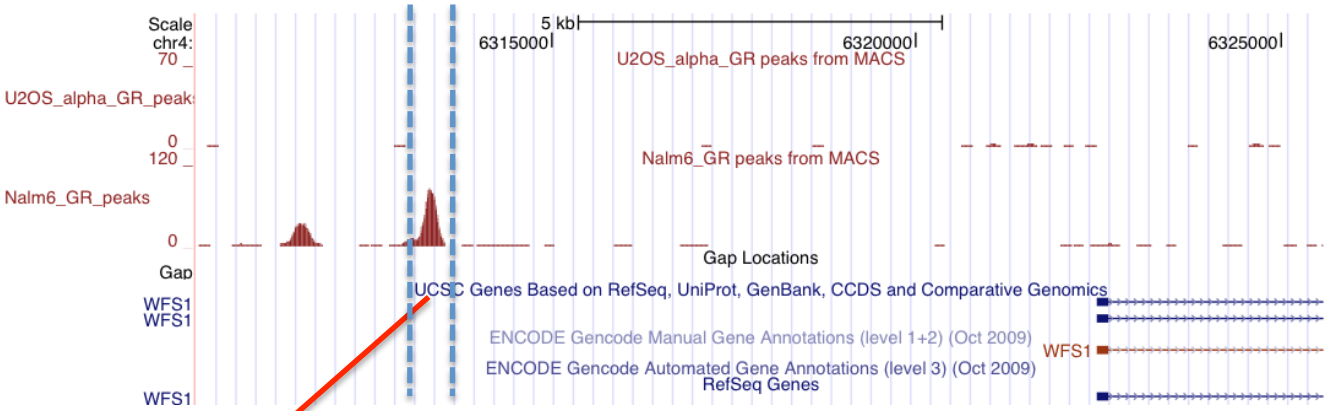


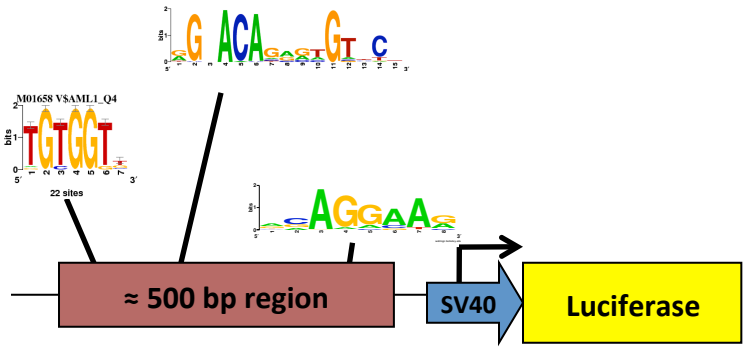


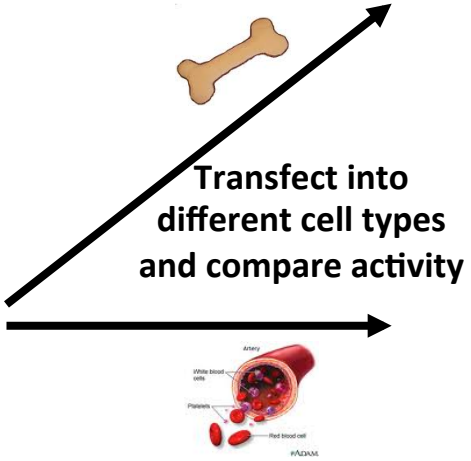
Hypothesis: Identified motif contributes to tissue specific regulation by GR

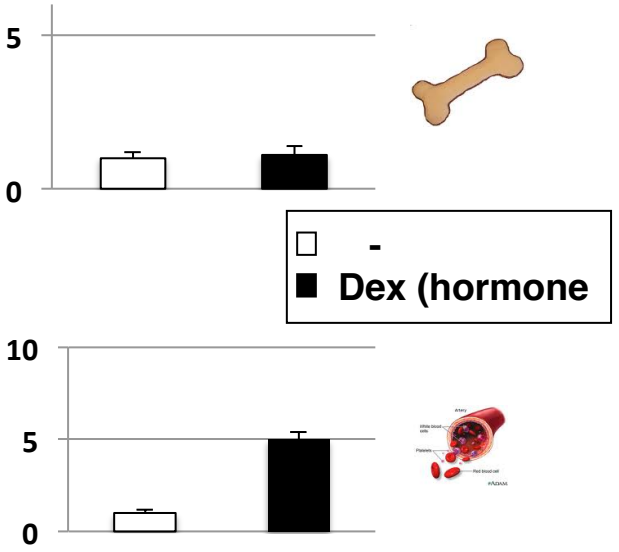
Experimental part of the experiment.....

Test if cis elements can recapitulate cell-type specific regulation



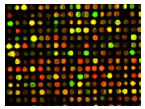






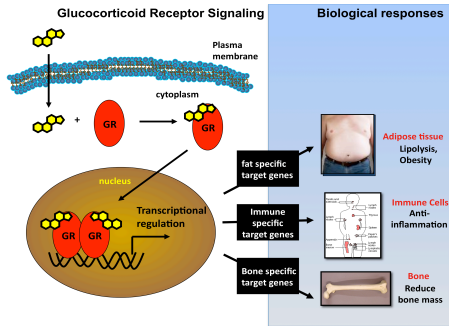
Question we're trying to answer:

- Glucocorticoid receptor (GR) expressed throughout the body
- Effects highly tissue specific



Experimental

Bioinfomatical



- Likely because different genes are regulated in different cell types

Experimental

Bioinfomatical

- Likely because cell-type specific binding of transcription factor

What causes cell-type specific binding?

Bioinfomatical

- Different sequence motifs

Experimental

- Test/validate role of identified sequence

Other examples of how data sets can be analyzed the possibilities are endless……..

→This is (I think) the interface where biologists & bioinformatitions need to meet to prioritize

What can be done ←→ What analysis is usefull.....

- Correlate binding and regulation of nearest gene

- Integrate other inputs (epigenetic landscape)

- Motif search for subsets of peaks (e.g. peaks with highest binding/ involved in metabolism……..)

- Etc etc etc

Examples of integrating experimental and computational approaches:

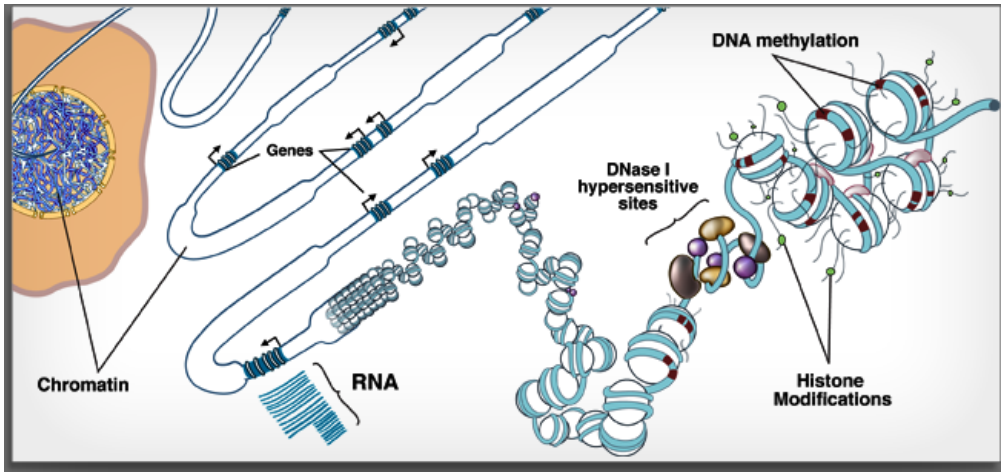
Tissue specific regulation

Negative regulatory sequences….?

Getting gene dosage right

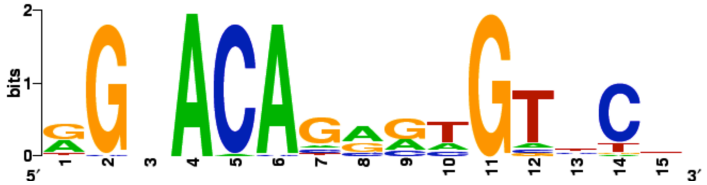
Conclusions/Outlook:

How does the glucocorticoid receptor “know” where to go in the genome ?

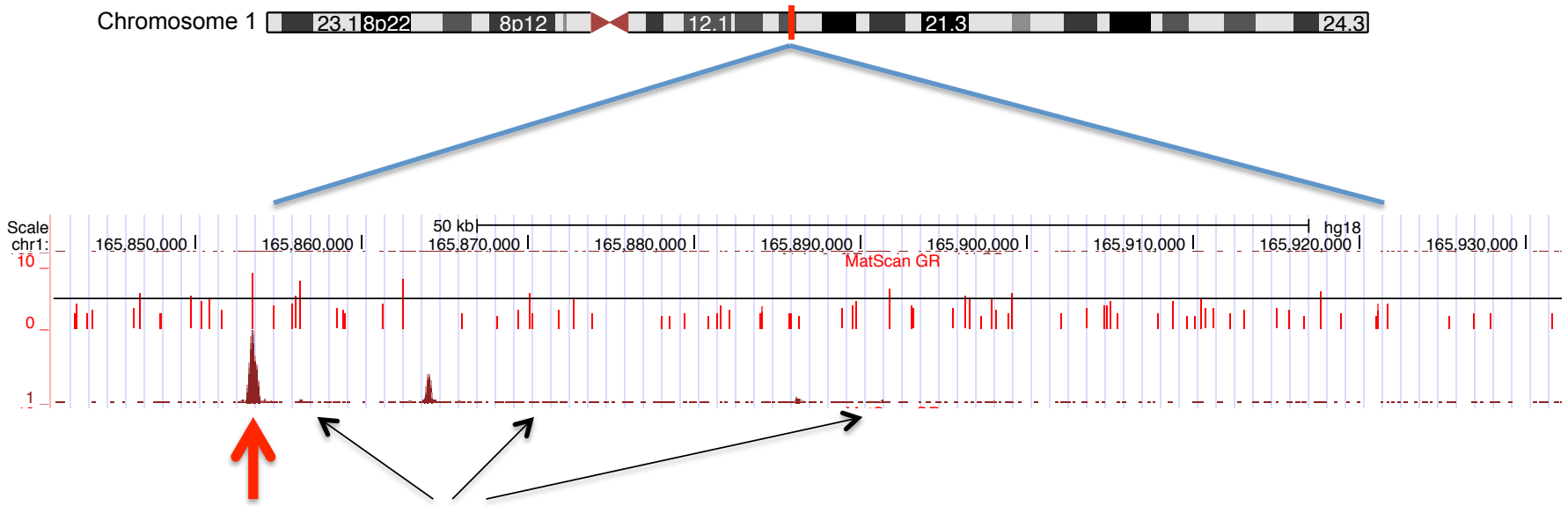


DNA sequence and genomic binding by GR

ChIP-seq in IMR90 cells (primary human epithelial lung fibroblast)

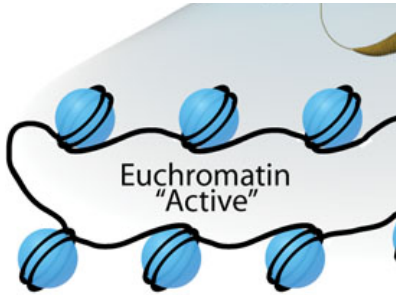


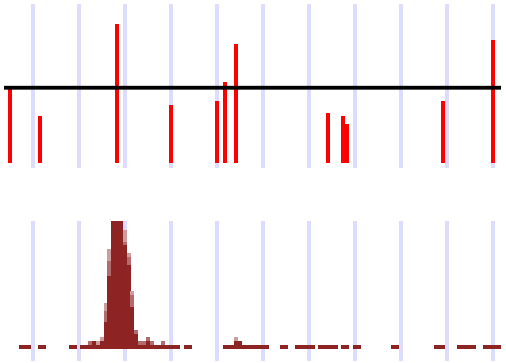
Sequence found in genome more than once every 1000bp



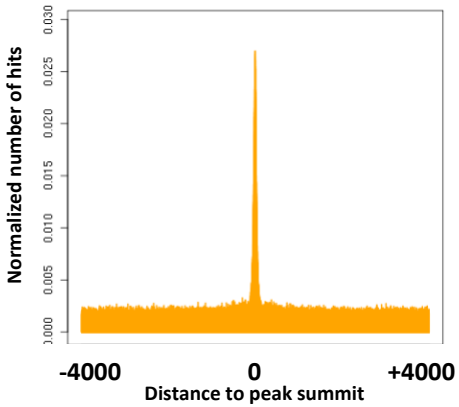
Question: Why binding to this site (and not to others) ?

Signals positively correlating with GR binding

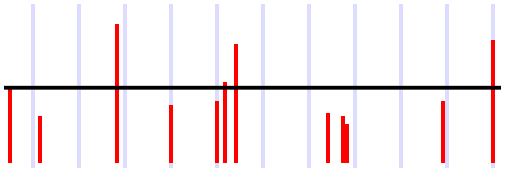


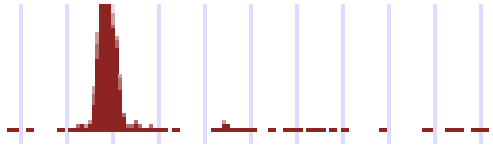


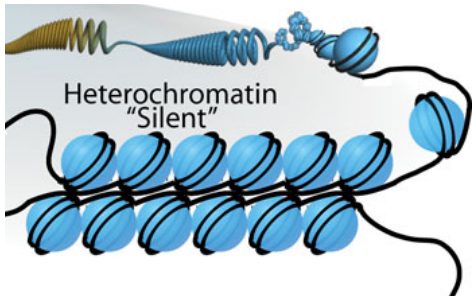
John et al., *Mol. Cell* (2008).



Signals negatively correlating with GR binding



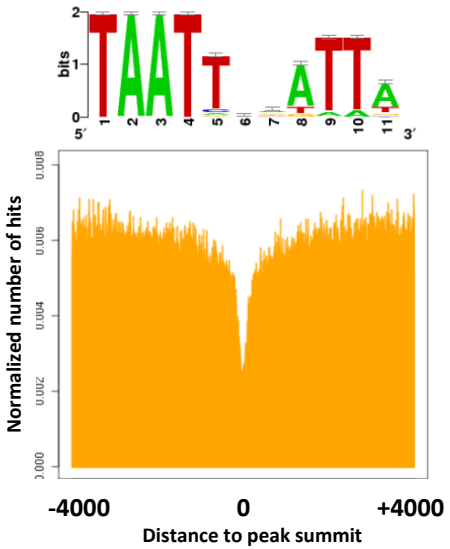


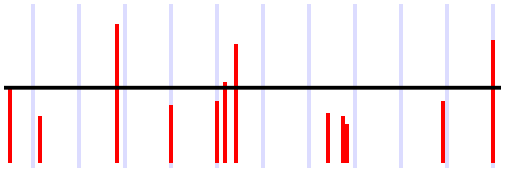


Sequence signals ?

Jonas Telorac / Morgane Thomas-Chollier

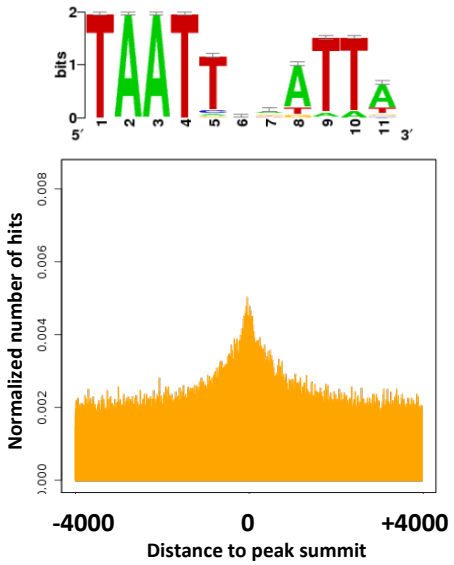
Signals negatively correlating with GR binding

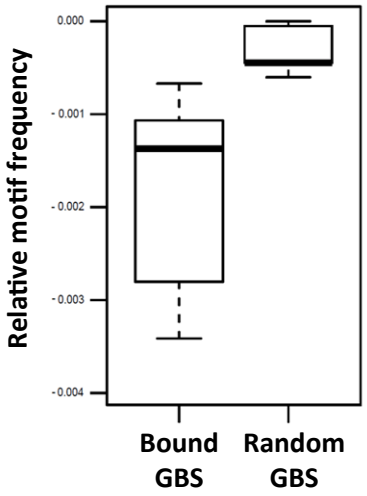




Peak-motifs: Looks for sequence motifs with significant changes (both enriched & depleted)

Thomas-Chollier et al., *N.A.R.* (2011).

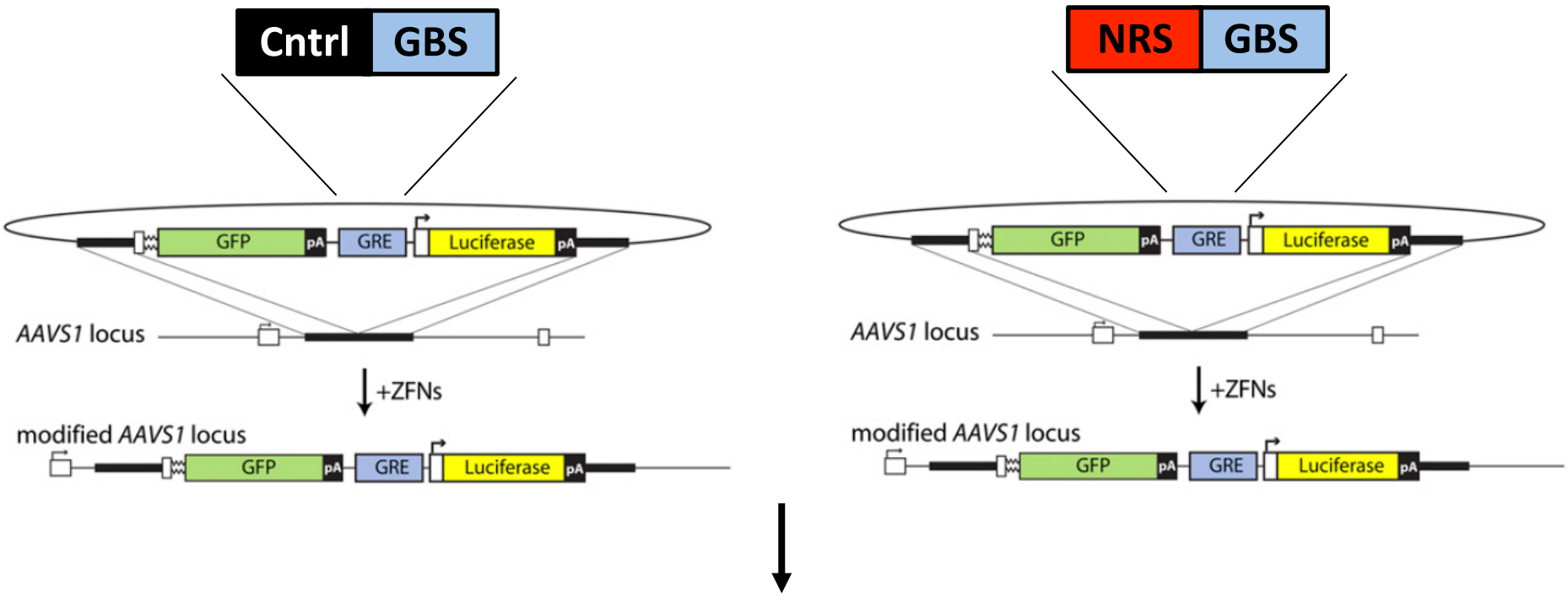




Jonas Telorac / Morgane Thomas-Chollier

Testing these Negative Regulatory Sequences (NRSs)

Targeted genomic integration using ZFNs

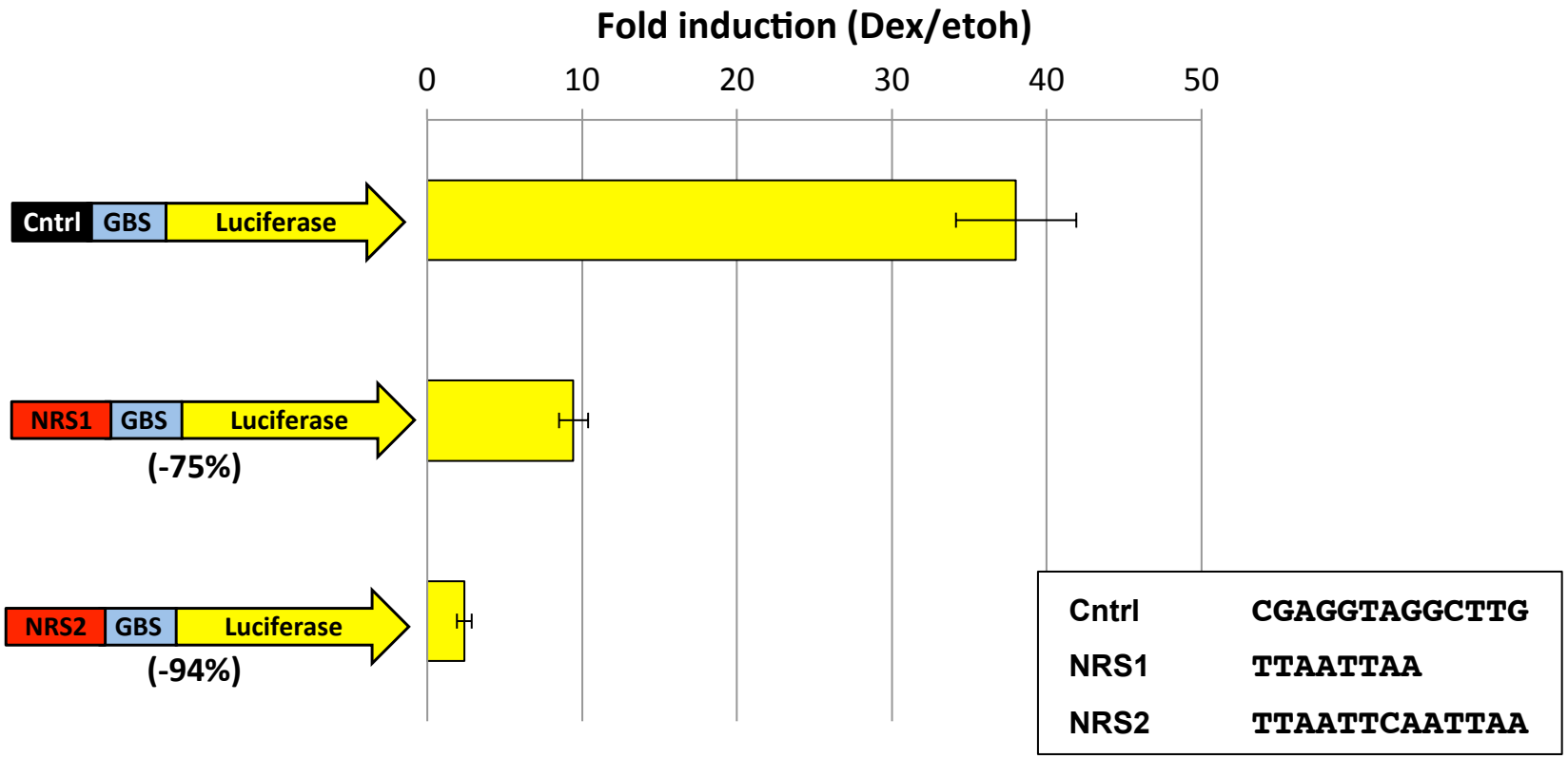


Isogenic cell lines with integrated NRS reporters

Dekelver, Meijsing et al., *Genome Research* (2010).

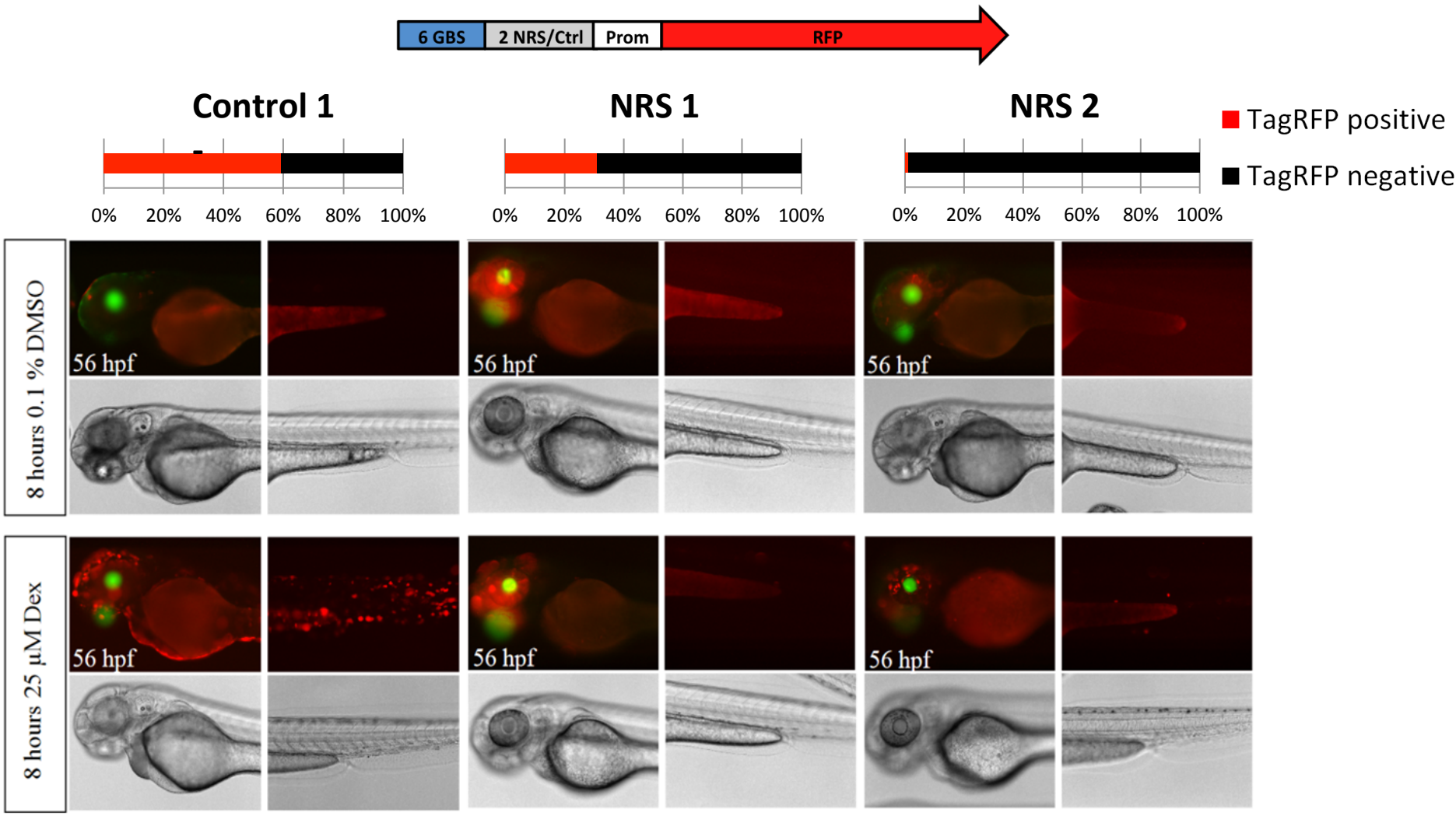
NRSs interfere with GR activity

Transcriptional regulation:



- NRSs interfere with transcriptional regulation

Conservation / tissue specificity??

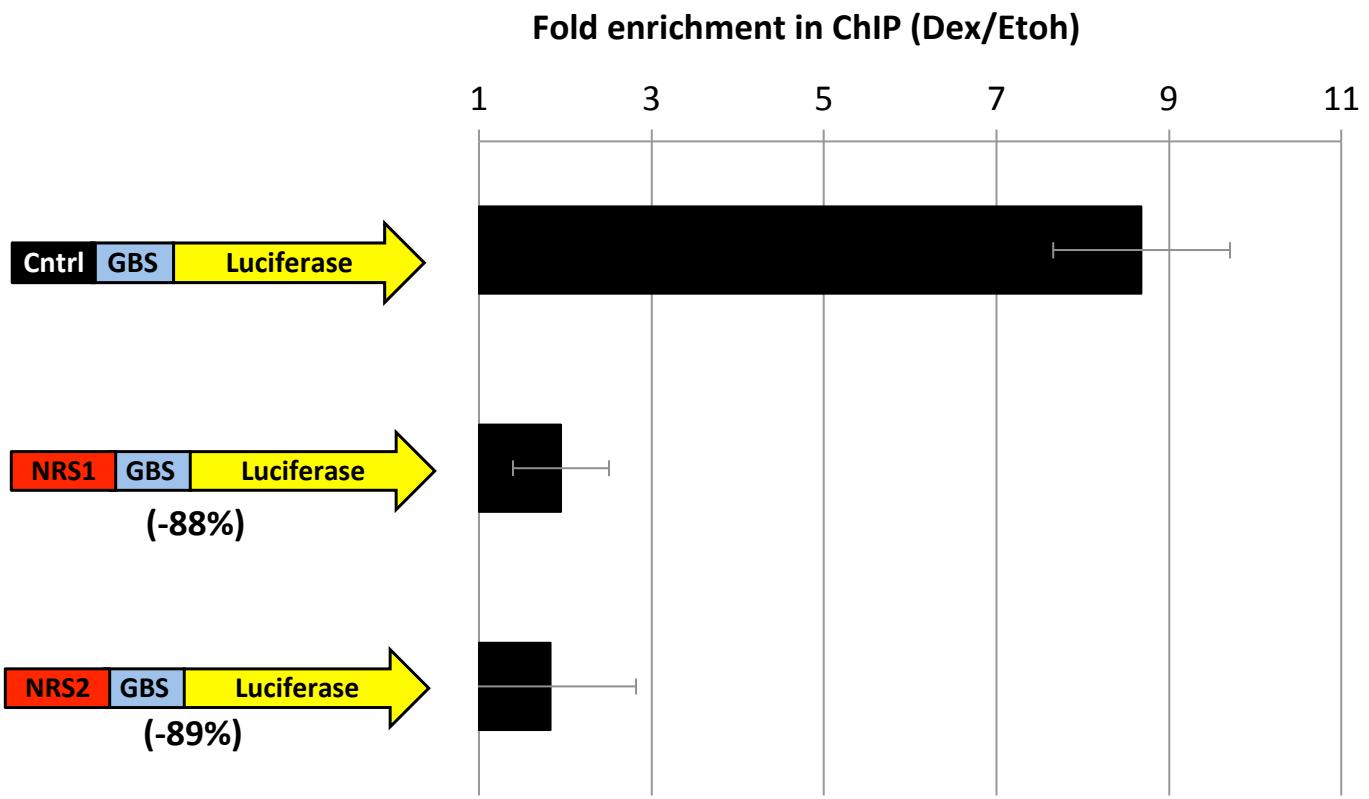


- Conserved & effect observed in all tissues

Sergey Prykhozhij/ Jonas Telorac

NRSs disrupt DNA binding by GR

Genomic binding (ChIP):

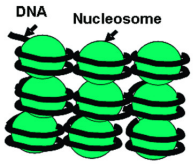


- NRSs interfere with DNA binding by GR

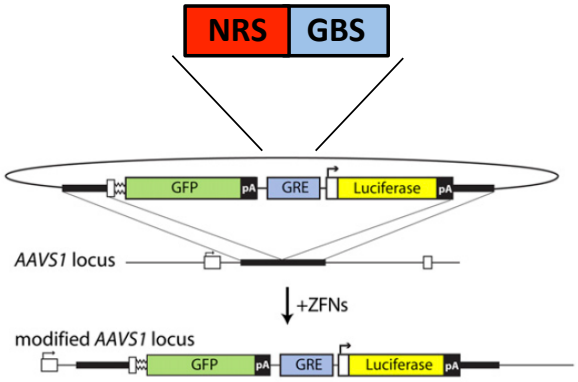
Mechanism?

Chromatin accessibility?

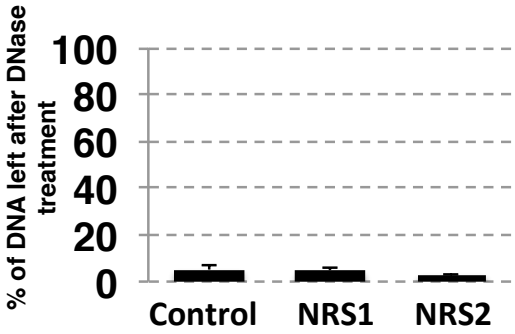
FKBP5 control region
DNAseHS "open":

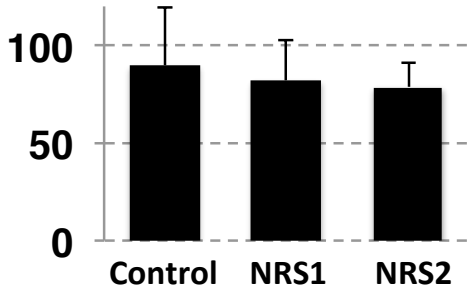


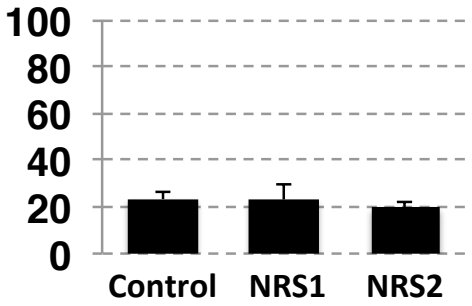
IGFBP1 control region
DNAseHS "closed":



NRS-GBS-Luc region
Similar "open-ness"





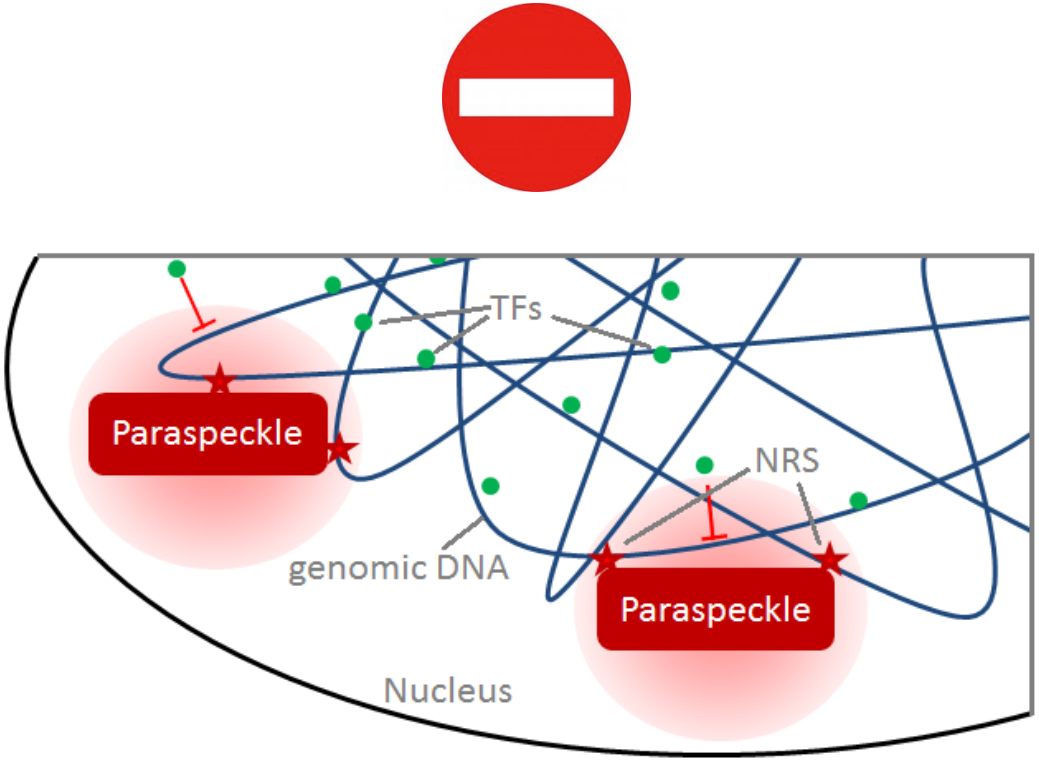


- Chromatin accessibility does not explain reduced binding

Mechanism?

- NRSs associate with paraspeckle components (mass-spec)
- Knockdown of paraspeckle components (partially) reverse repressive effect of NRSs

Sub-nuclear positioning ??



Examples of integrating experimental and computational approaches:

Negative regulatory sequences….?

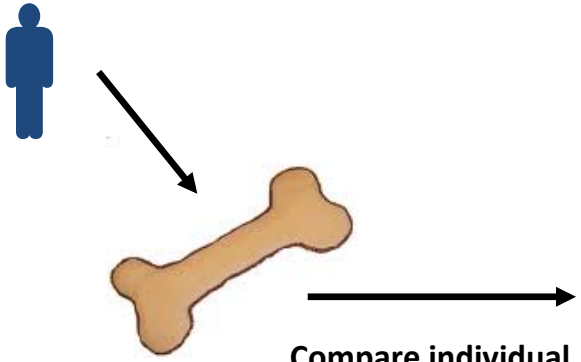
Tissue specific regulation

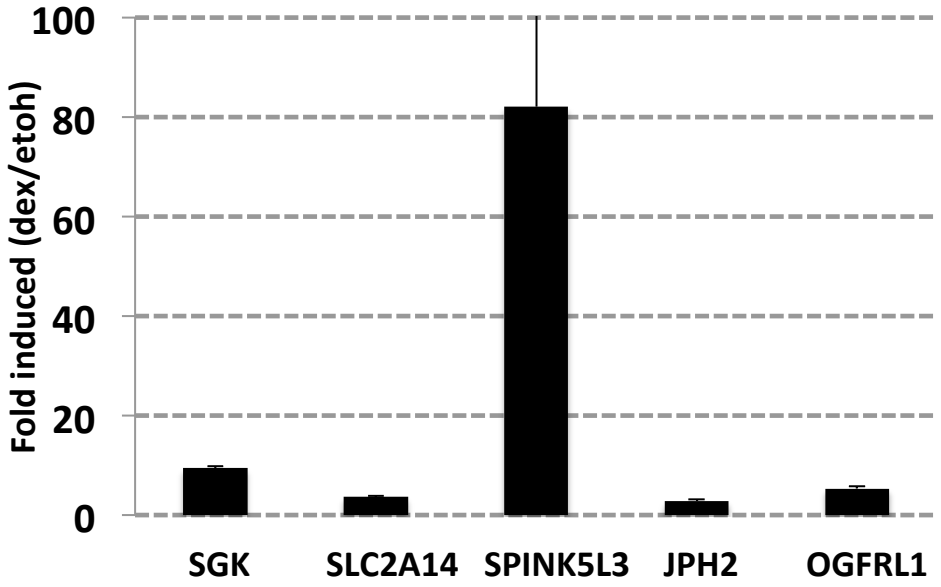
Getting gene dosage right

Conclusions/Outlook:

Part-II

- Glucocorticoid receptor (GR) expressed throughout the body
- Effects highly tissue specific
- Effects gene specific, expression fine-tuned for individual genes within a tissue





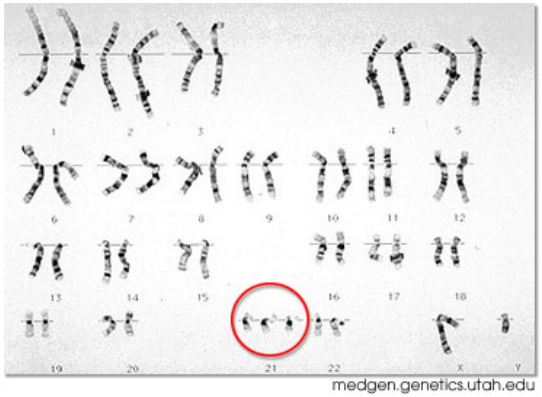
Getting gene dosage just right is important

Too much: Trisomy 21 (down syndrome)

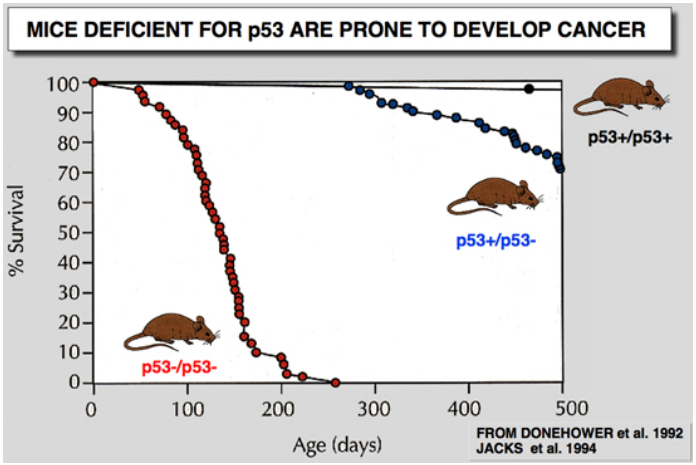
(X-inactivation woman)

Copy number variation linked to various disease:

autism, schizophrenia, systemic lupus erythematosis, Crohn's disease and psoriasis



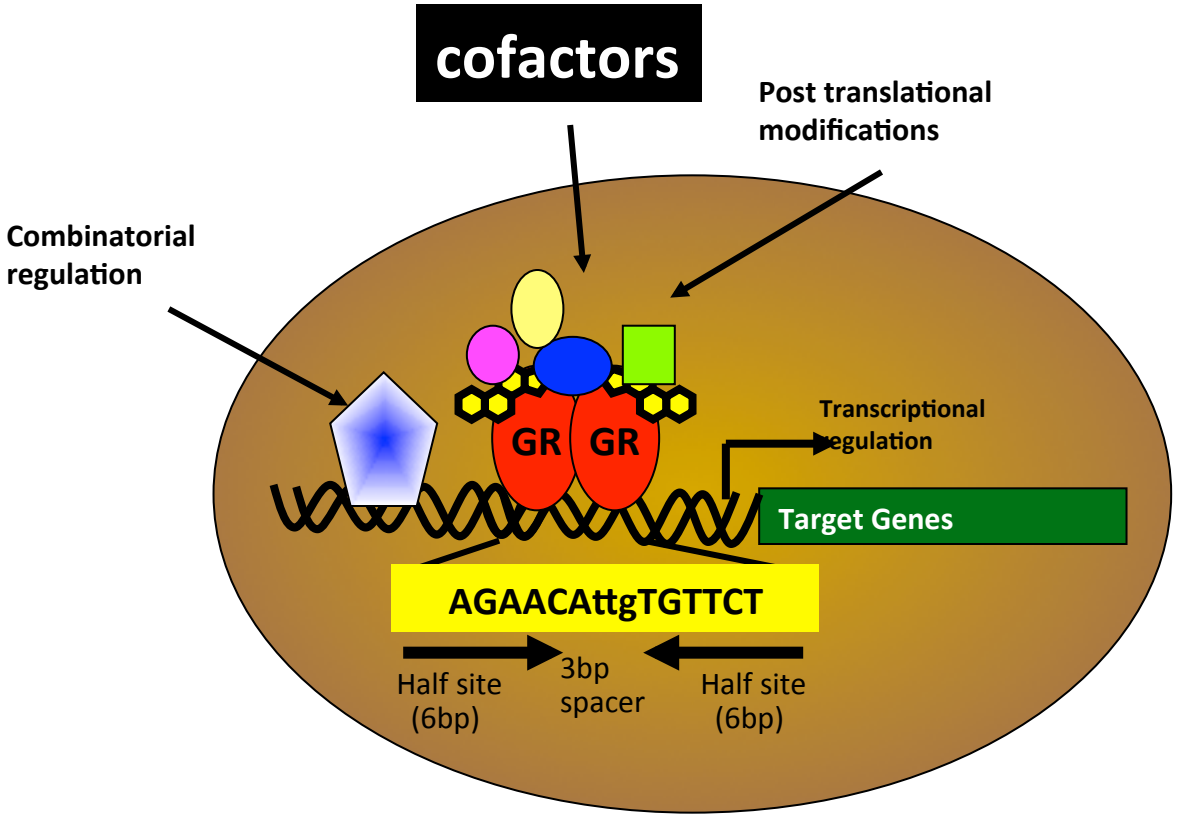
Too little: p53 and cancer



PartII:

Q: How is activity fine-tuned for individual genes?

A: Several inputs are integrated

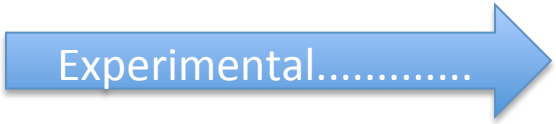


Inputs that specify:

→ Which genes to regulate (in each cell types/tissues)

→ **By how much (within individual cell types/tissues)**

Objective: Identify genes that modulate the activity of GR



Experimental: Y2H for interaction partners etc etc

Alternative: Data mining and bioinformatical analysis to identify candidates….

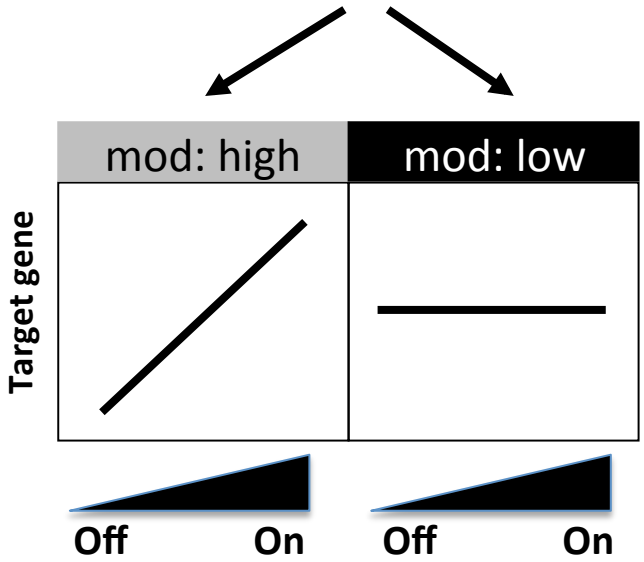
Objective: Identify modulators that modulate of GR activity

Approach: Identify genes whose expression correlates with activity of GR

Compare regulation of GR target genes in cell lines derived from different tissues

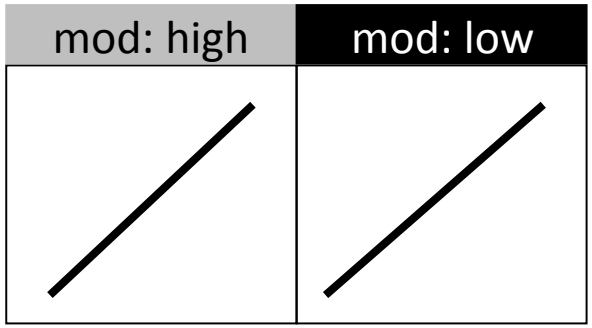
Determine if expression of candidate modulators correlates with activity of receptor

Sort experiments according to the expression level of candidate modulator genes

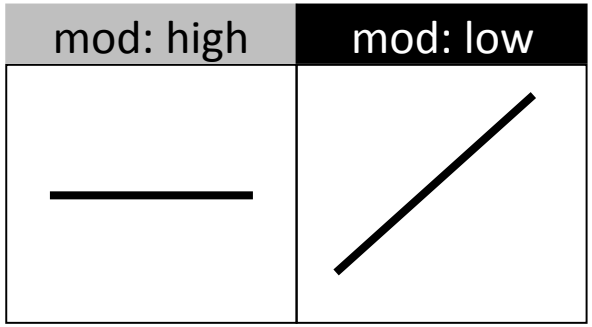


Transcription factor (GR)

Positive modulator



Low mutual information → not a modulator



Negative modulator

Akdes Serin

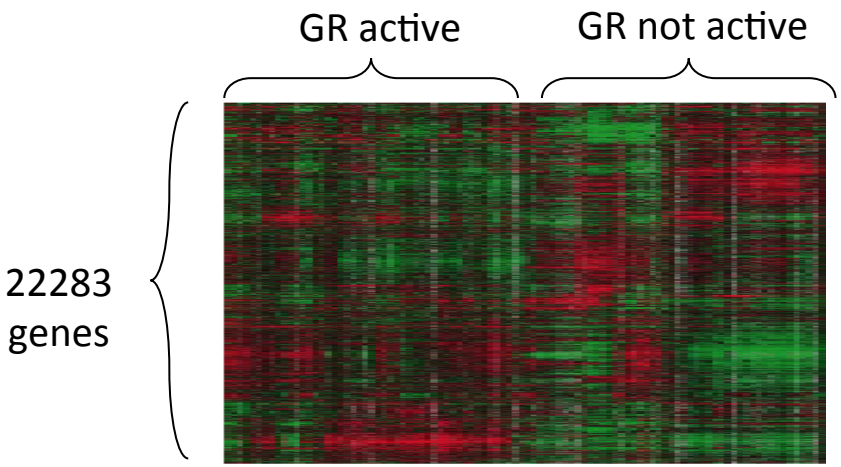
Modulators of GR activity (with **Akdes Serin**/**Katja Borzym**/Matthias Heinig and Edda Einfeldt)

Data sets: **C-map data Broad institute:**

Tested the effext of > 6000 compounds on different cell lines

Several of these compounds are glucocorticoid hormones

180 data sets from 4 cell lines used for analysis

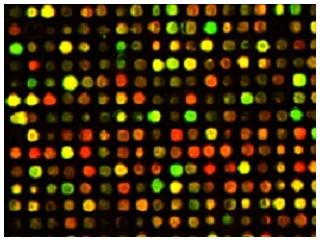


Experimental part of the experiment.....

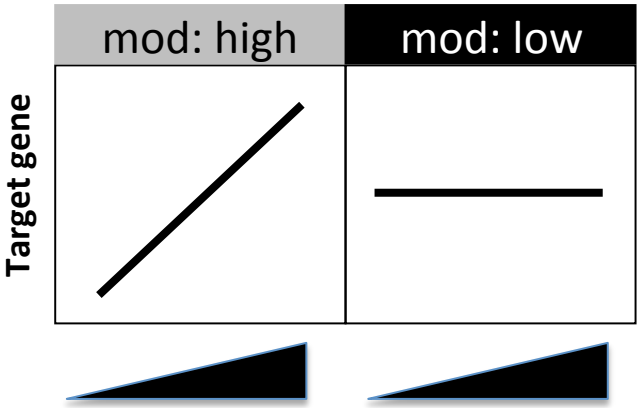
Bioinfomatical analysis.....

CONNECTIVITY MAP 02

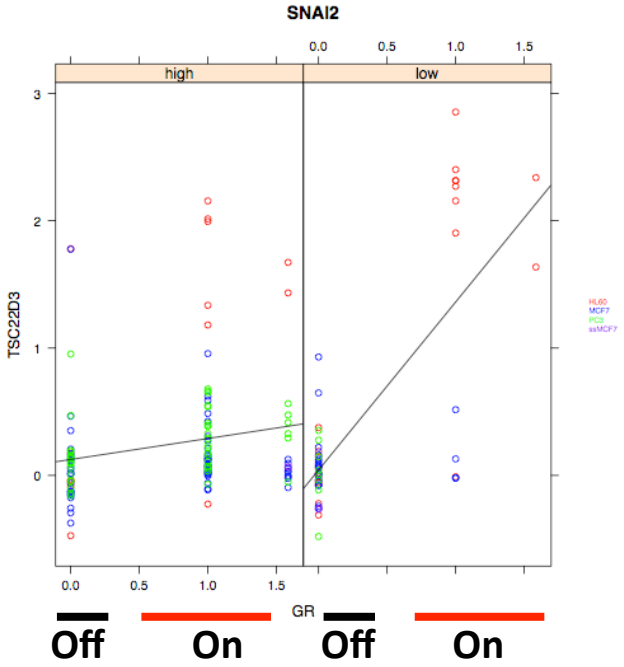
Treat cells with hormone isolate RNA hybridize wash/scan etc.....

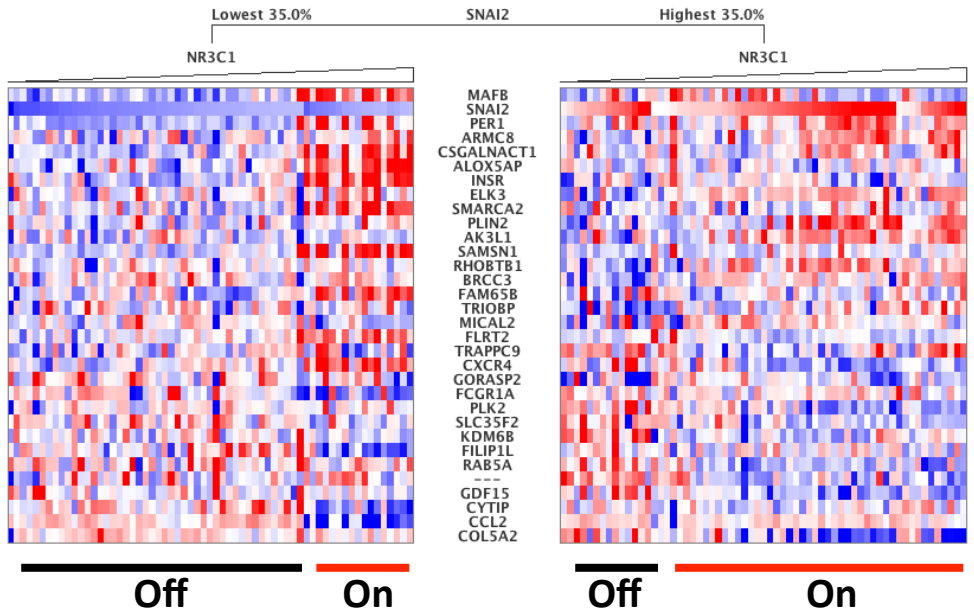


Normalize Quality control data/ statistical analysis/ Mutual information analysis



.....Bioinfomatical analysis.....





Wang et al., Nature Biotechnol. 2009

F-test:

Snai2 (Slug) negative

JMJD4 positive, putative histone demethylase

S100p negative

FDFT1 negative*

Mindy:

Snai2 negative

KLF9 **

Per1*

RBL1 *

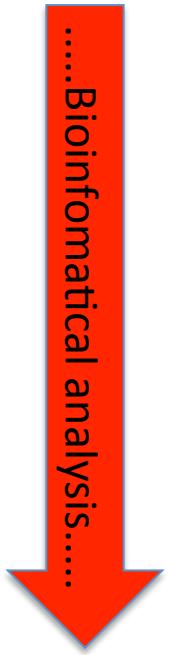
MAFB

TRIM66 **

-> Different methods identify overlapping and non-overlapping candidates

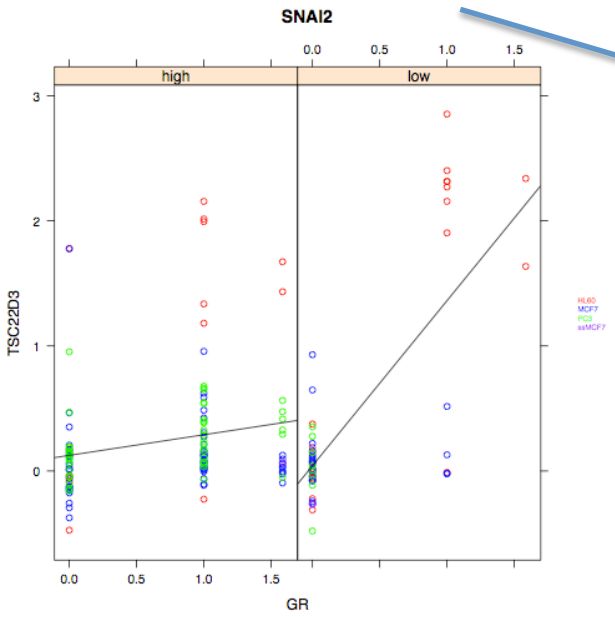
* Known interaction with GR

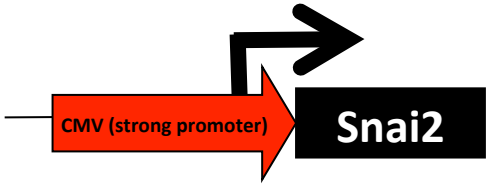
** related proteins interact



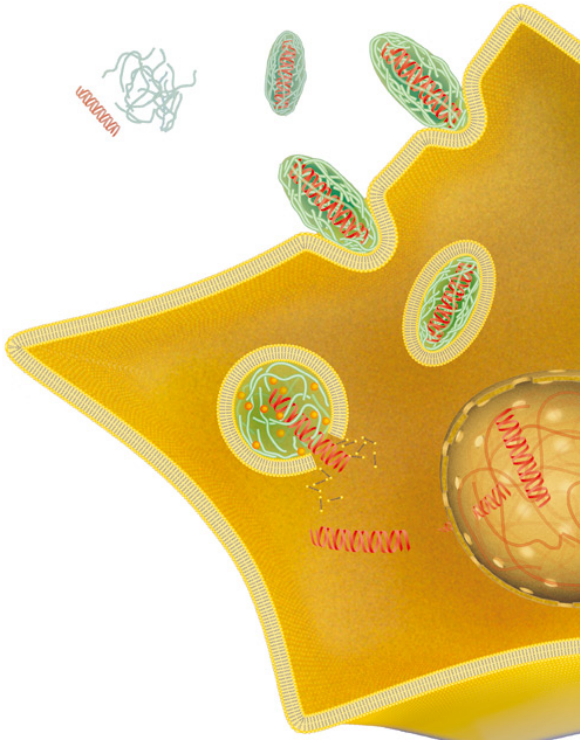
Hypothesis: Snai2 can modulate activity of GR

How can we test if Snai2 acts as a negative modulator of GR activity?





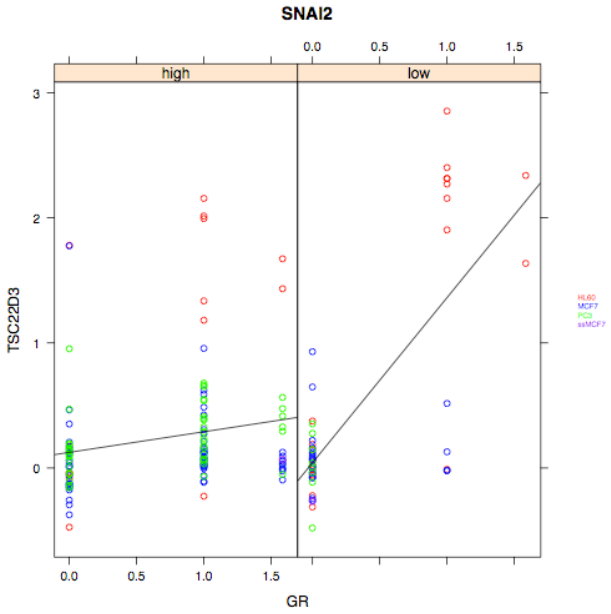
Artificially change Snai2 levels in cells

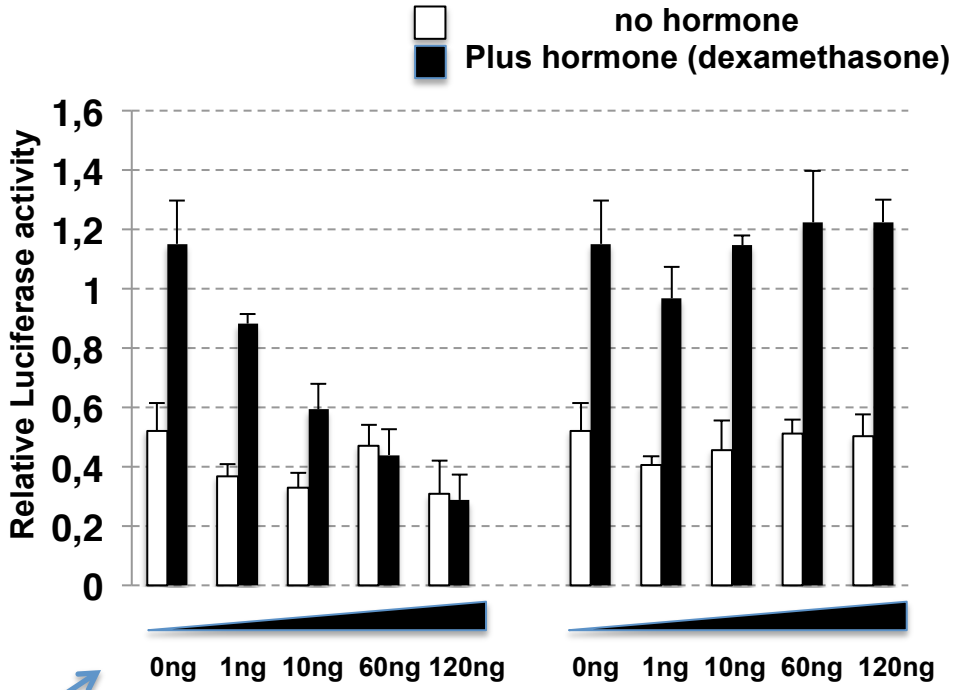


http://www.baria.cz/jetprime-dna-sirna-transfection-reagent

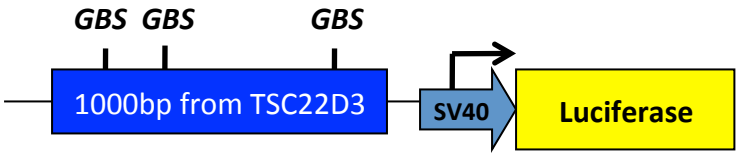
Assay effect of changed Snai2 levels?

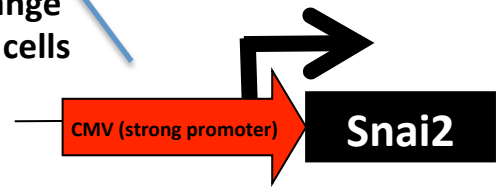
Can/does Snai2 act as a negative modulator of GR activity?





Artificially change
Snai2 levels in cells



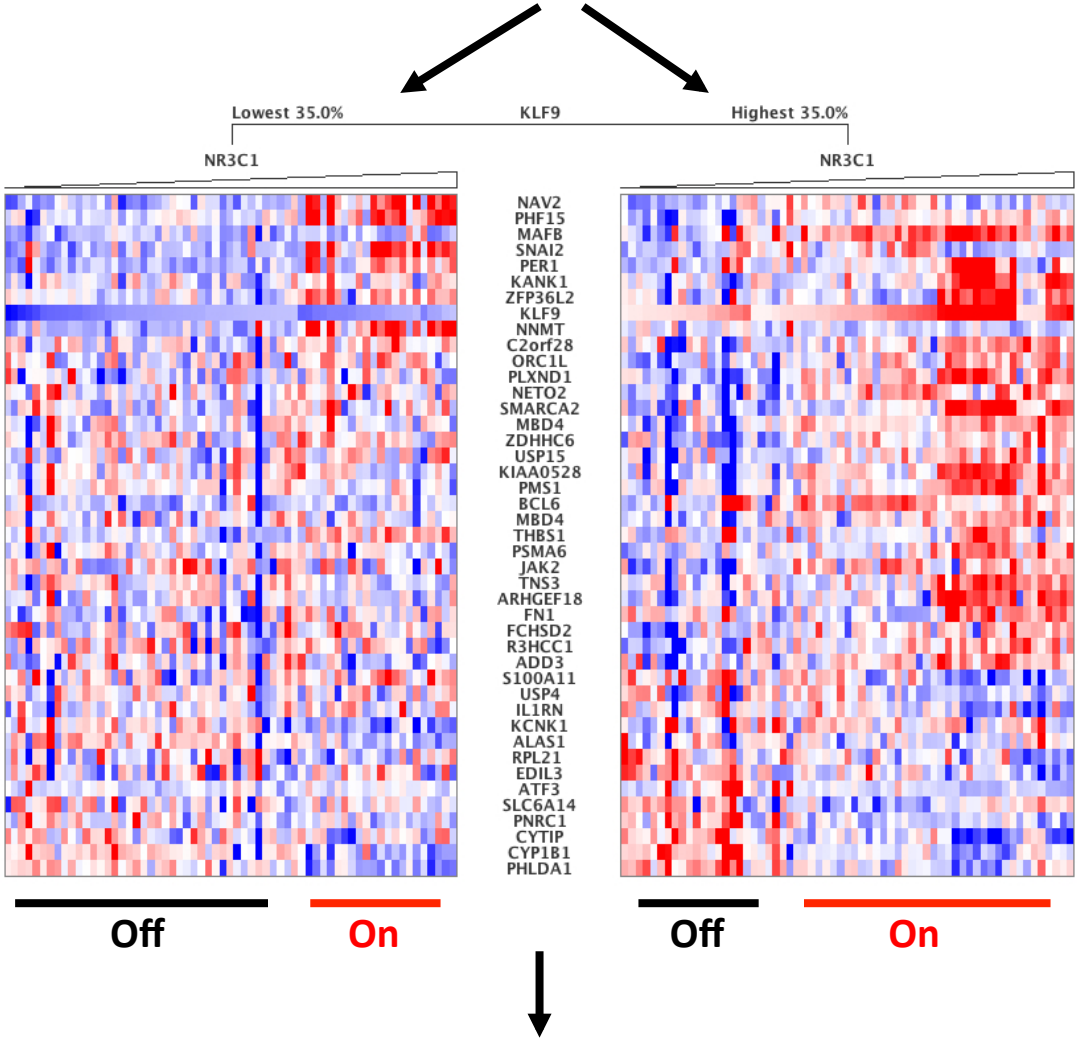


→ Snai2 behaves as a negative modulator of GR activity

→ GR independent activity not effected → selectivite modulator

.....Bioinfomatical analysis.....

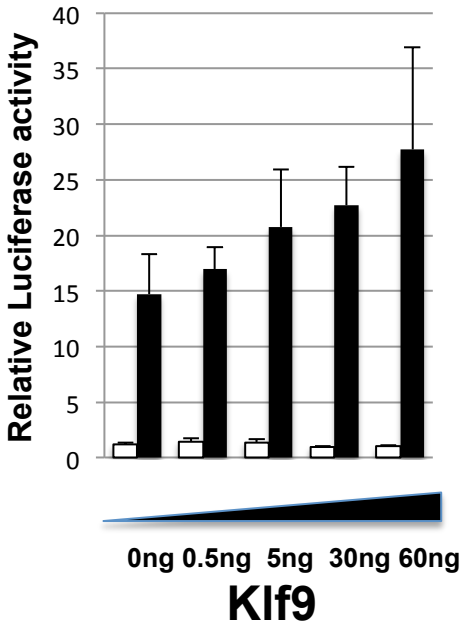
Sort experiments the expression level of a candidate modulator gene

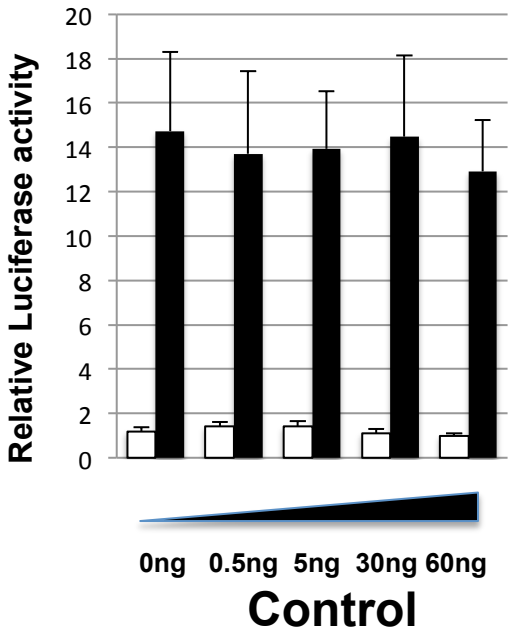


KLF9 positive modulator?

Experimental validation

KLF9 modulator activity???





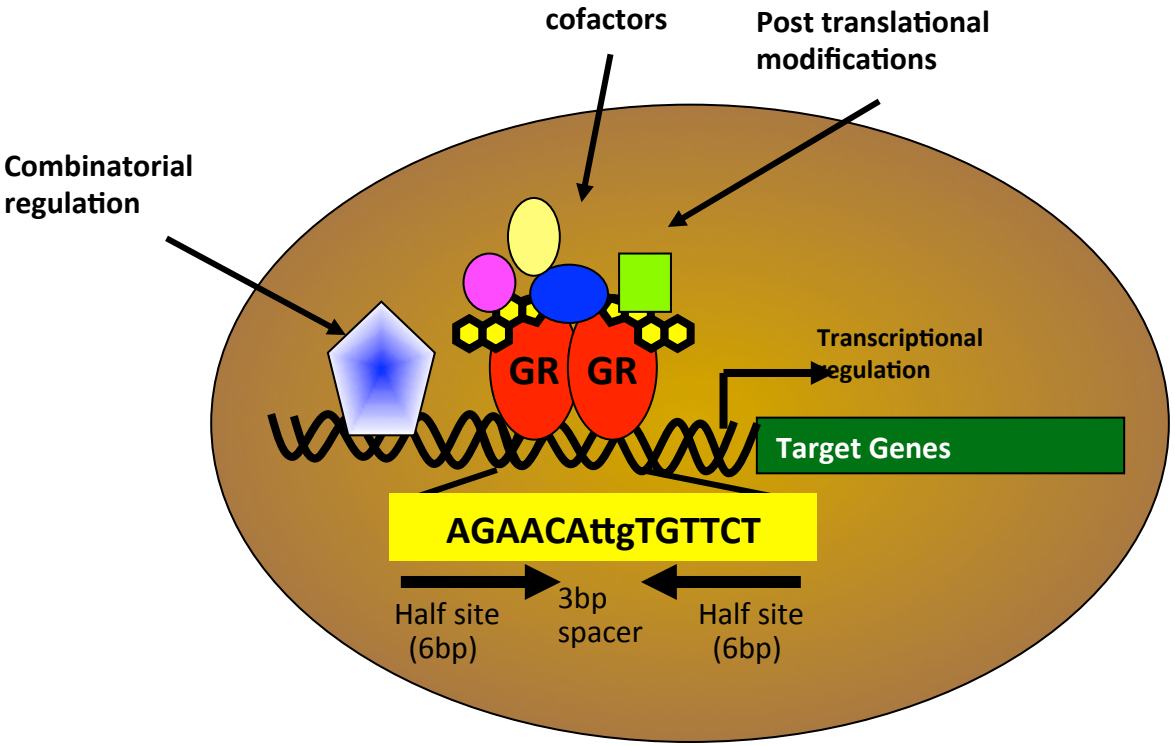
no hormone
Plus hormone (dexamethasone)

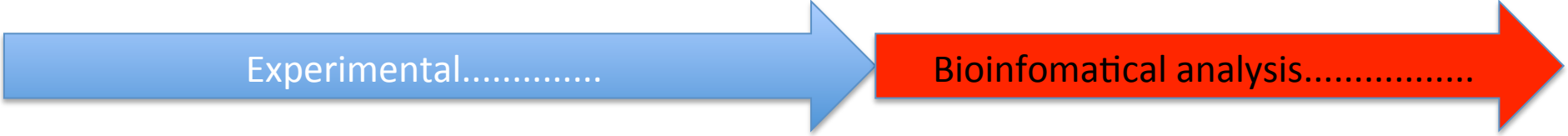
→ Klf9 behaves as a positive modulator of GR activity

Katja Borzym
Edda Einfeldt

Future directions:

Understand mechanistically how modulators modulate





Summary/Take-home messages:

- **Bioinformatical approaches can help generate hypothesis (Data analysis/integration)**
- ***Experimental validation to test hypothesis***
- ***Intimate understanding of the underlying algorithm is not essential***
- ***Knowing what can be done and guiding what kind of analysis is useful is essential***