Cornell
Baker Institute for Animal Health

Dissection of core promoter architecture through single nucleotide-resolution modeling of transcription initiation

Adam He & Charles Danko

Mechanisms of Eukaryotic Transcription (2023)

Central question

How does the sequence at enhancers and promoters control local transcription initiation?

CIS-BP

Lots of efforts to map binding motifs for sequence-specific TFs, core promoter motifs

JASPAR

Central question

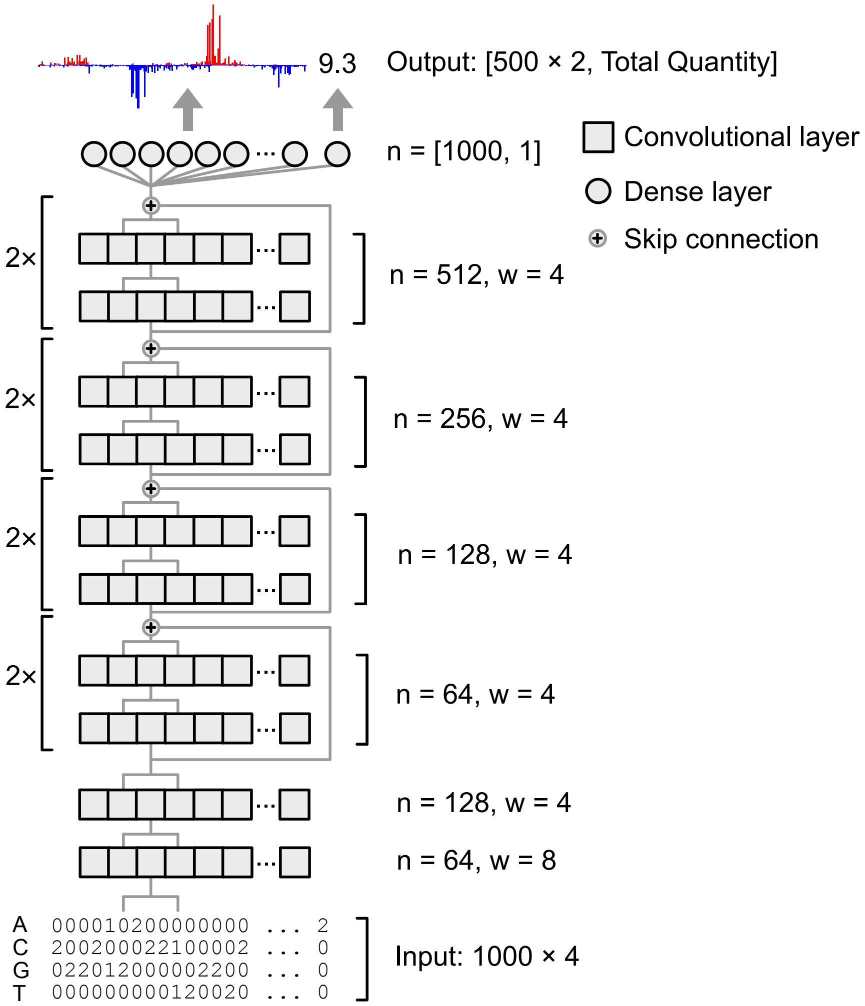
How does the sequence at enhancers and promoters control local transcription initiation?

But a complete picture of how initiation is determined remains elusive.

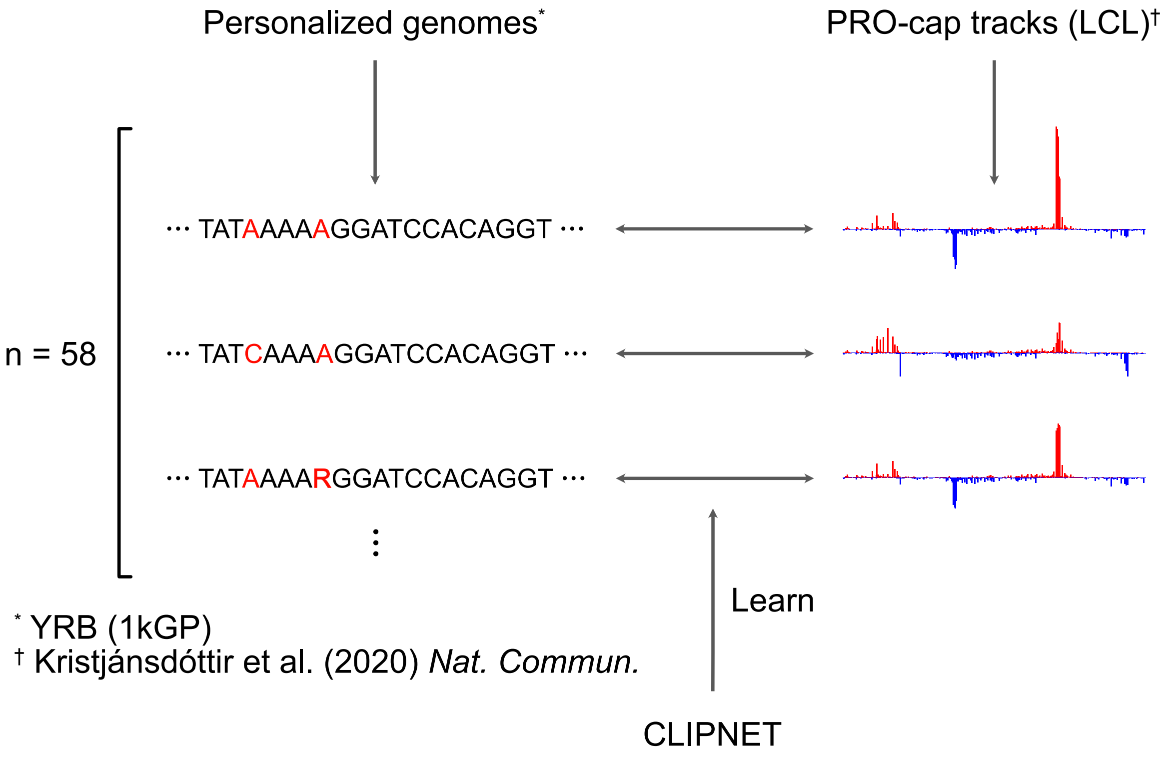
CLIPNET

Convolutionally Learned, Initiation Predicting neural NETwork

1 kb sequence -> 500 bp transcription initiation @ single nucleotide resolution



Learning the genotype-phenotype map



Personalized genomes[*]

PRO-cap tracks (LCL)[†]

n = 58

··· TATAAAAAGGATCCACAGGT ···

··· TATCAAAAGGATCCACAGGT ···

··· TATAAAARGGATCCACAGGT ···

⋮

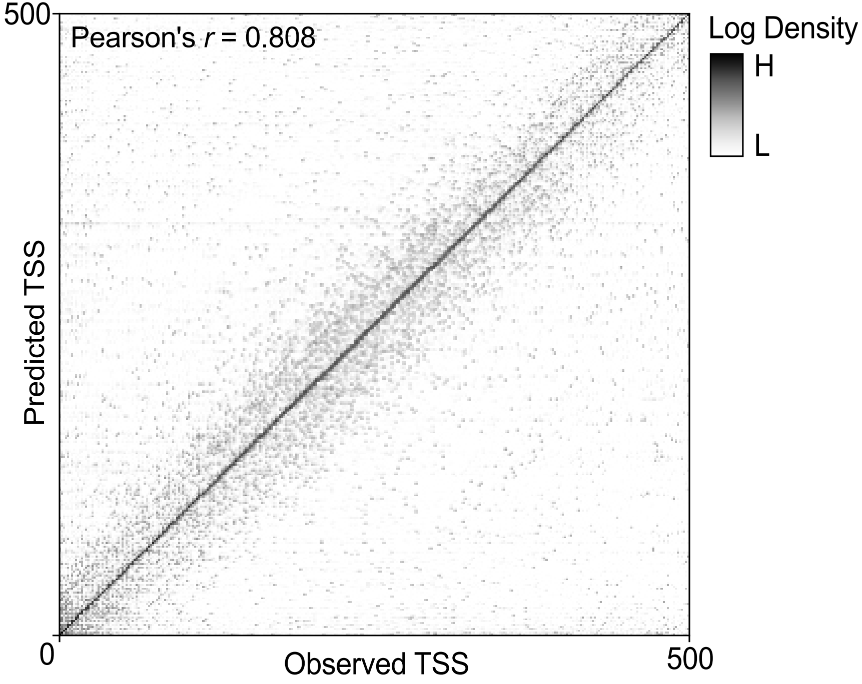
Learn

CLIPNET

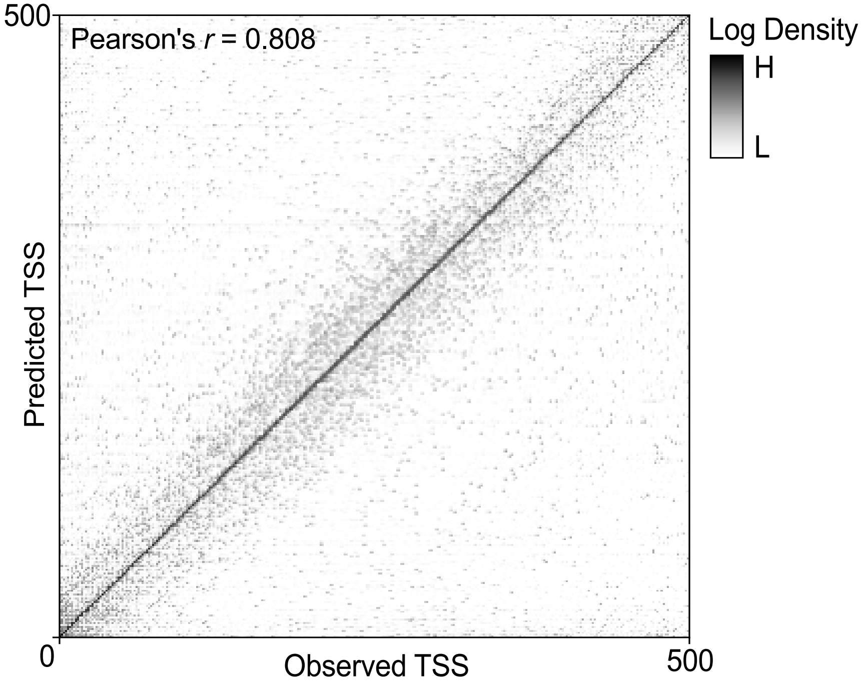
[*] YRB (1kGP)

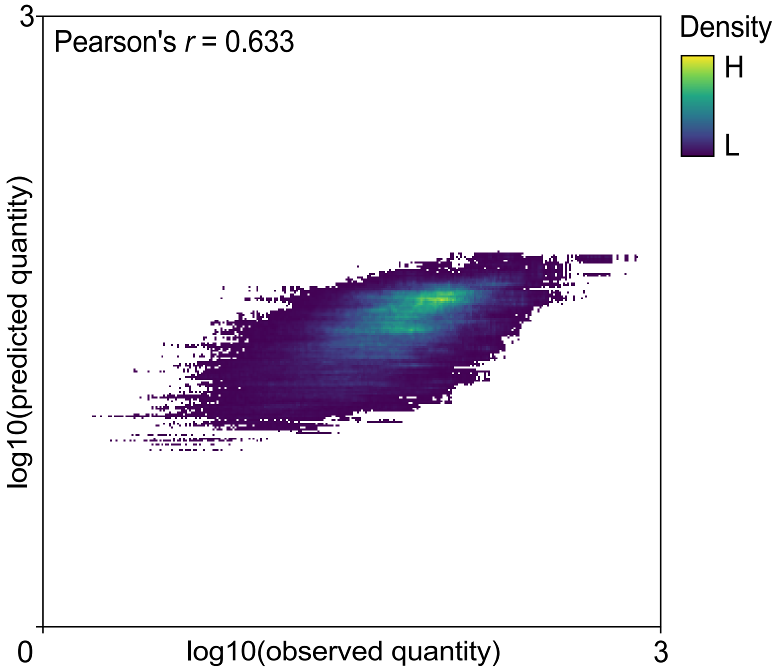
[†] Kristjánsdóttir et al. (2020) *Nat. Commun.*

CLIPNET accurately predicts TSN position



… and initiation quantity

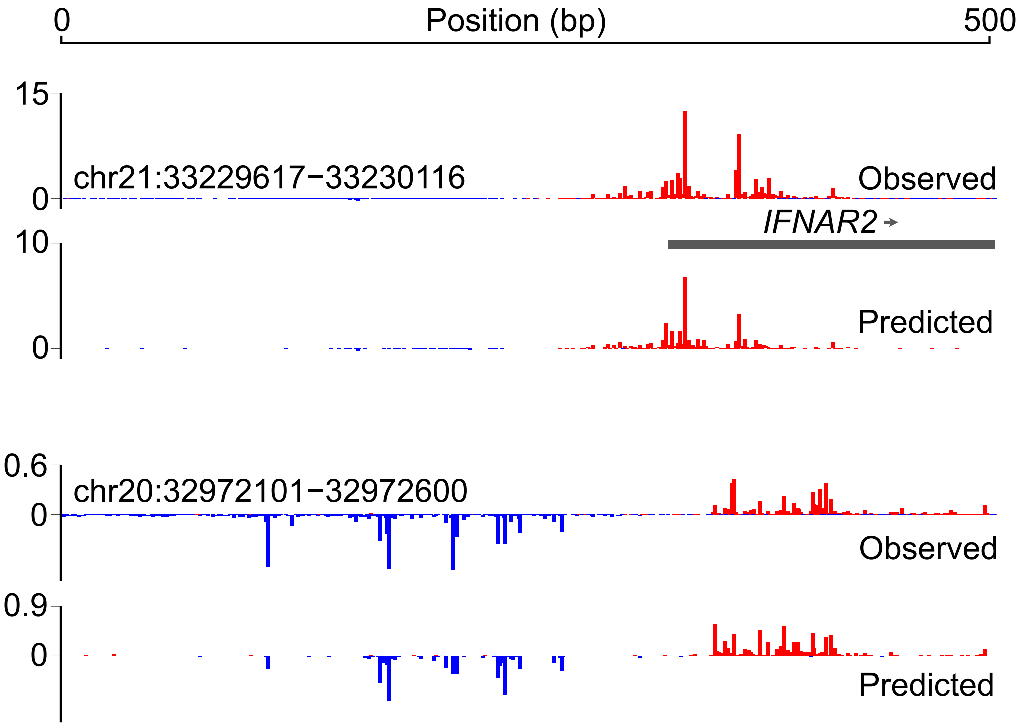




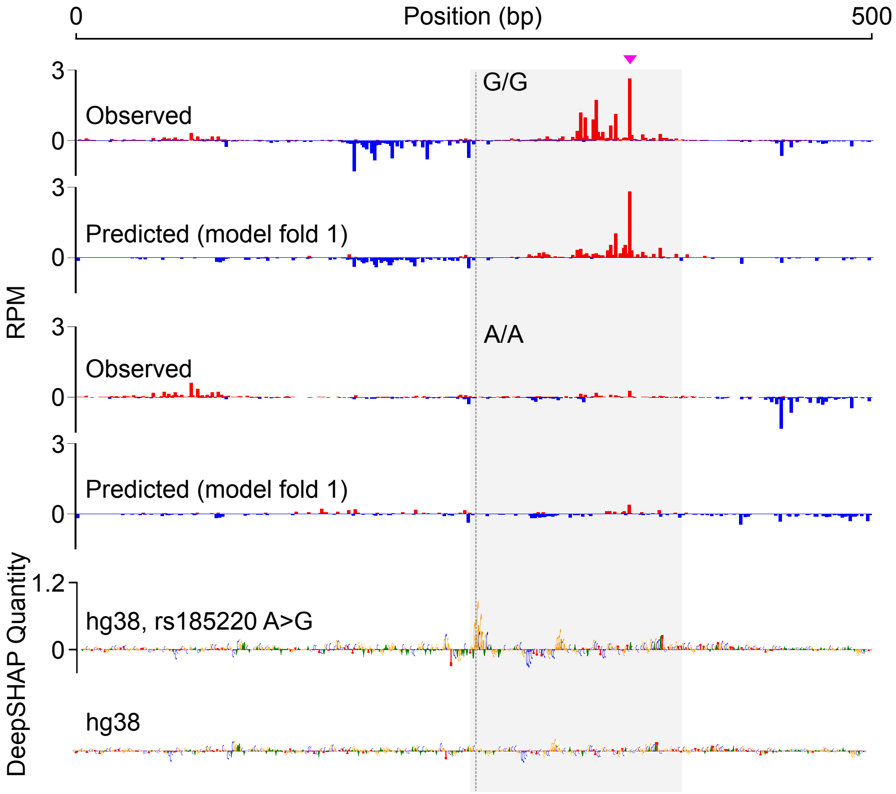
Example tracks

Promoter

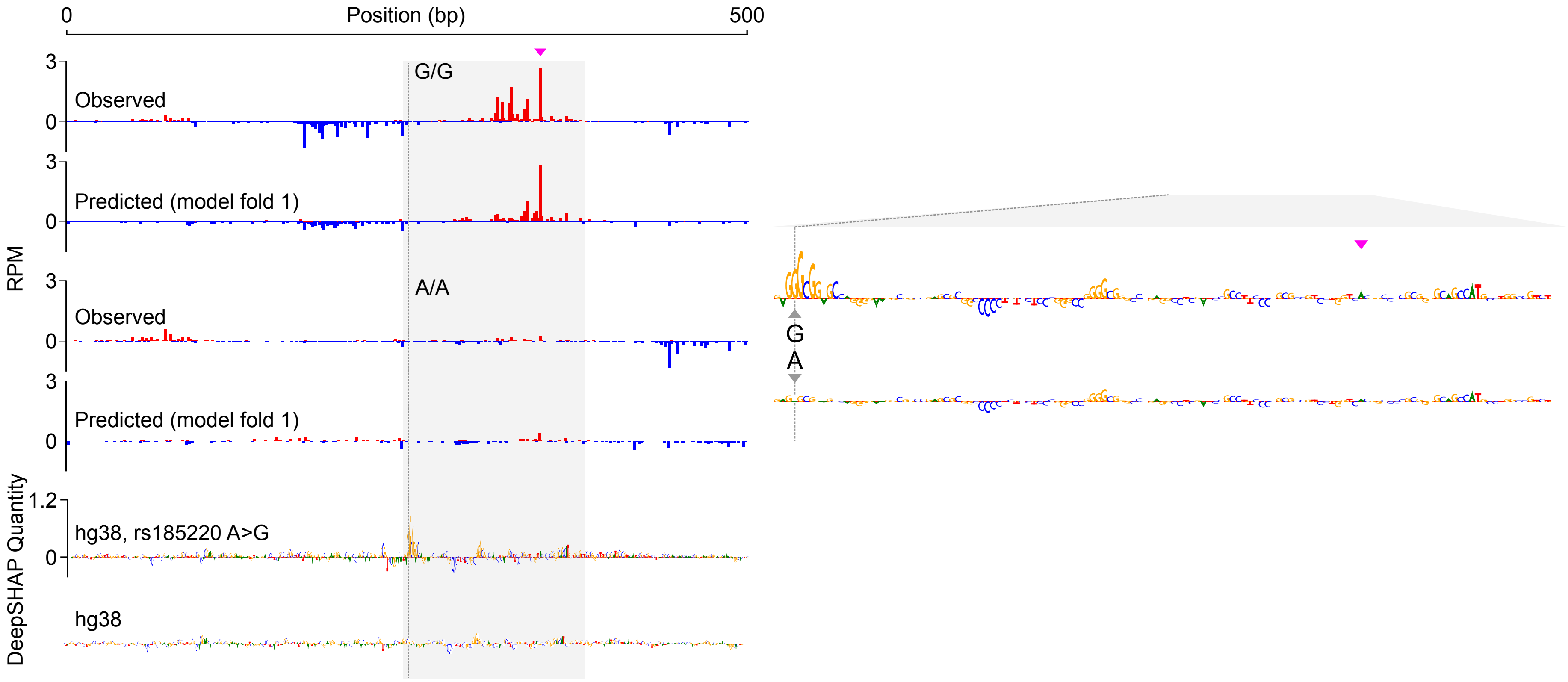
Enhancer



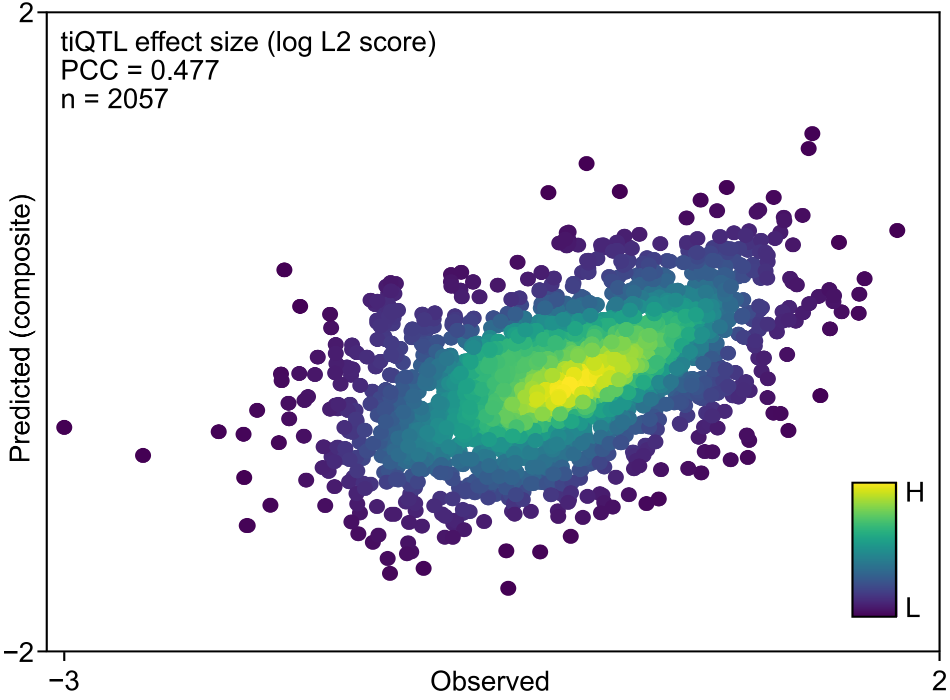
How well does CLIPNET predict QTL effects?

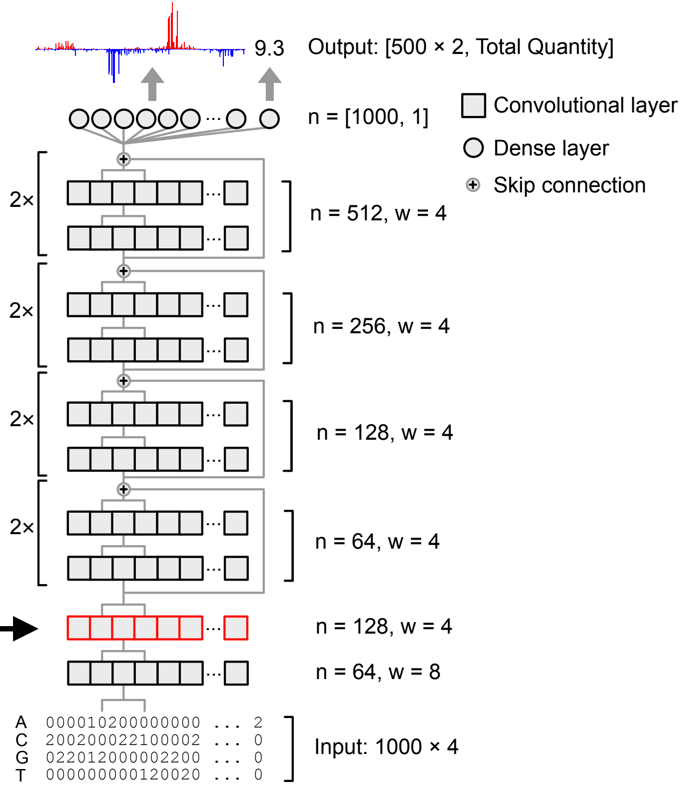


How well does CLIPNET predict QTL effects?



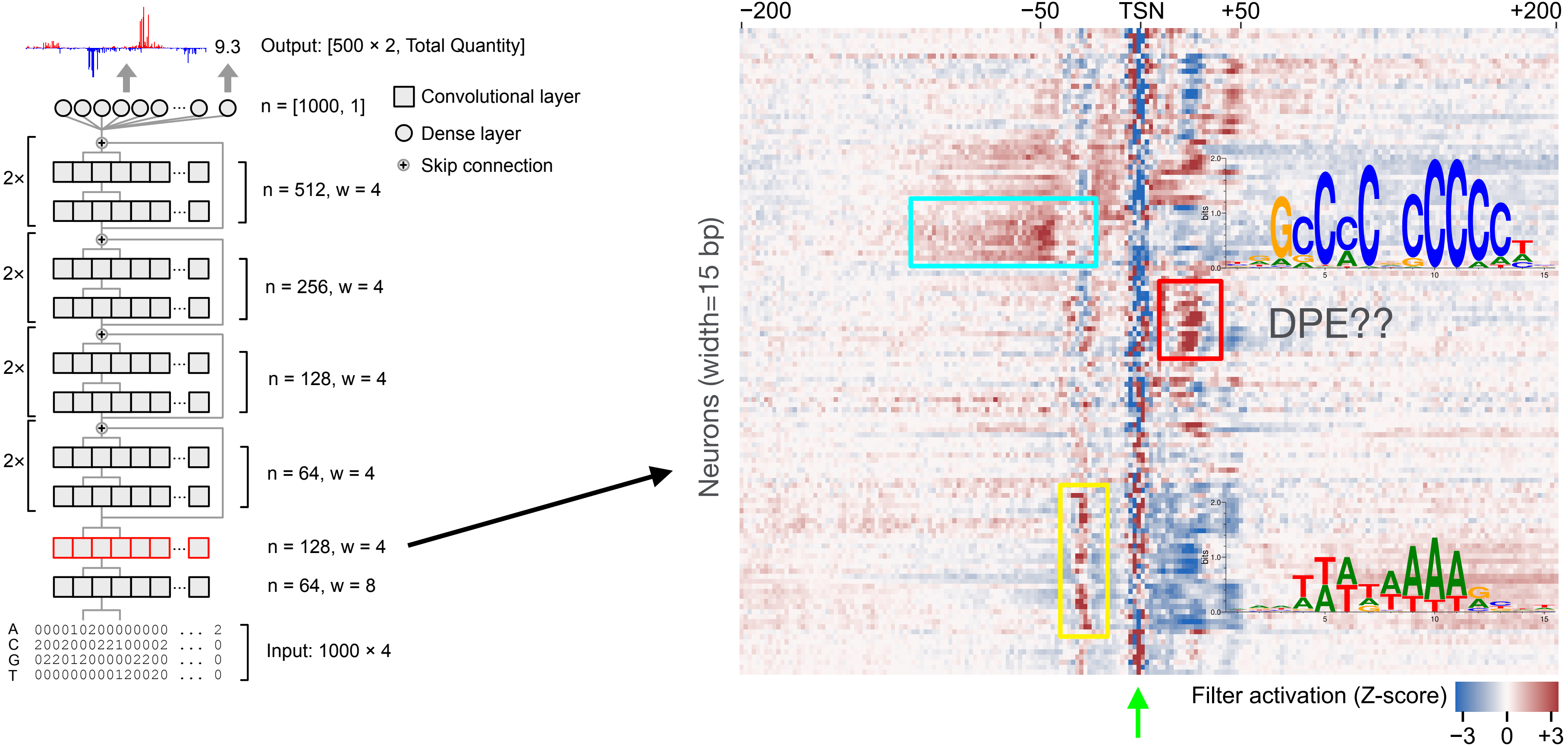
How well does CLIPNET predict QTL effects?





Where are these neurons most active?

Activation patterns suggest multiple motif classes



9.3

Output: [500 × 2, Total Quantity]

n = [1000, 1]

Convolutional layer

Dense layer

Skip connection

2× n = 512, w = 4

2× n = 256, w = 4

2× n = 128, w = 4

2× n = 64, w = 4

n = 128, w = 4

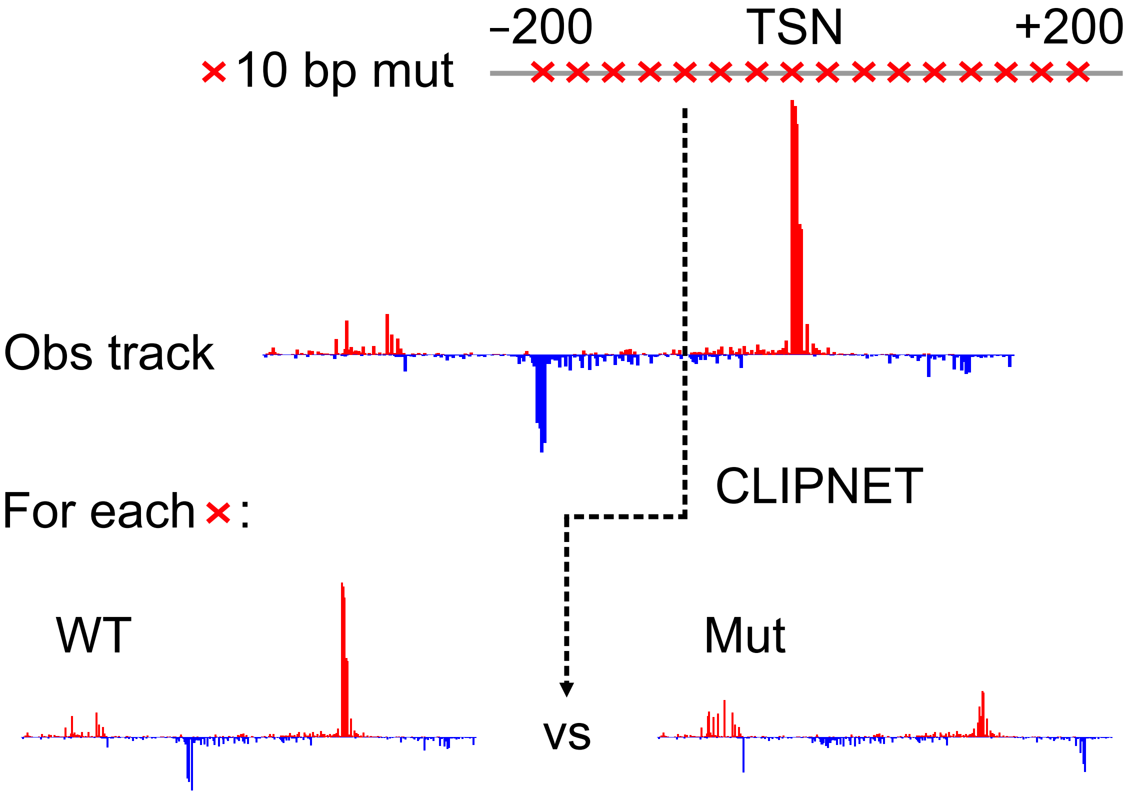
n = 64, w = 8

```
A  000010200000000  . . .  2
C  200200022100002  . . .  0
G  022012000002200  . . .  0
T  000000000120020  . . .  0
```

Input: 1000 × 4

Filter activation (Z-score)

Quantify sequence importance by tiling mutagenesis



–200 TSN +200

× 10 bp mut

Obs track

CLIPNET

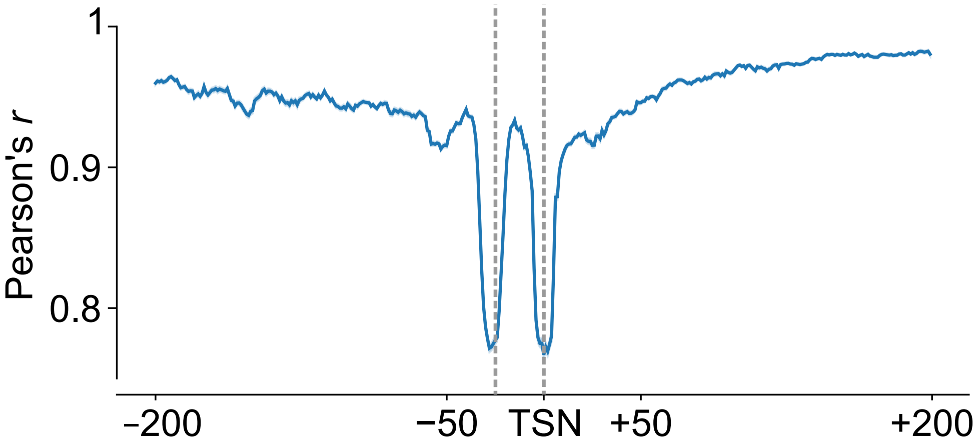
For each × :

WT

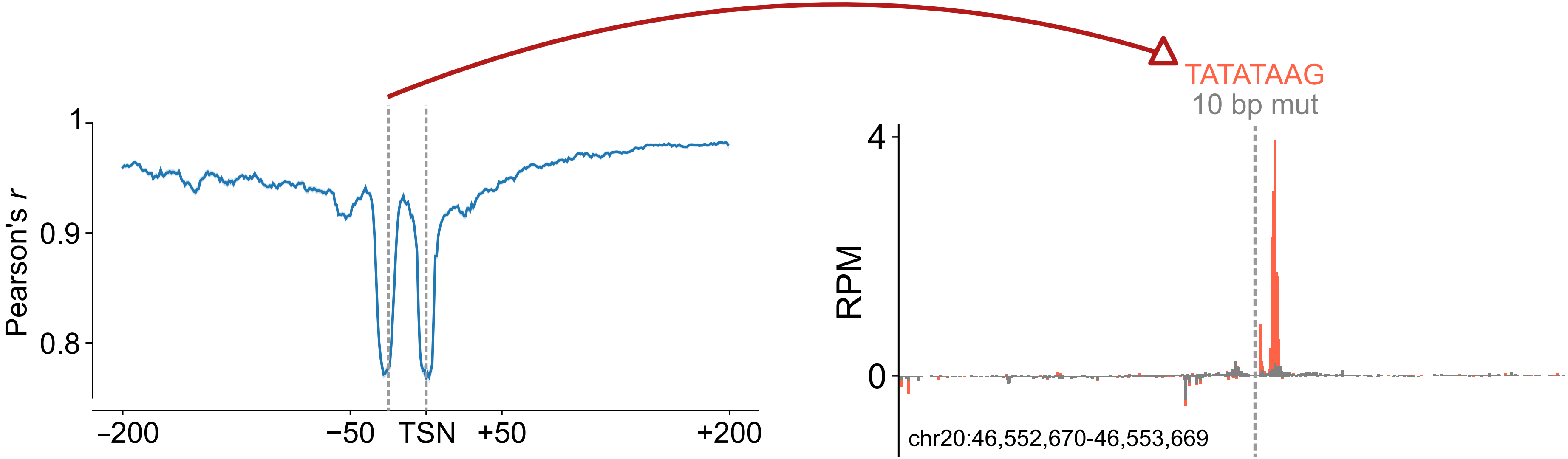
Mut

vs

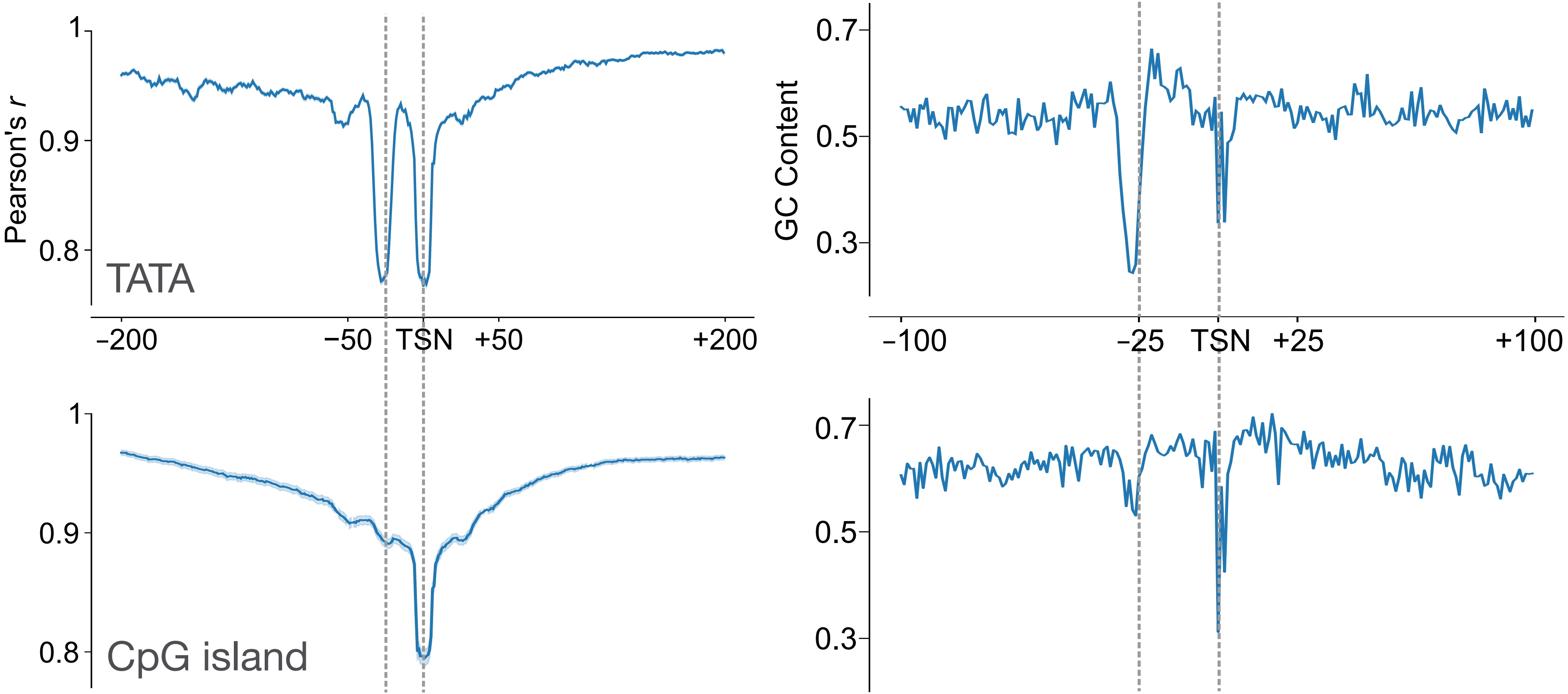
Tiling mutagenesis of TATA CREs



Tiling mutagenesis of TATA CREs

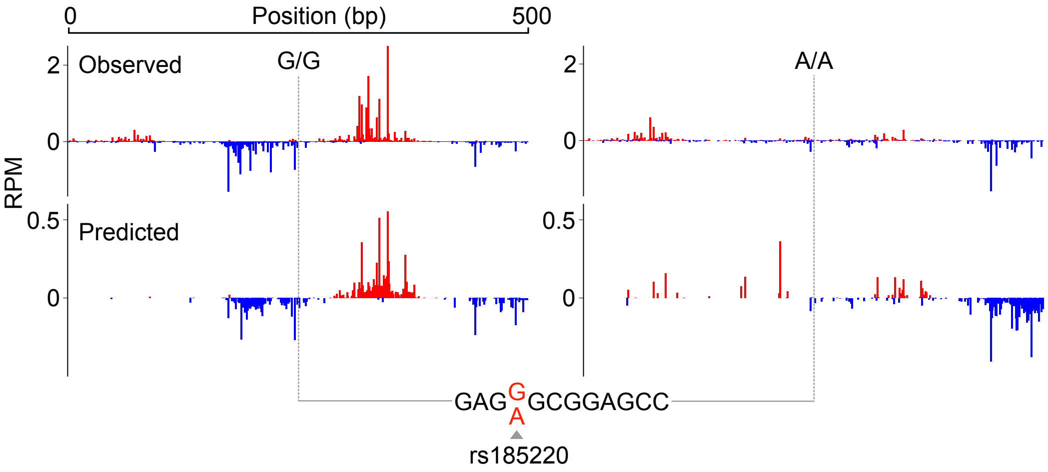


Tiling mutagenesis of TATA CREs vs CpG island promoters

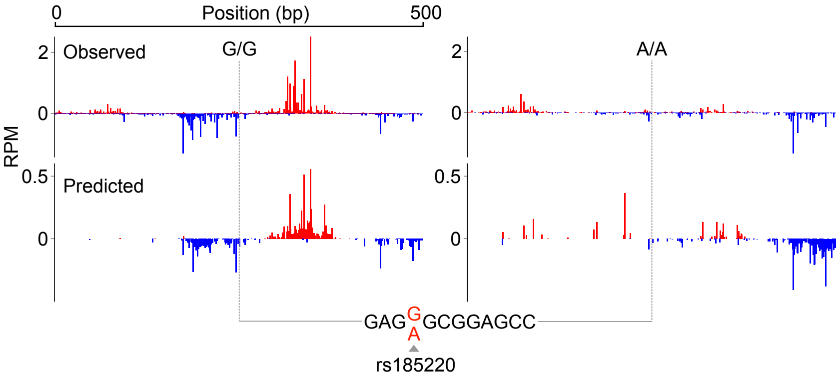


What about transcriptional activators?

Recall this QTL in an SP1 binding site:

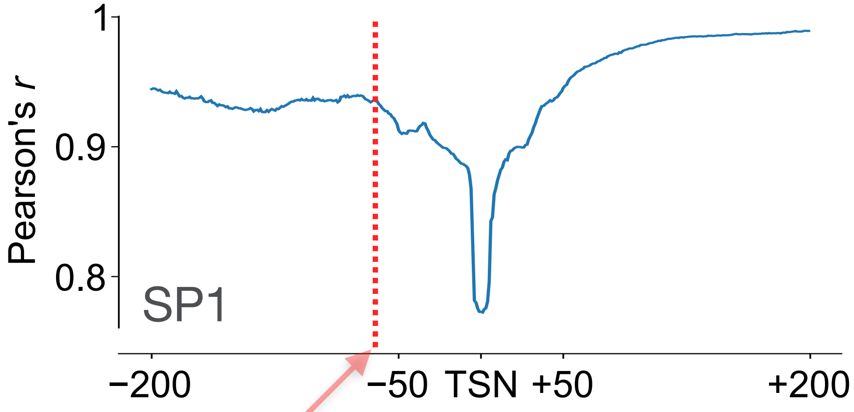


What about transcriptional activators?



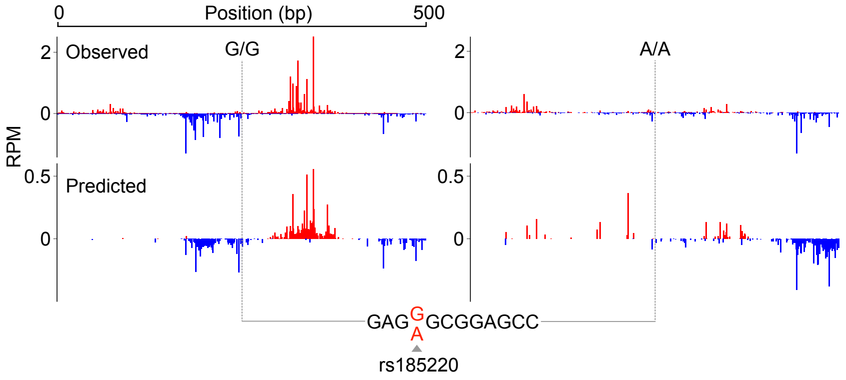
SP1 clearly affects quantity

Tiling mutagenesis of SP1 CREs:



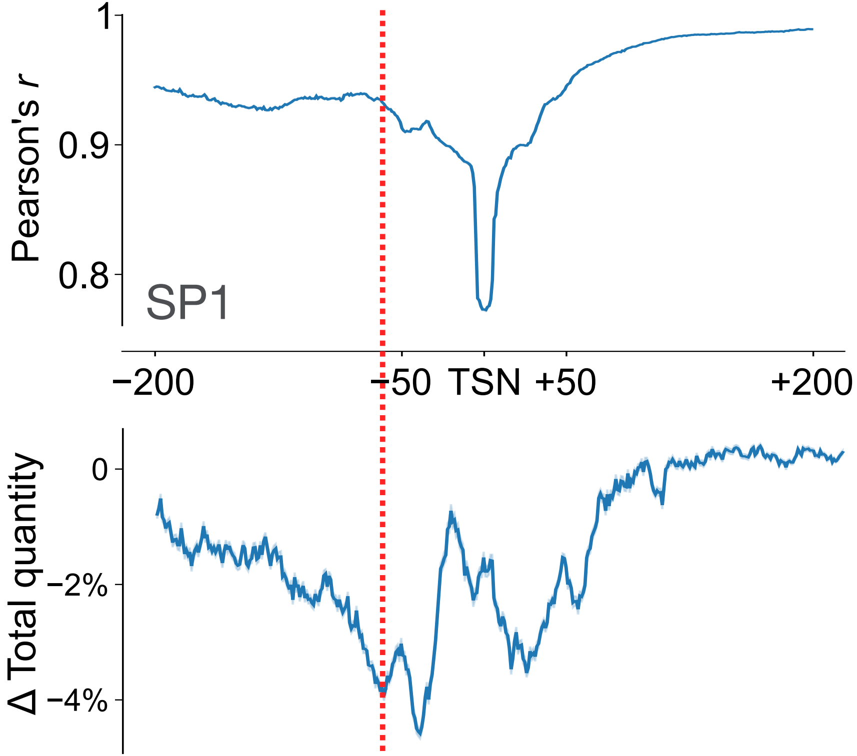
Median position of SP1 motif

What about transcriptional activators?

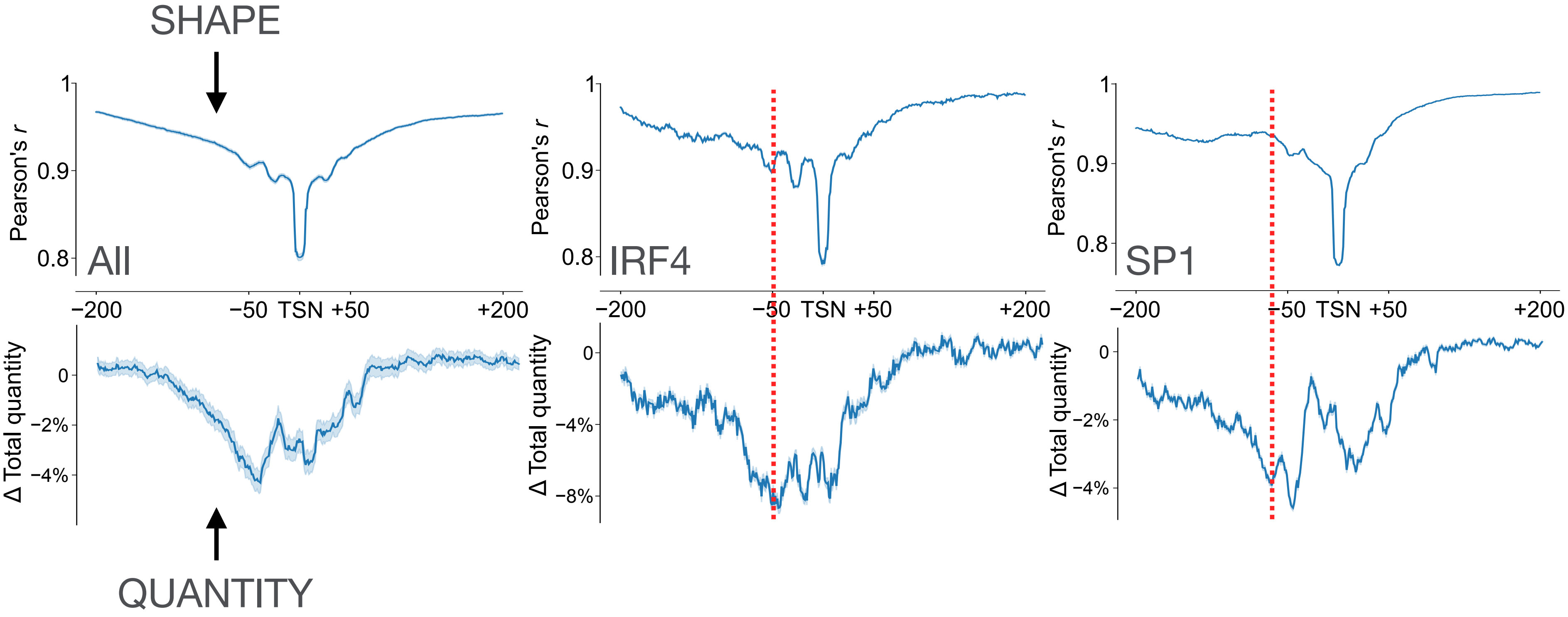


SP1 clearly affects quantity

Tiling mutagenesis of SP1 CREs:



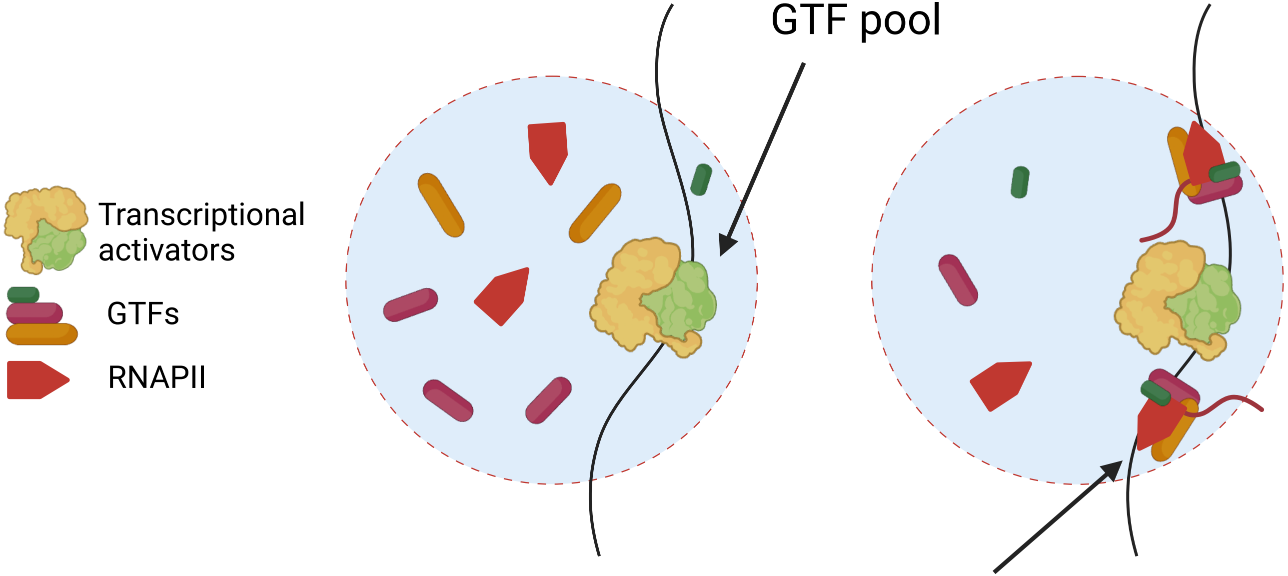
What about transcriptional activators?



SHAPE

QUANTITY

Distinct classes of motifs



Activator binding controls size of GTF pool

GTFs position RNAPII, establish TSS

Cornell
Baker Institute for Animal Health

- Funding / Resources:
  - NIH T32HD057854
  - XSEDE / ACCESS BIO210011P
- Data / Code:
  - Hojoong Kwak
  - Li Yao
  - Haiyuan Yu

NIH
National Institutes of Health

ACCESS | Advancing Innovation

Cornell
Baker Institute for Animal Health

ayh8@cornell.edu / @missing_a_rib
cgd24@cornell.edu / @charlesdanko

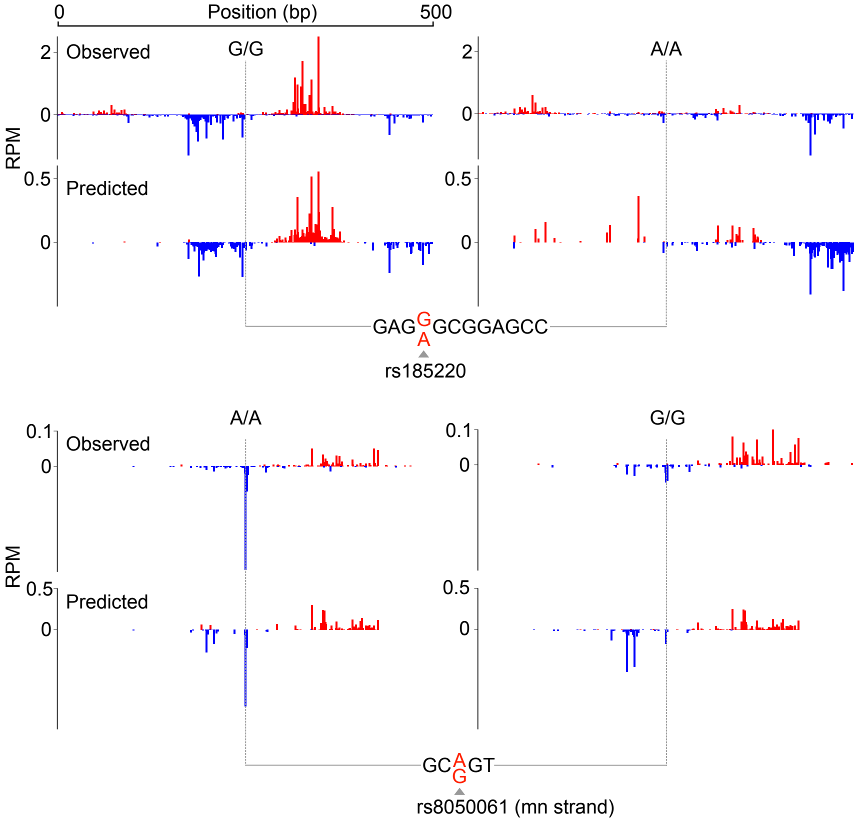
github.com/Danko-Lab
dankolab.org

Look for our preprint to come out shortly!

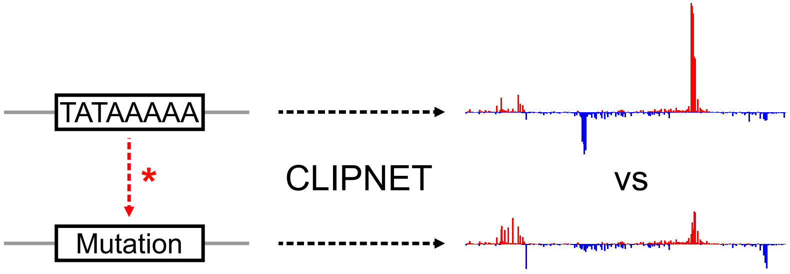
Divider slide

- Backup slides after this point

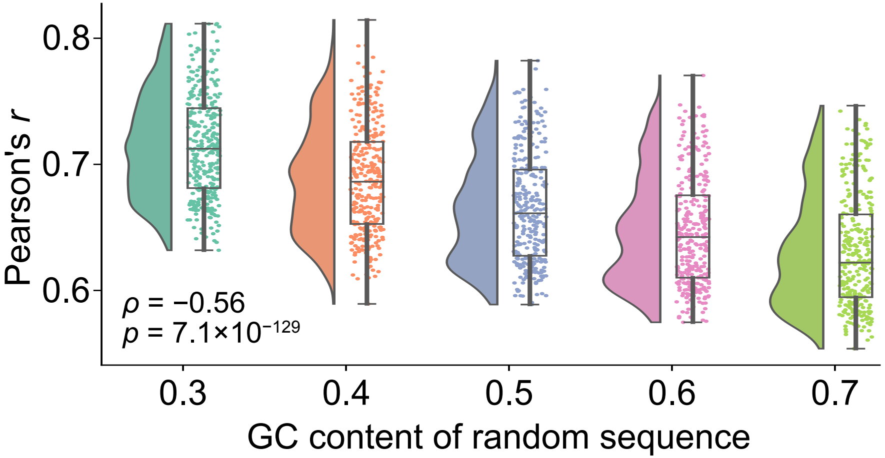
How well does CLIPNET predict QTL effects?



Higher AT% mutations less disruptive



* Random 8-mers: CGACATTC



What about transcriptional activators?