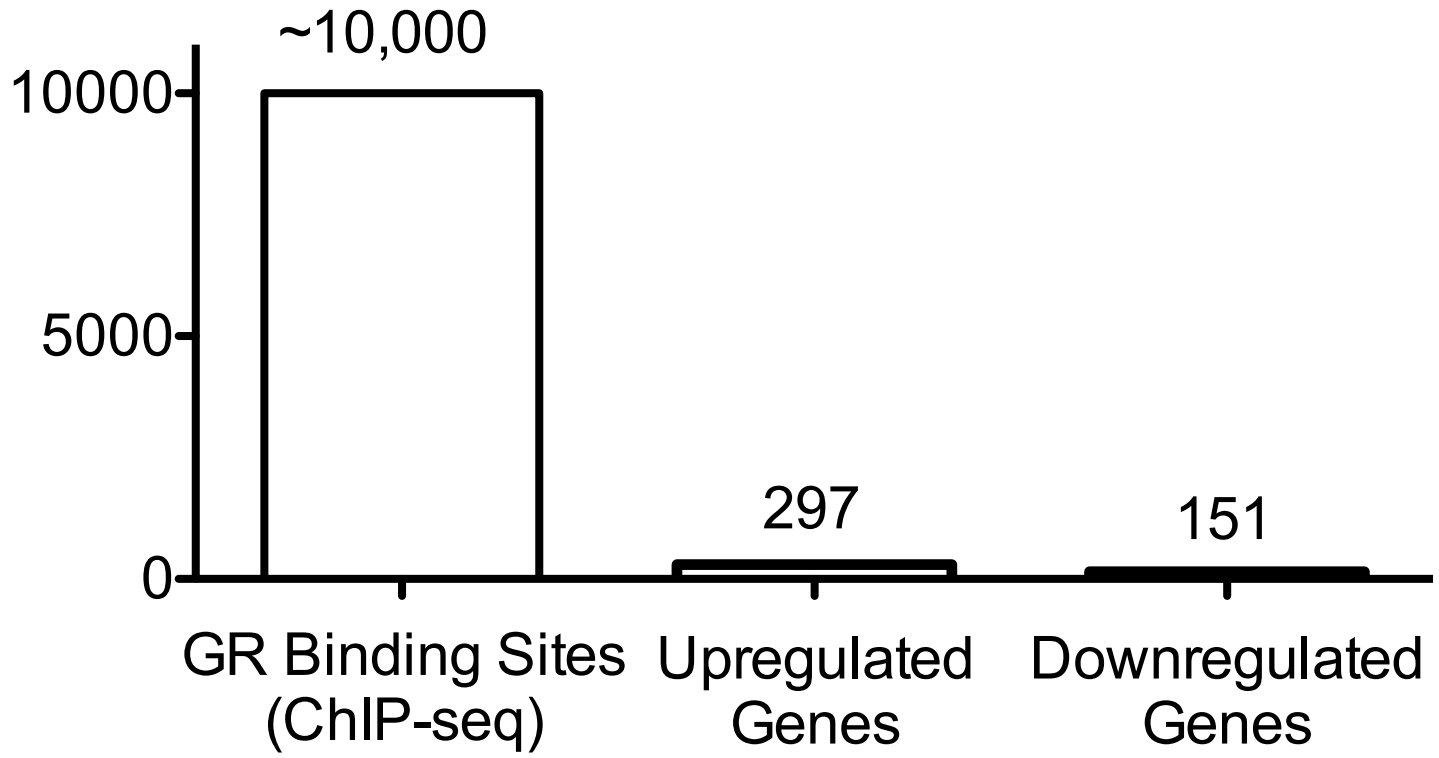
Comprehensive functional testing of ChIP-seq binding sites with ChIP-reporter assays

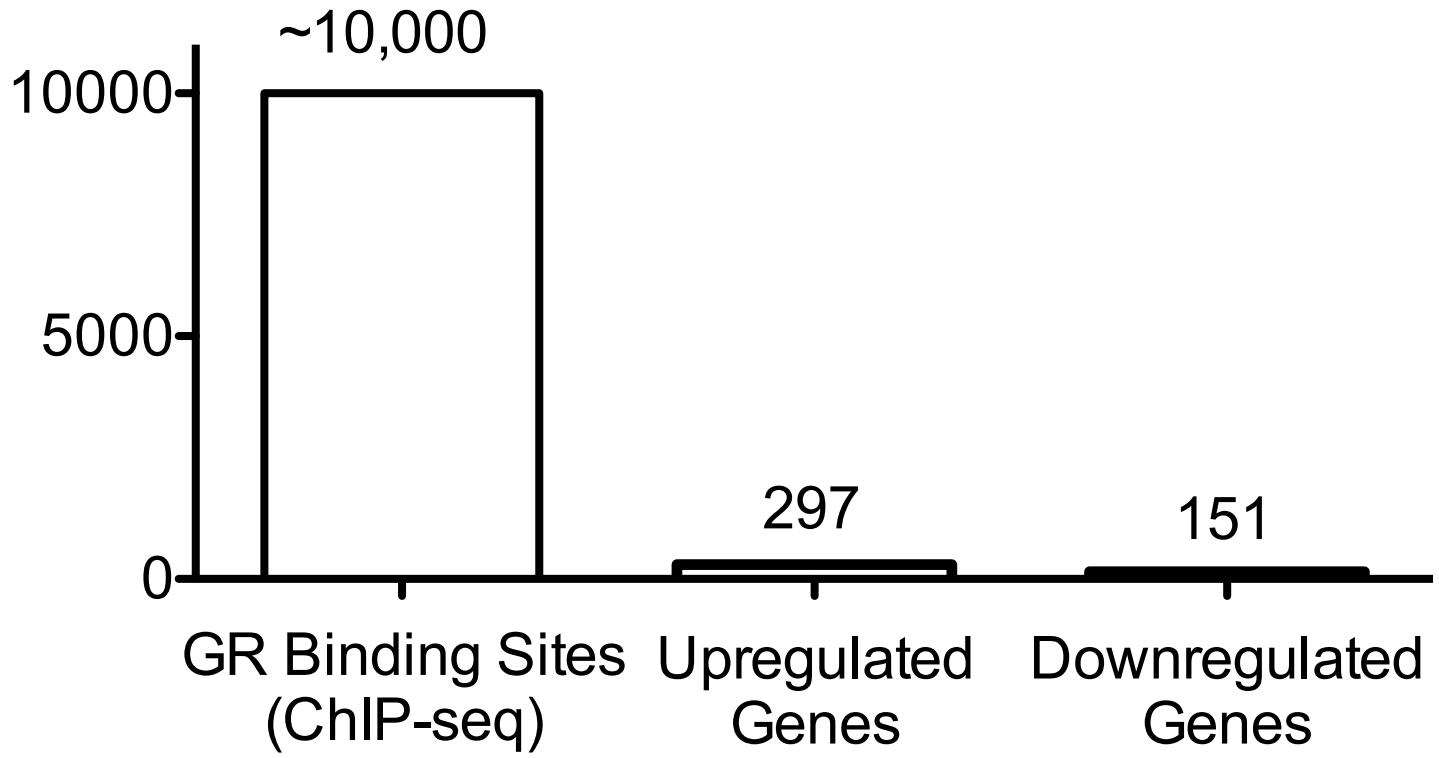
Tim Reddy
Duke University
ENCODE Users Meeting
June 9, 2016

ChIP-seq has revealed tens of thousands of TF binding sites in the human genome, far in excess of the regulated genes.



A549 cells (lung epithelial cell line), 3 h, 100 nM Dexamethasone

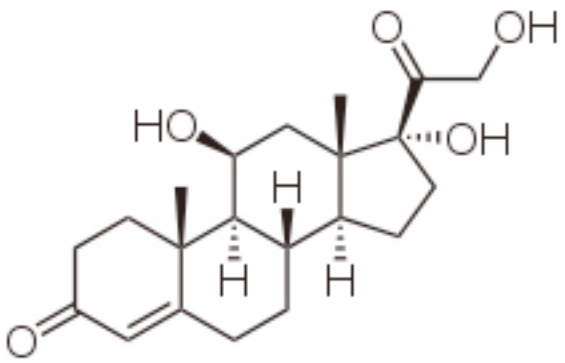
Major question post ChIP-seq:
What are all of these TF binding sites doing?



A549 cells (lung epithelial cell line), 3 h, 100 nM Dexamethasone

Our model system: Cortisol

(glucocorticoid steroid hormone)



Suppresses immune system and reduces inflammation

Increases blood pressure and blood sugar

Synthetic GC's used clinically to treat Psoriasis, Crohn's Disease, Rheumatoid Arthritis

Major component of response to stress response and metabolism

Our model system: Cortisol

(glucocorticoid steroid hormone)

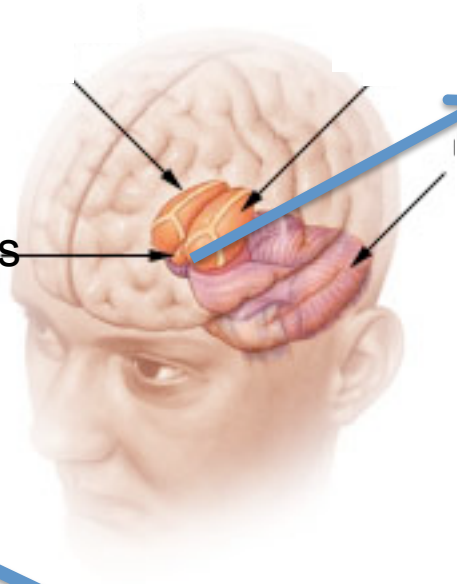
Suppresses immune system and reduces inflammation

Increases blood pressure and blood sugar

Synthetic GC's used clinically to treat Psoriasis, Crohn's Disease, Rheumatoid Arthritis

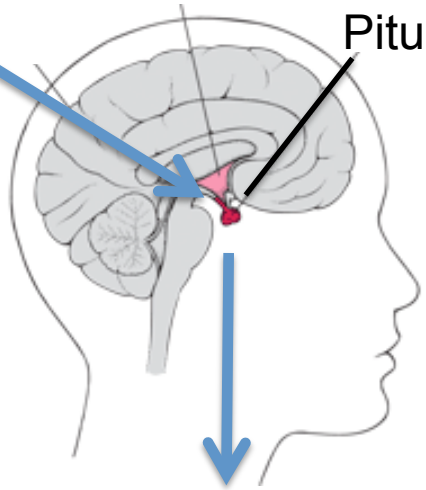
<u>Major component of response to stress response and metabolism</u>

Cortisol Production
(HPA axis)



Hypothalamus

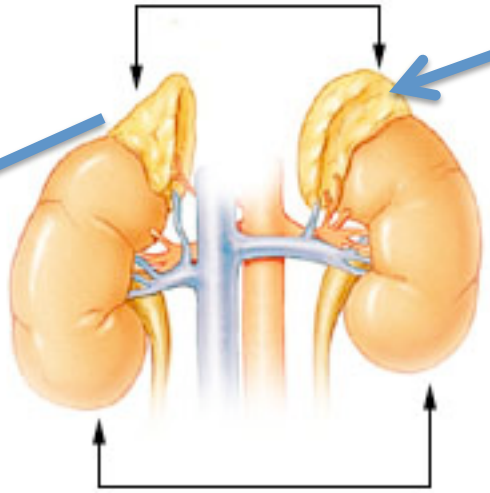
Corticotrophin-releasing Hormone (CRH)



Pituitary

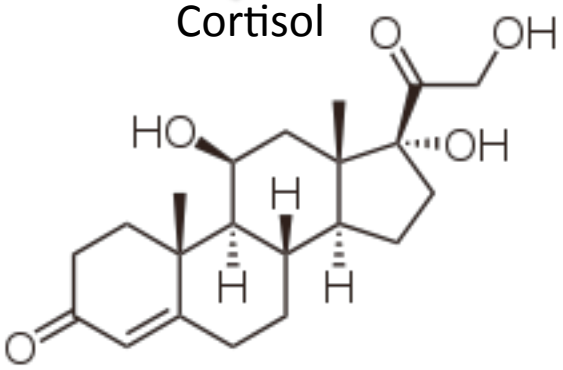
Adrenocorticotropic Hormone (ACTH)

Adrenal Glands

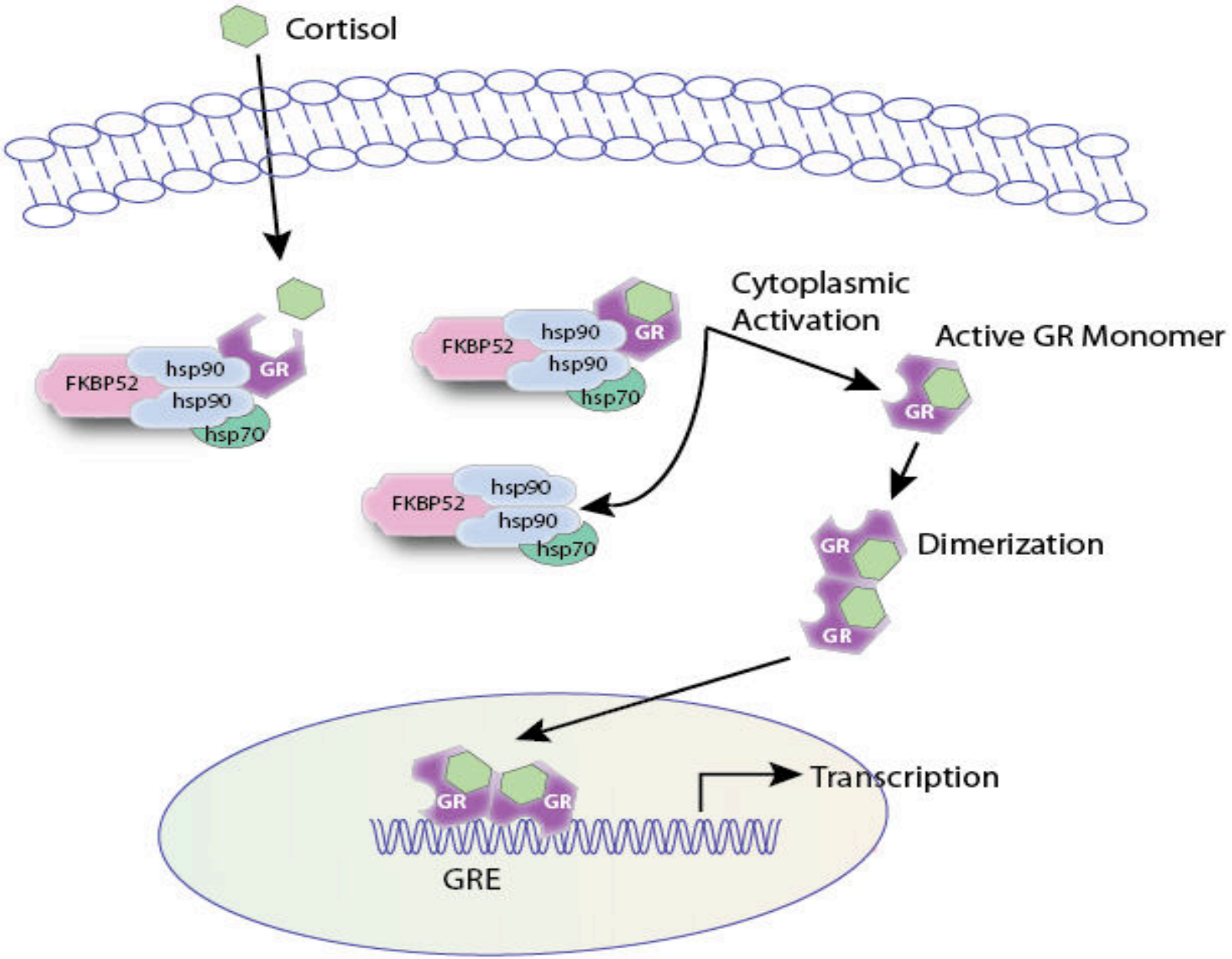


Kidneys

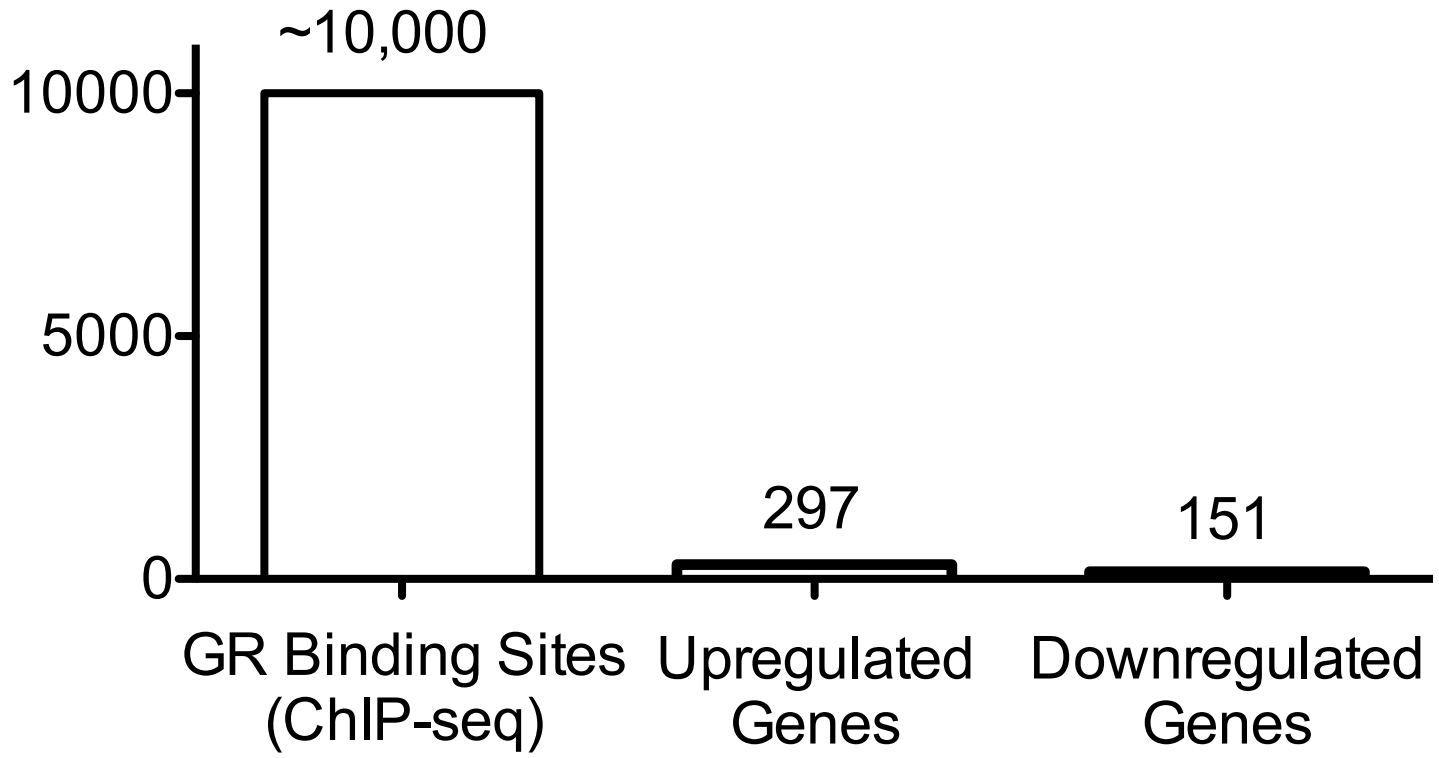
Cortisol



The Glucocorticoid Receptor (GR)

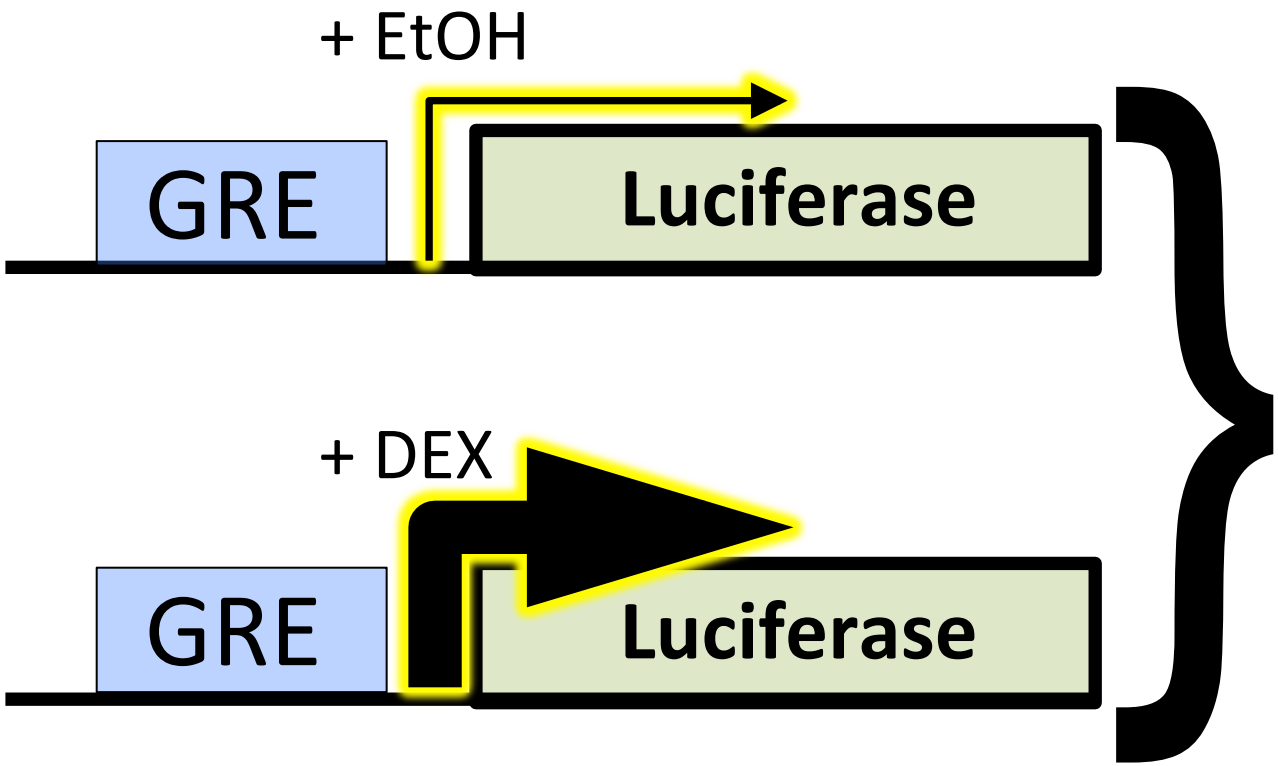


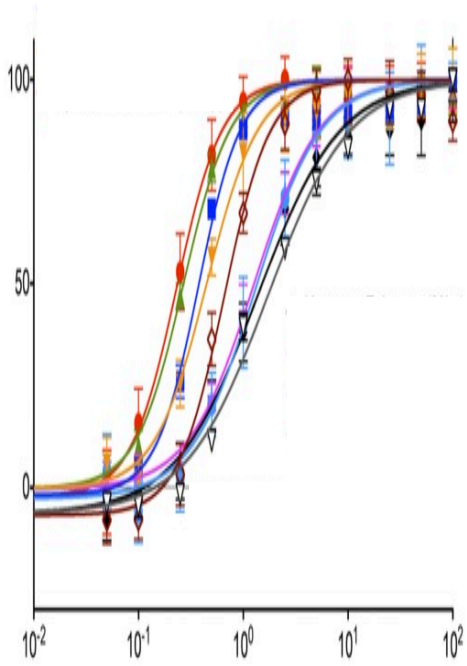
GR binds >10,000 sites in the human genome and regulates hundreds of genes.



A549 cells (lung epithelial cell line), 3 h, 100 nM Dexamethasone

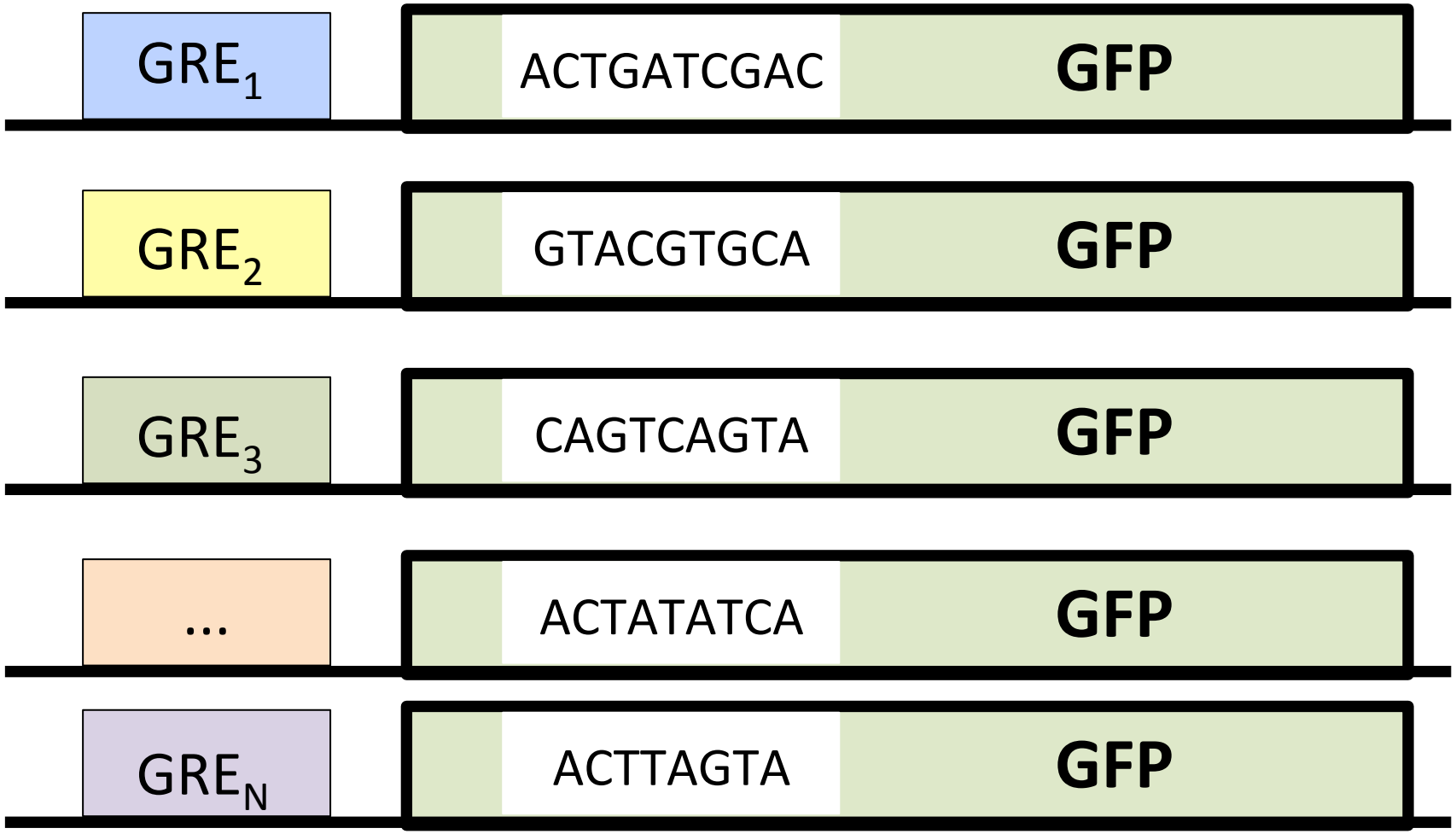
Reporter assays to quantify the activity of GC response elements





DEX = Dexamethasone, a synthetic glucocorticoid

Using high-throughput sequencing to make reporter assays high throughput



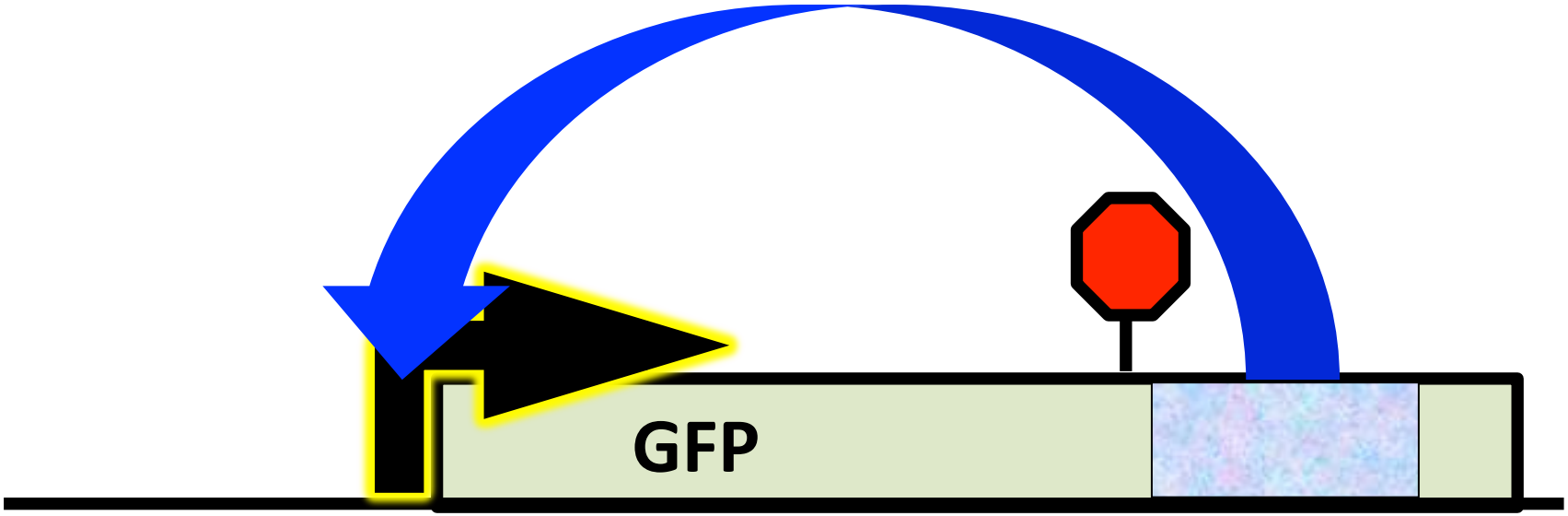
Patwardhan et al, 2012, Sharon et al, 2012, Kheradapour, 2013, Kwasnieski et al, 2012 and 2014, Melnikov et al, 2012 and 2013,

STARR-seq reporter assays

Regulatory elements located in the 3’ UTR of the reporter gene.

STARR-seq: Arnold et al, *Science*, 2013

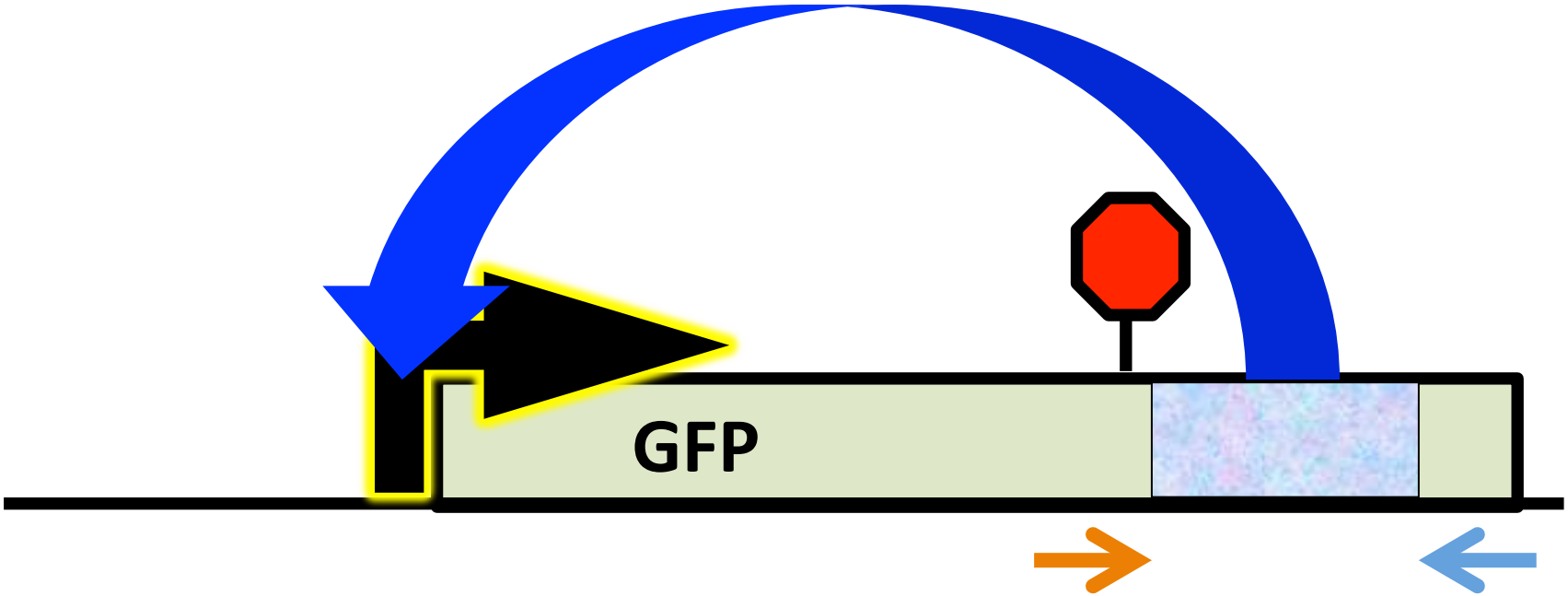
STARR-seq reporter assays



From that position, the elements regulate their own expression.

STARR-seq: Arnold et al, *Science*, 2013

STARR-seq reporter assays

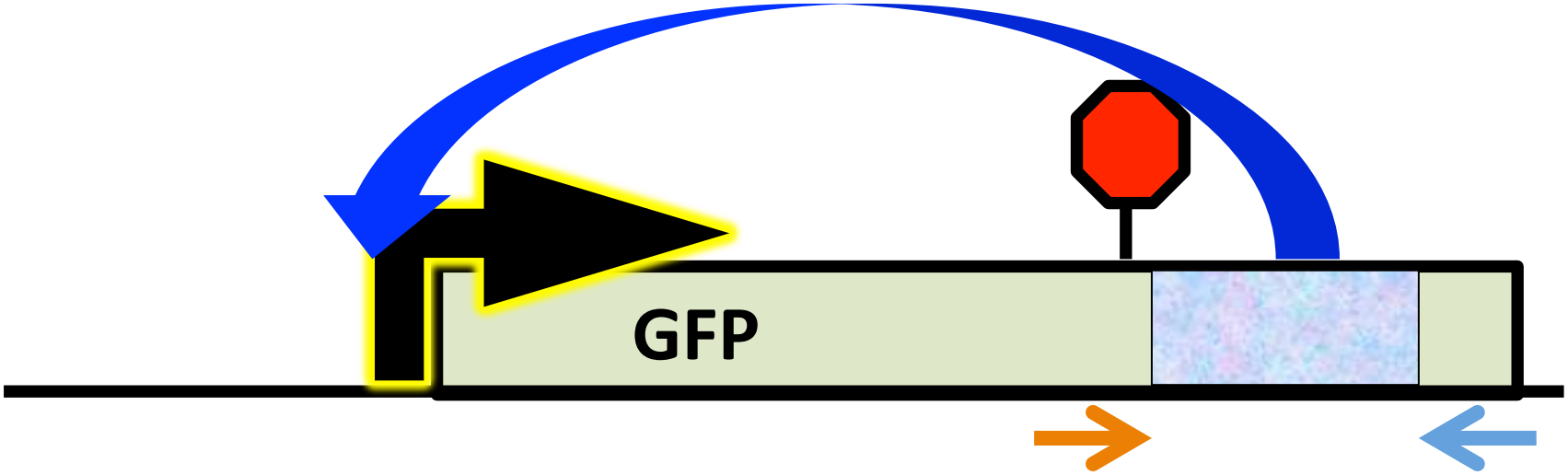


High-throughput sequencing of the cloning site in the expressed reporter gene can then be used to quantify regulatory element activity

STARR-seq: Arnold et al, *Science*, 2013

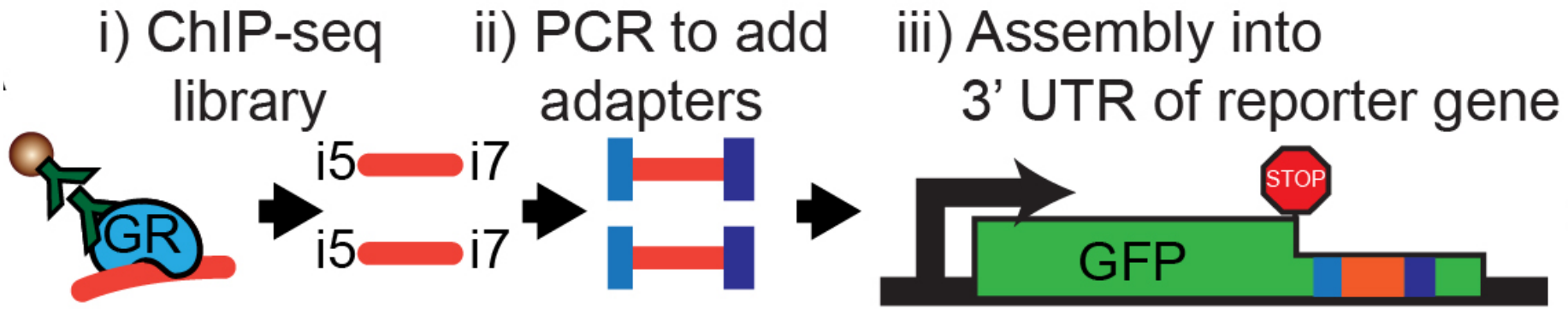
STARR-seq reporter assays

- Can assay millions of fragments at once
- Elements can be hundreds to thousands of bp
- Allows direct ligation of captured DNA into high-throughput reporter assays

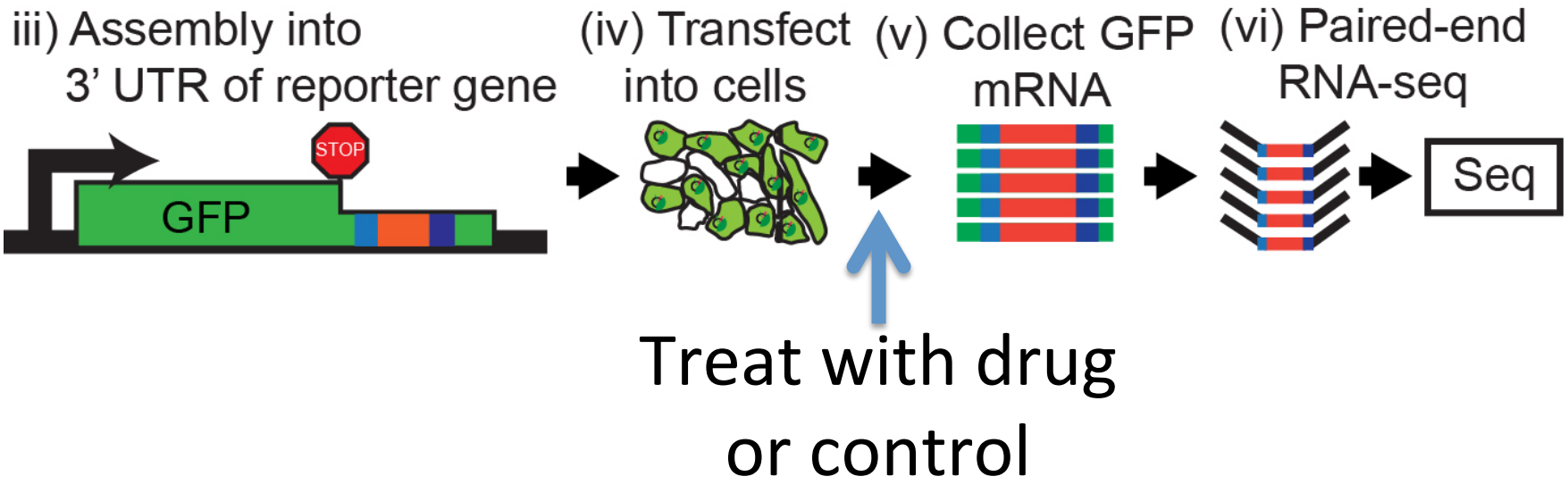


STARR-seq: Arnold et al, *Science*, 2013

Chromatin Immunoprecipitation + STARR-seq to quantify the activity of all TF binding sites

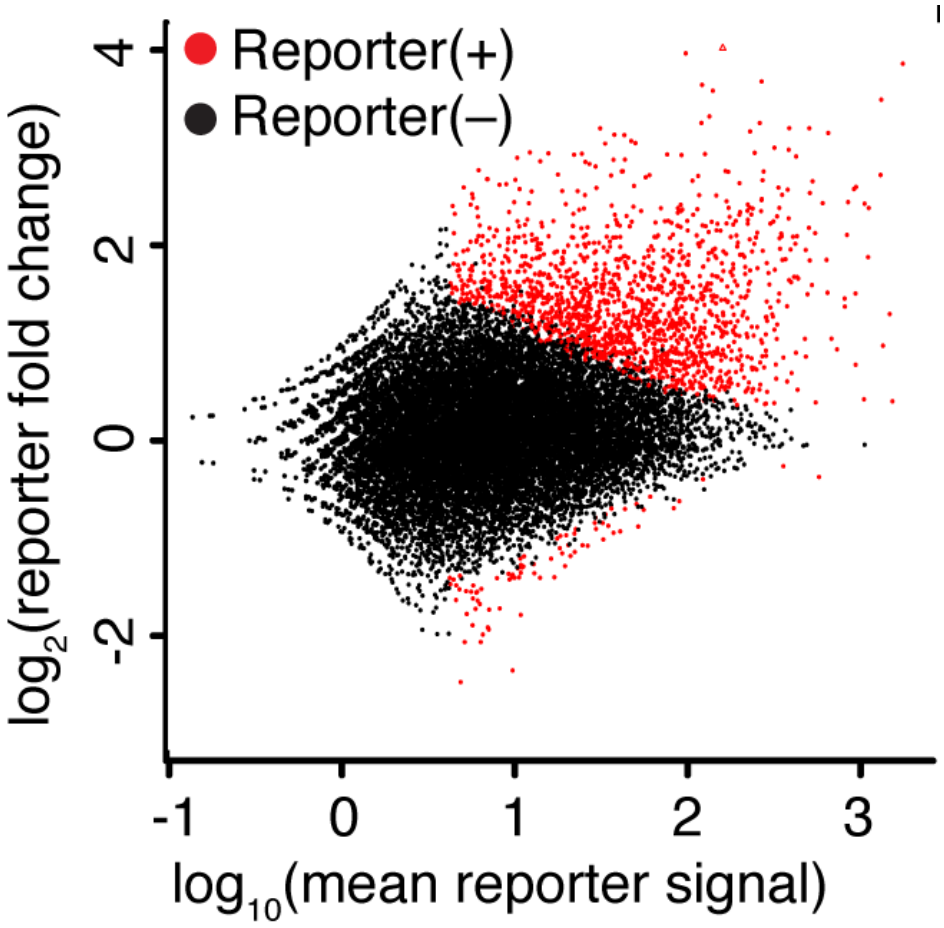


Chromatin Immunoprecipitation + STARR-seq to quantify the activity of all TF binding sites

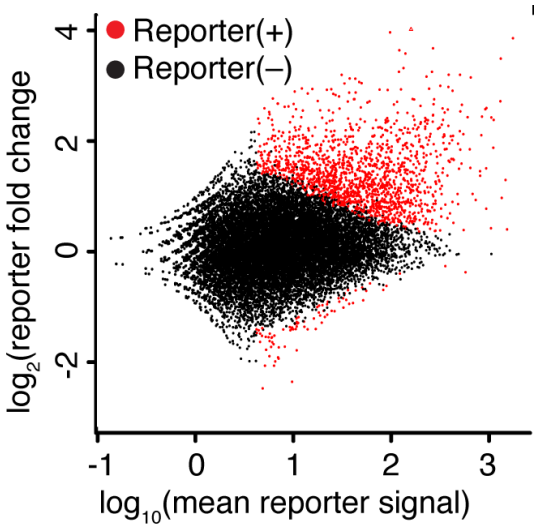


GR ChIP-reporter results

- Assayed >12,000 GR binding sites
- 10% were DEX-responsive
- 95% of regulatory elements increased reporter expression
- Validated with standard dual-luciferase (r = 0.77)



Possible interpretations:

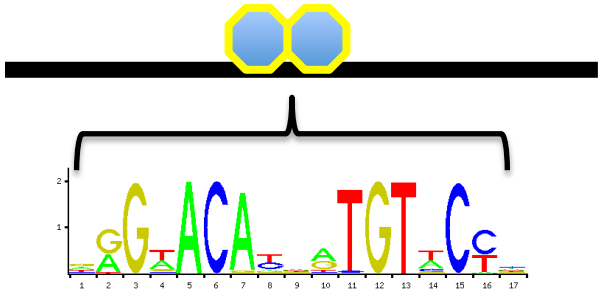


One possibility: Only 10% of the GR binding sites have activity.

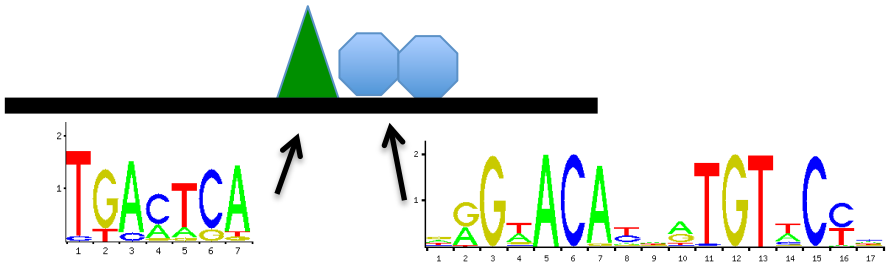
Alternative: Other GR binding sites require additional genomic context.

Some ways that GR binds the genome

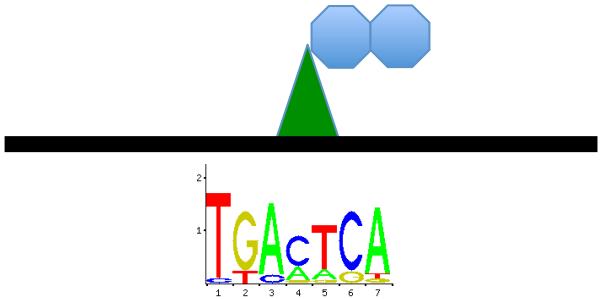
Direct binding to a Glucocorticoid Response Element (GRE)



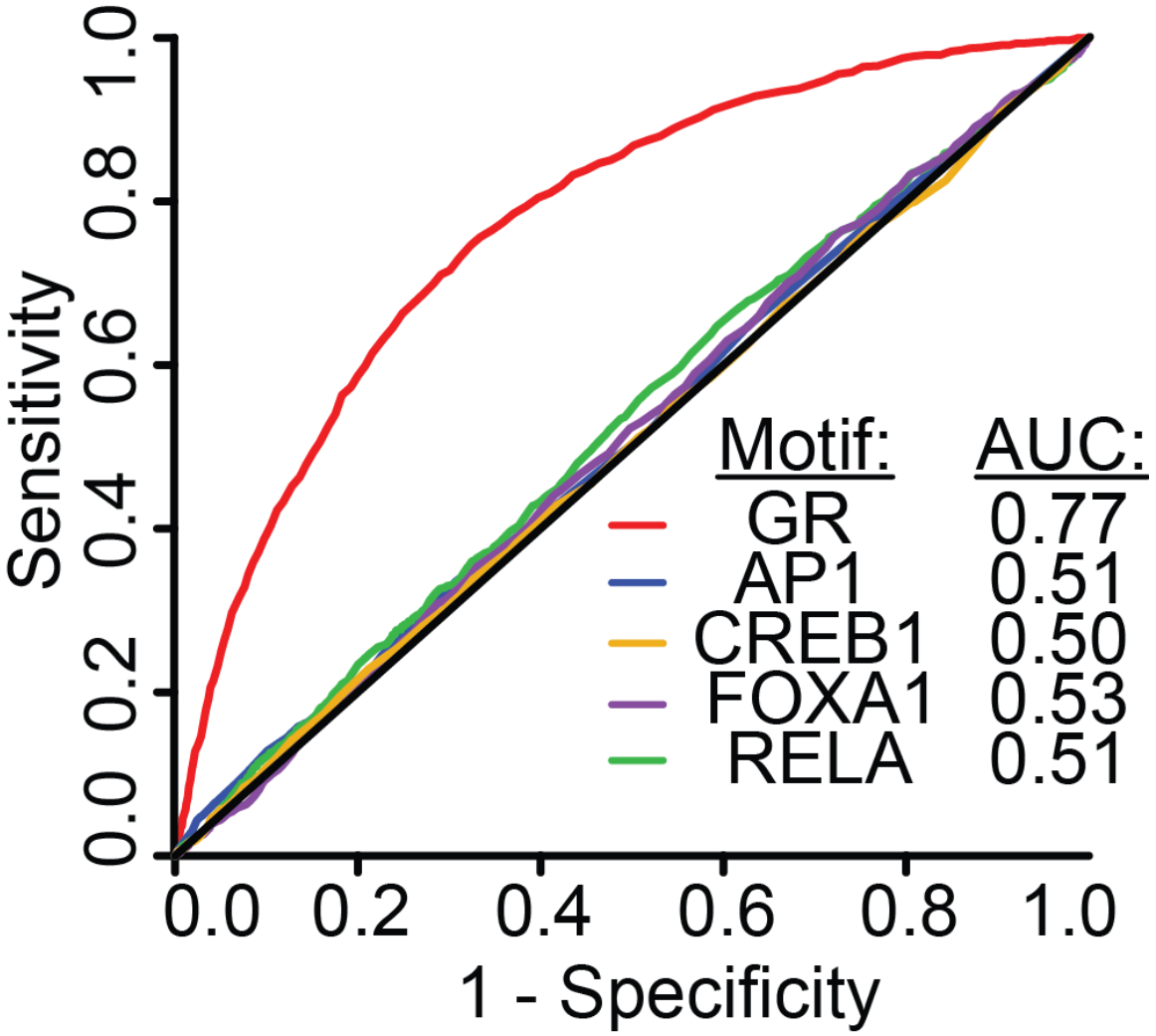
Cooperative binding to a composite regulatory element



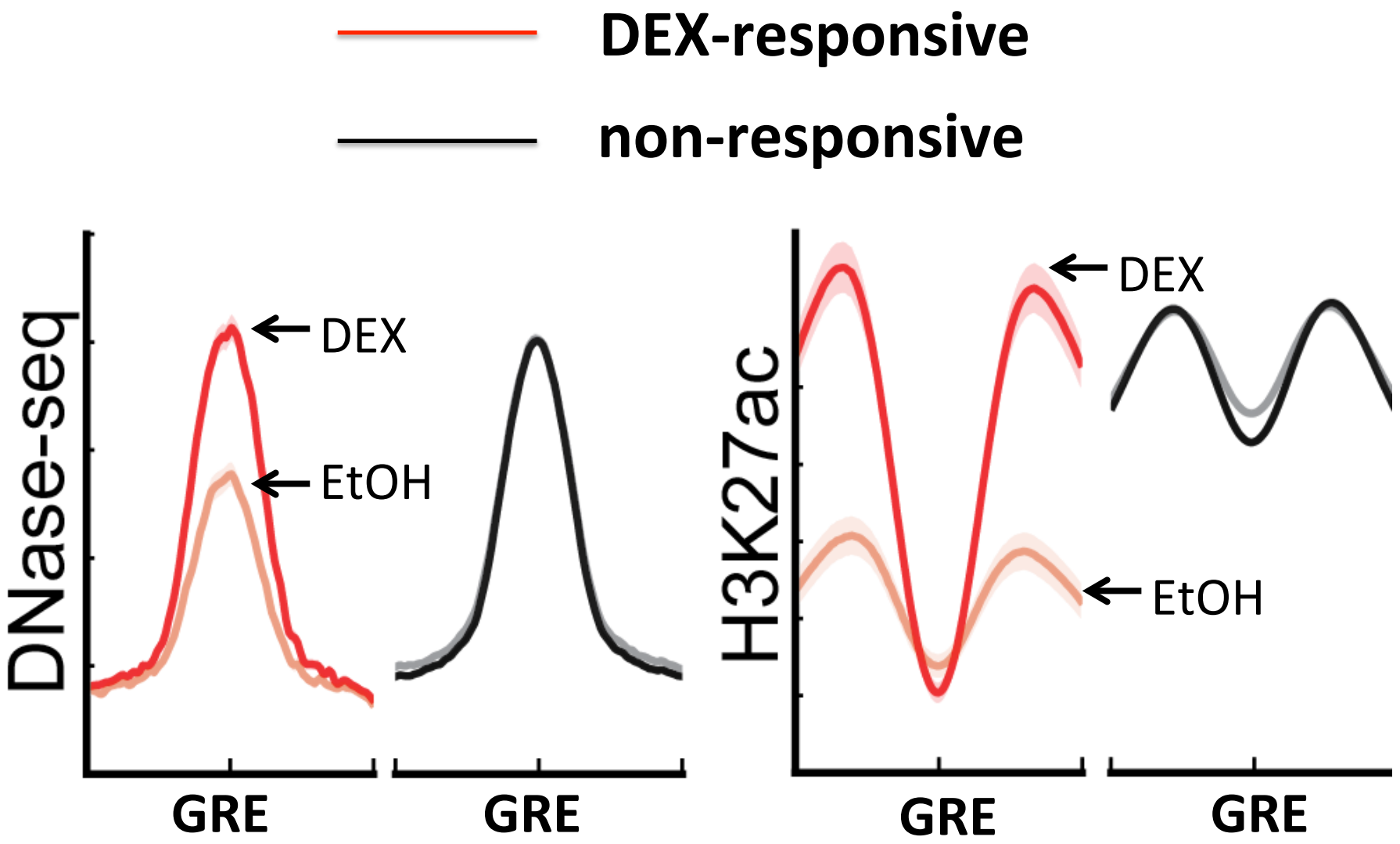
Tethered binding via other TFs such as AP-1



Presence of a GRE explains the response
Co-binding TF motifs do not

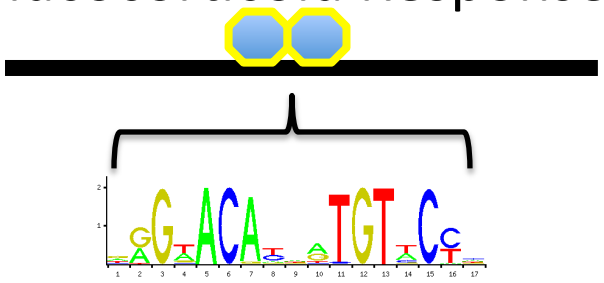


DEX-responsive sites also have epigenetic state changes that reflect activity in the genome



Direct binding explains DEX-responsive reporter activity

Direct binding to a Glucocorticoid Response Element (GRE)



Cooperative binding to a composite regulatory element

Tethered binding via other TFs such as AP-1

The problem

- The set of GC-responsive genes varies dramatically between cell types
- Those differences can largely be attributed to changes in co-factors, particularly AP-1

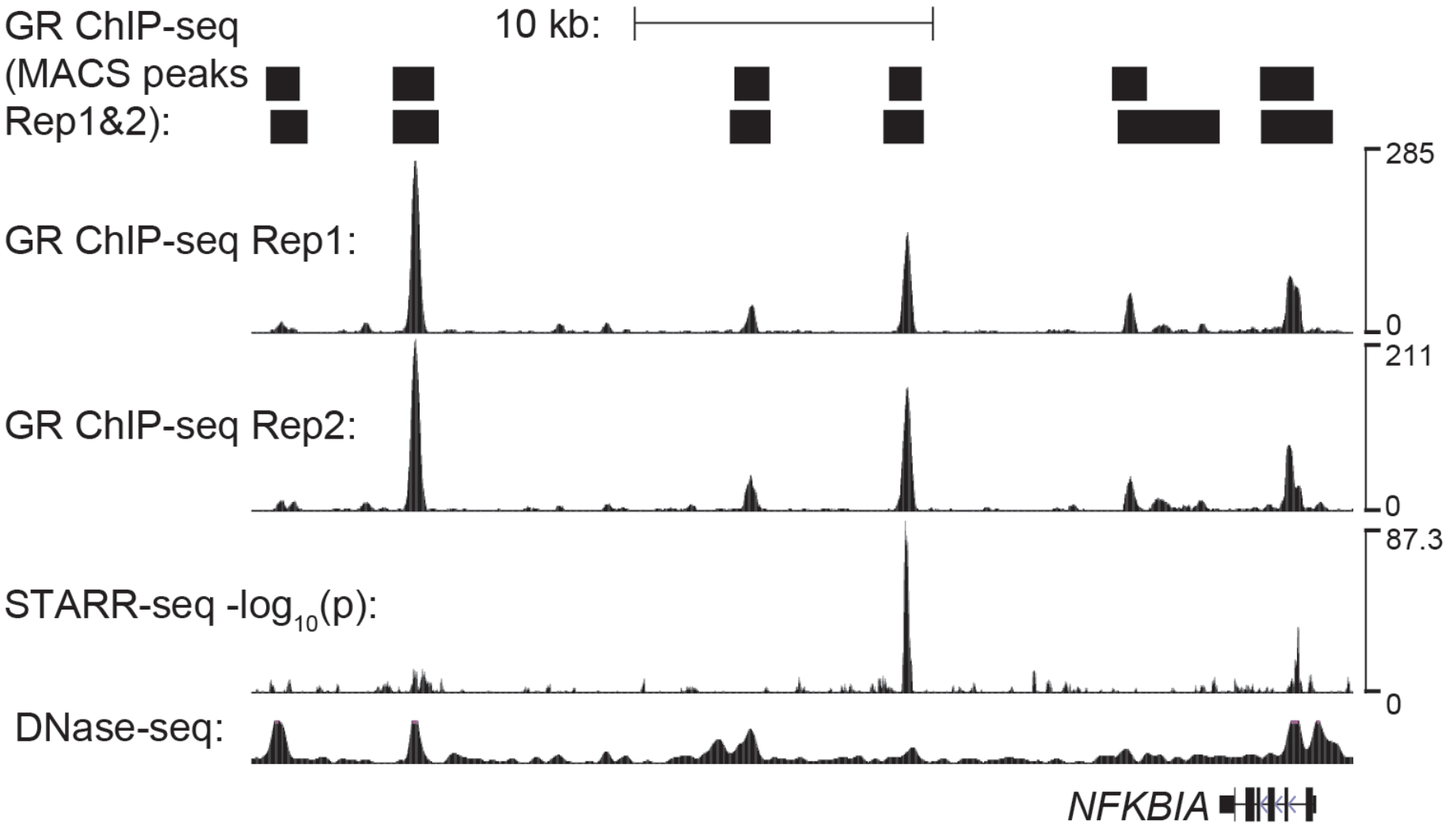
  (e.g. Biddie et al, 2011; John et al, 2011; Gertz et al, 2013)

- *If GR binding at AP-1 sites does not have activity, then we struggle to explain cell-type-specific differences in GC-responses*

Our model

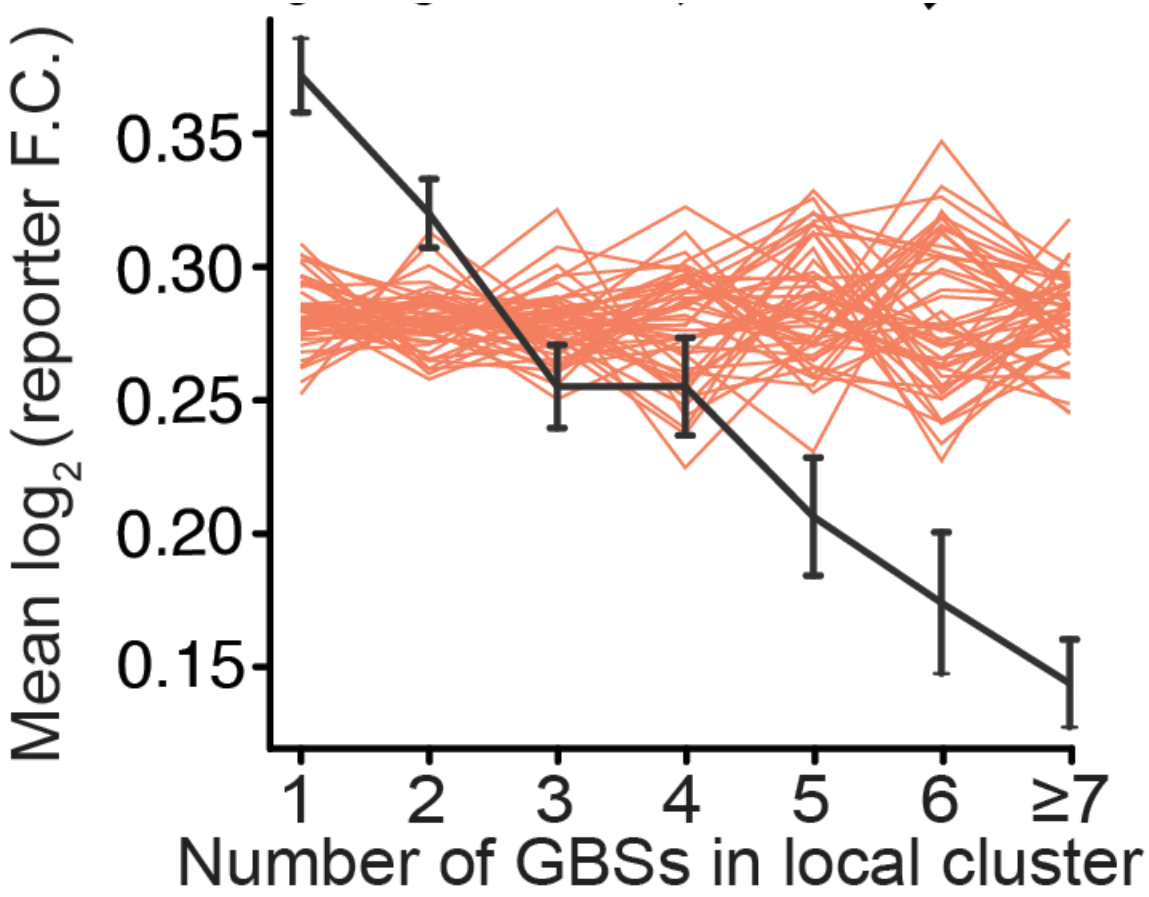
- AP-1 sites modulate the activity of direct GR binding sites in the genome
- However, the AP-1 sites are not sufficient for DEX-responsive regulatory activity
- *In this model, we expect GR and AP-1 to co-cluster in the genome*

GR binds the genome in clusters

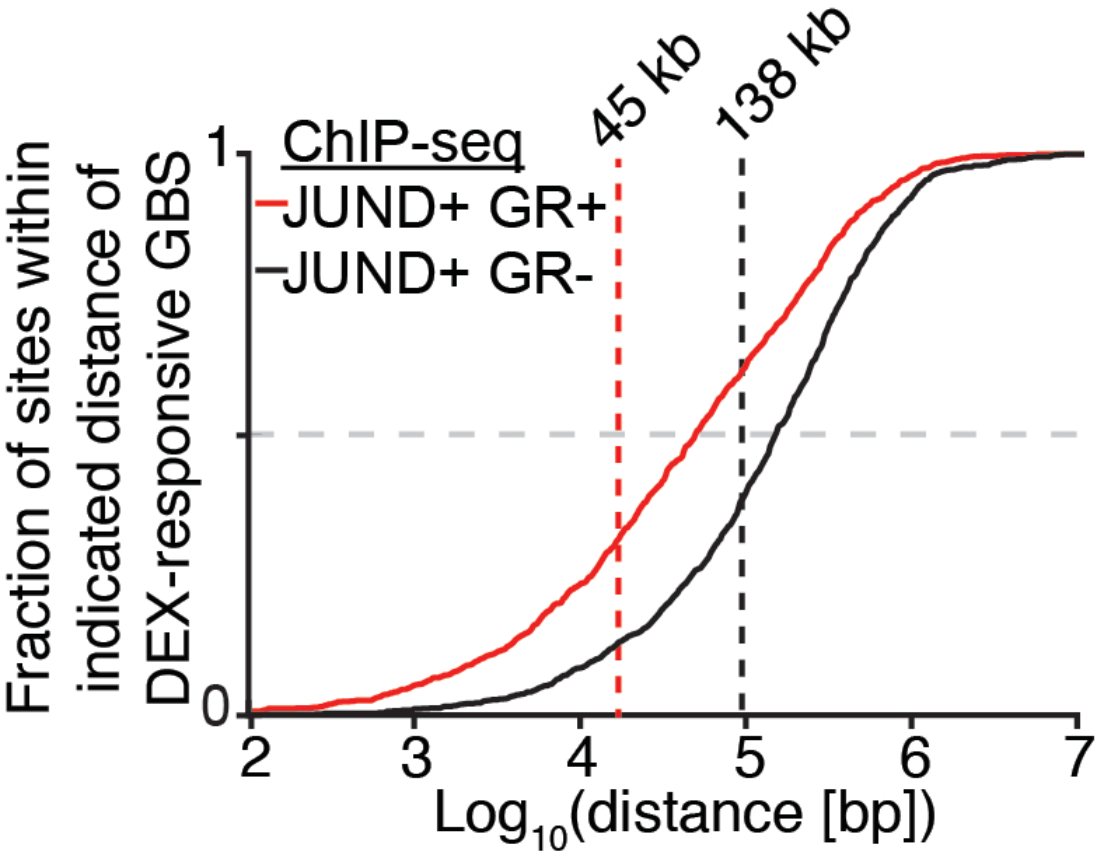


More numerous clusters have a smaller fraction of DEX-responsive sites

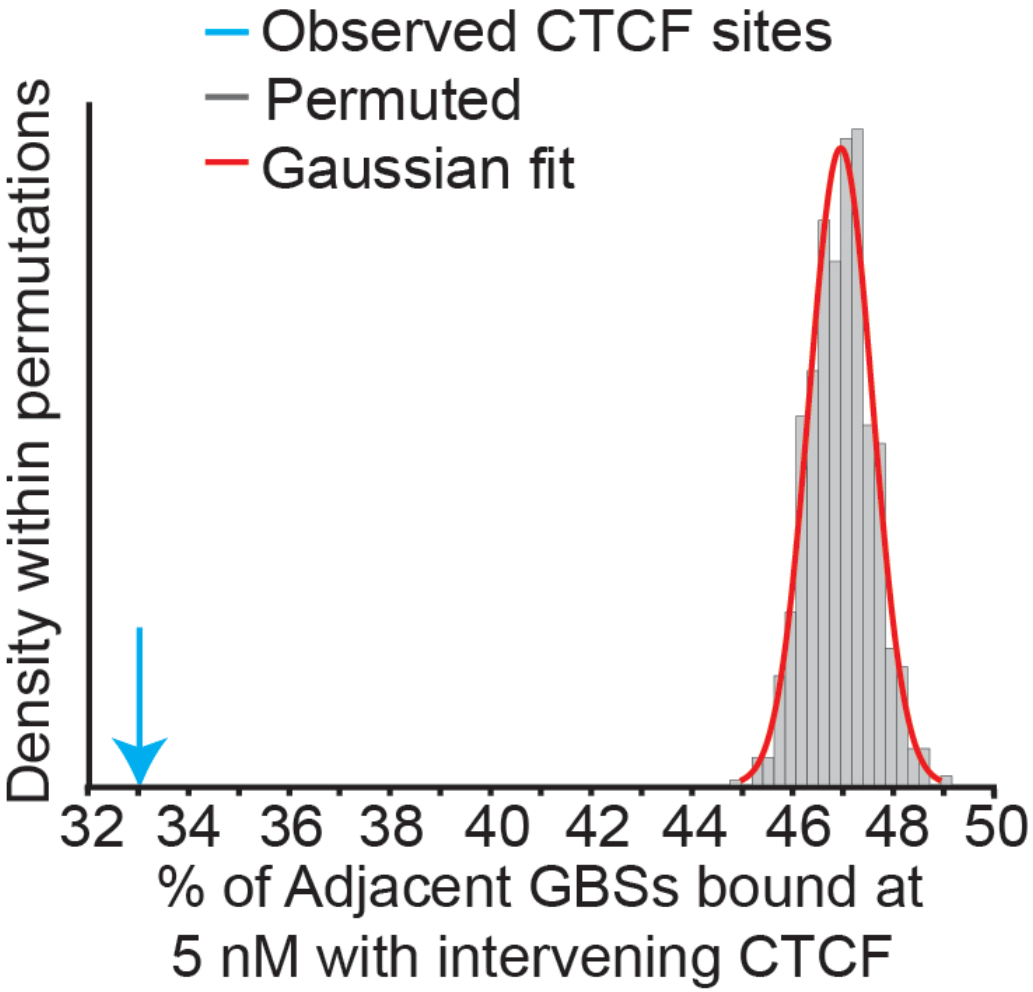
(i.e. few direct sites can nucleate large clusters)



AP-1 sites, represented by JunD, that gain GR are closer to direct GR binding sites



Clusters of GR binding are depleted for intervening CTCF

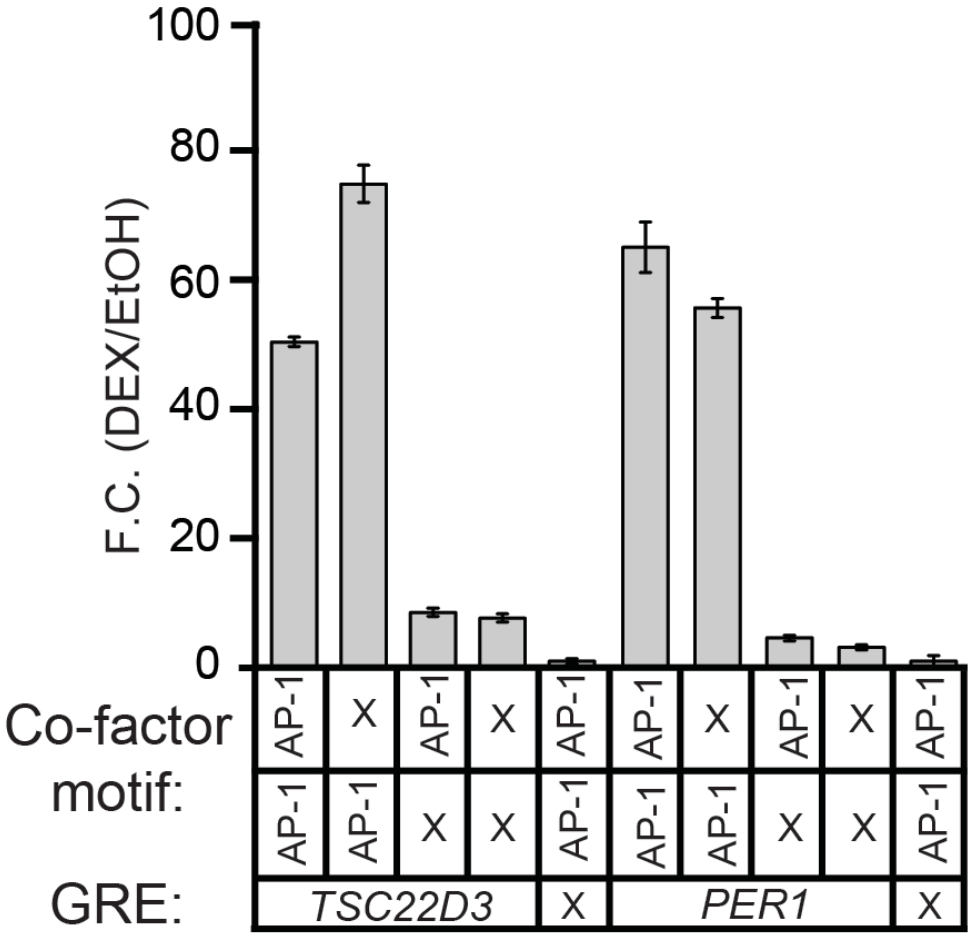


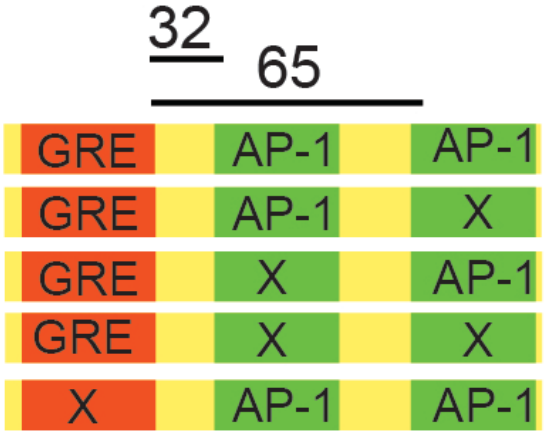
Conclusion

- GR biding at non-responsive site likely reflects looping interactions with distal direct GR binding sites

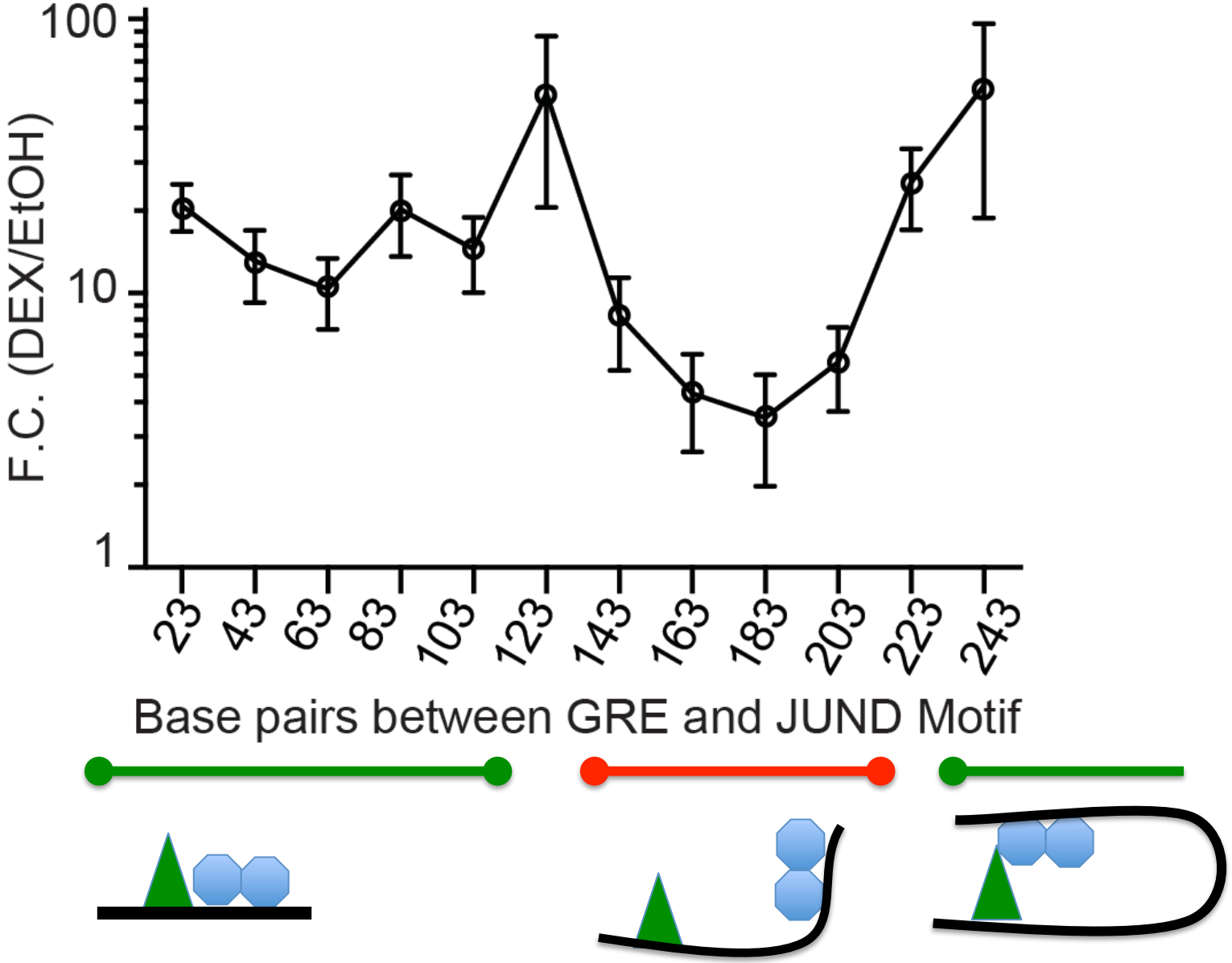
- Does this alter DEX-responsive regulatory activity, potentially explaining cell-type specificity?

Adding AP-1 sites amplifies GC-response by >10-fold

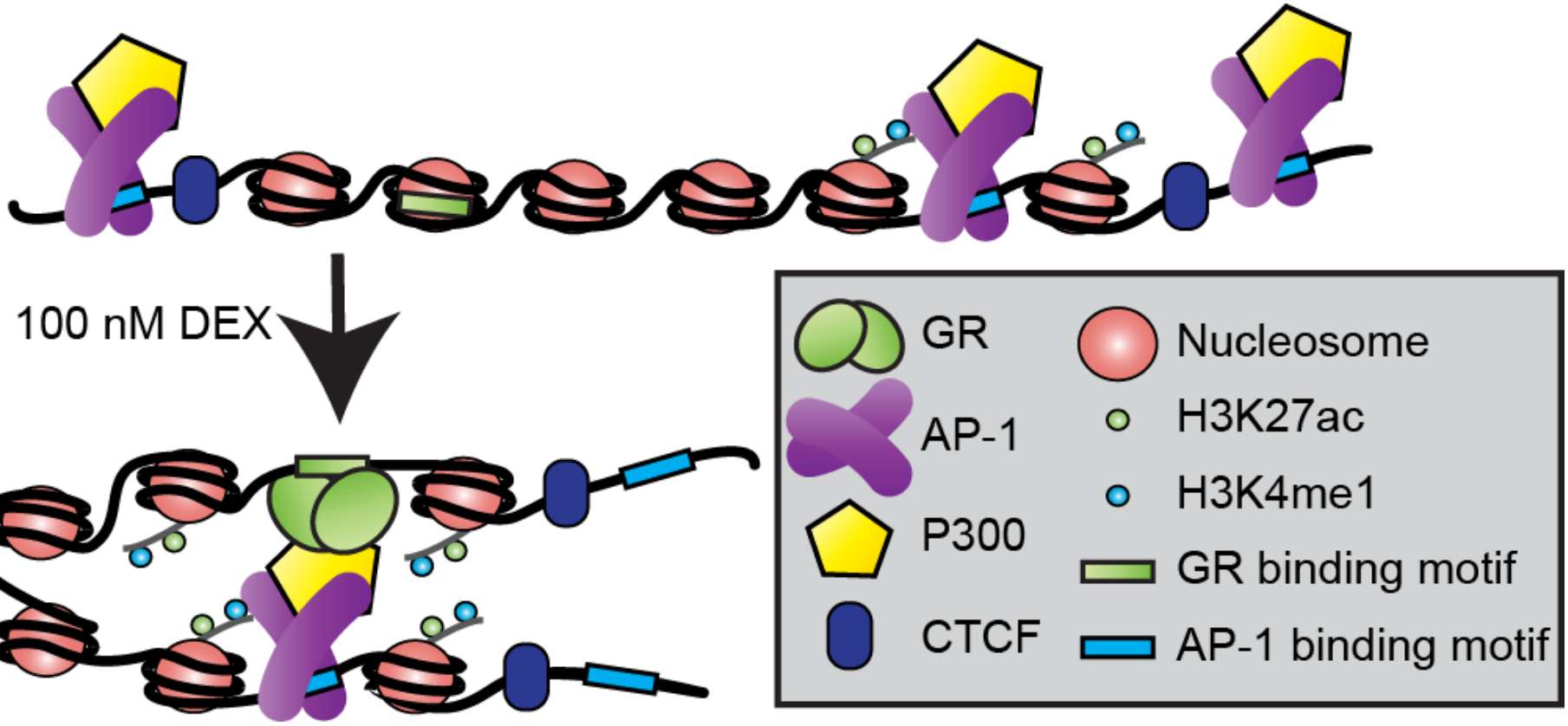




That amplification is distant dependent



Summary of our model: GR binding clusters reflect extensive GR:AP-1 tethering interactions.

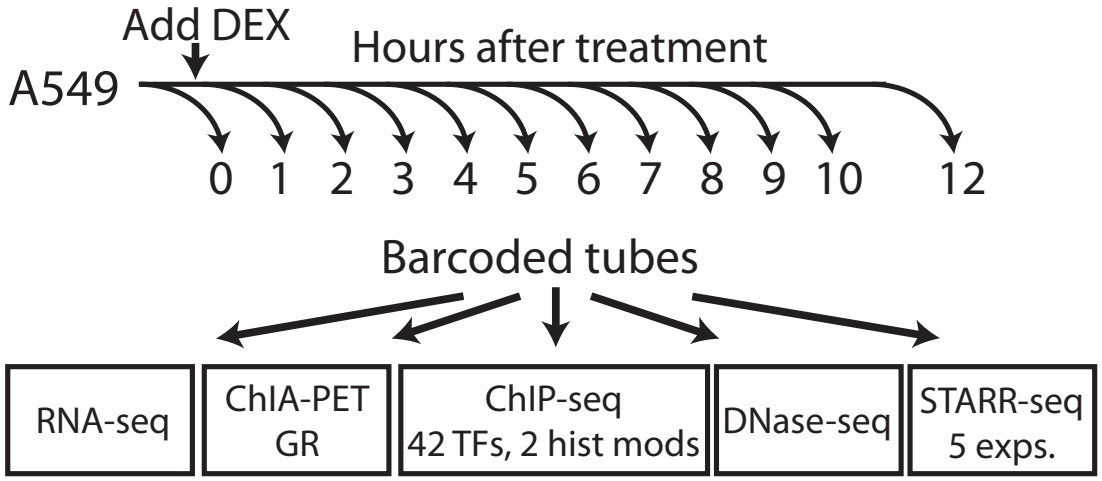


Summary

- We developed a high-throughput approach to measure the activity of every GR binding site in a reporter assay

- The results reveal a functional diversity of GR binding sites, and suggest that interactions between direct and tethered GR binding sites are the basis for cell-type specific GC responses

Ongoing work:
Genomics of Gene Regulation

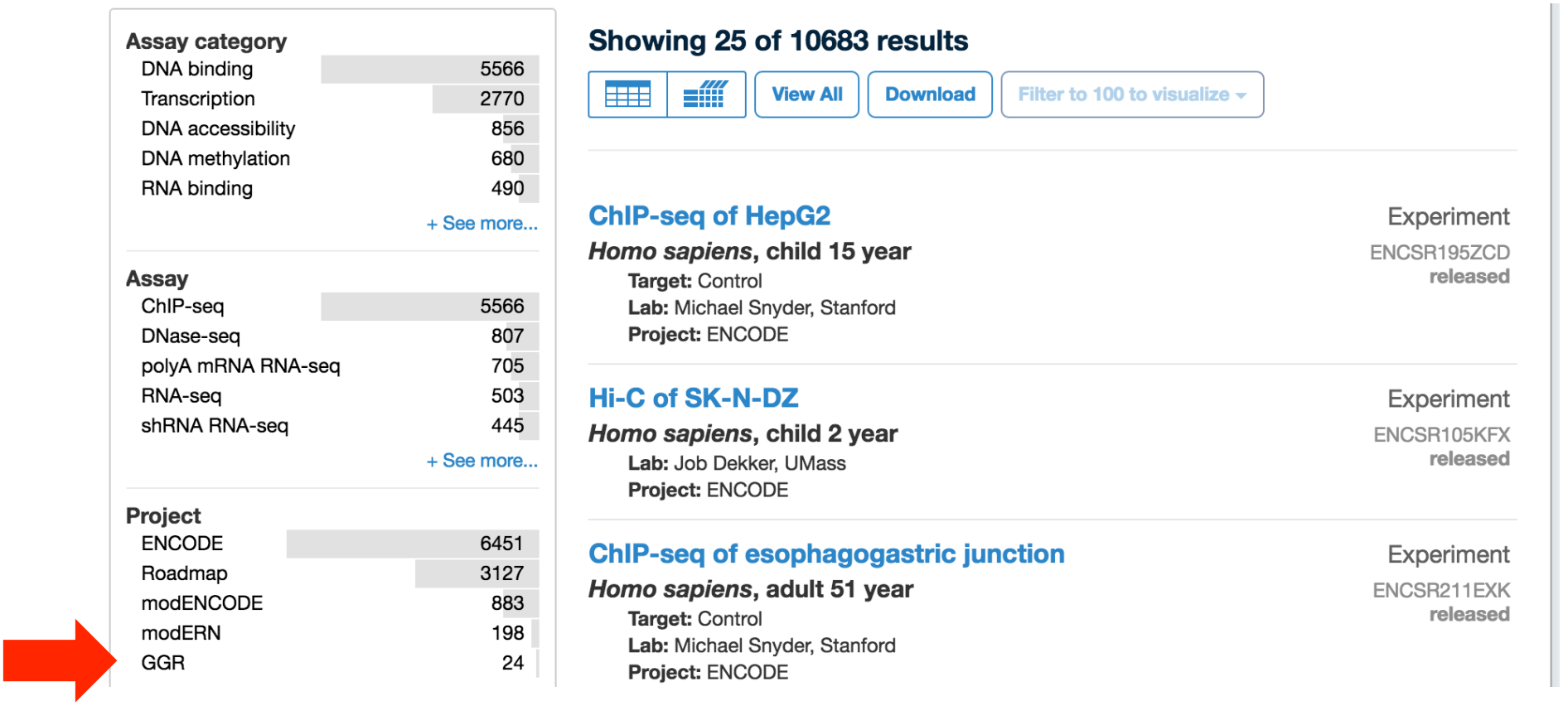


Expanding upon these findings with extensive and coordinated ChIP-seq, DNase-seq, RNA-seq, STARR-seq, Hi-C

Matched with Bayesian models of regulatory networks, and epigenome and genome editing to test predictions

RNA-seq and DNase-seq released via ENCODE DCC

- 12 time points, coordinated to the second
- At least four replicates of each time point
- ChIP-seq / Hi-C / STARR-seq data forthcoming



Looking for postdocs to join the lab

- Looking for both experimental and computational scientists

- Genomics and Genetics projects featuring high-throughput reporter assays, genome and epigenome editing, network modeling, experimental design, and specific disease studies.

Acknowledgements

Reddy Lab:

Karl Guo, Ph.D.

Chris Vockley (Looking for a postdoc)

Anthony D'Ippolito

Bill Majoros

Ian McDowell

Graham Johnson

Linda Hong

Sarah Leichter

Luke Bartelt

Gersbach Lab:

Charlie Gersbach

Tyler Klann

Josh Black

Isaac Hilton

Dewran Kocak

Pratiksha Thakore

Ami Kabadi

Lauren Polstein

Northwestern

Bill Lowe

Geoff Hayes

Denise Scholtens

Brian Layden

Anton Ludvik

Michael Nodzenski

Duke:

Greg Crawford

Greg Wray

Alex Hartemink

Brigid Hogan

Princeton

Barbara Engelhardt

Penn

Casey Brown

Funding:

National Human Genome Research Institute

U01 HG007900 (Reddy)

NIDDK

NATIONAL INSTITUTE OF DIABETES AND DIGESTIVE AND KIDNEY DISEASES

R01 DK097534 (Lowe)

R01 DK099820 (Lowe)

NIAMS

R01 DA036865 (Gersbach)

Predoctoral Fellowships

F31 HD085666 (Guo)

F31 HL129743 (Vockley)