STANFORD SCHOOL of MEDICINE | Genetics

STANFORD COMPUTER SCIENCE

Not Just a Black Box: Interpretable Deep Learning for Genomics

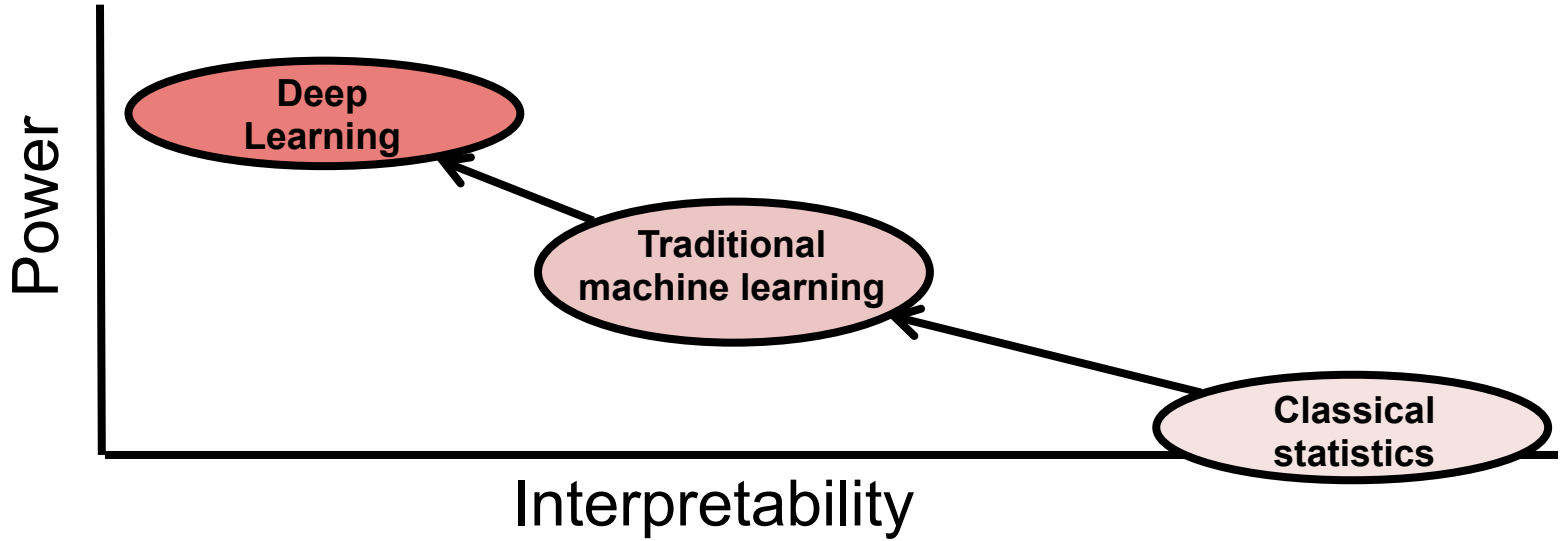
Presented by: Avanti Shrikumar

With great power comes really poor interpretability…

NATURE | NEWS FEATURE

Can we open the black box of AI?

Artificial intelligence is everywhere. But before scientists trust it, they first need to understand how machines learn.

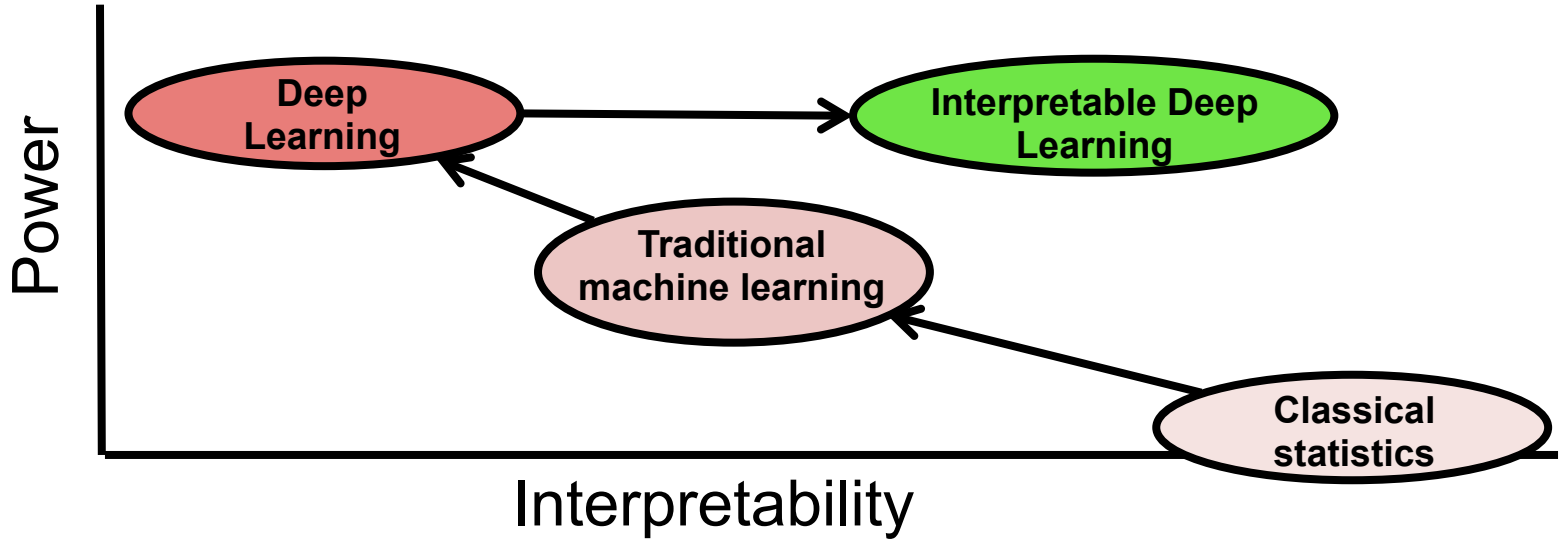


With great power comes really poor interpretability…

NATURE | NEWS FEATURE

Can we open the black box of AI?

Artificial intelligence is everywhere. But before scientists trust it, they first need to understand how machines learn.



Questions for the model

- Which parts of the input are the most important for making a given prediction?

Questions for the model

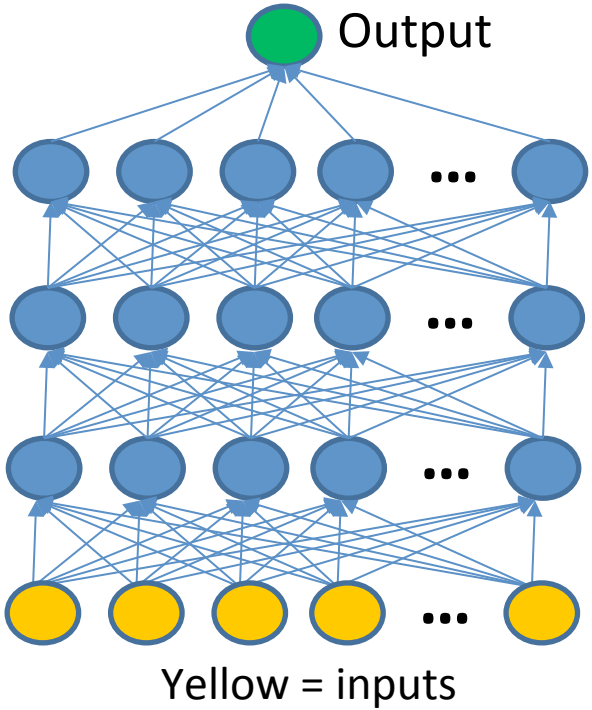
- Which parts of the input are the most important for making a given prediction?
- What are the recurring patterns in the input?

Questions for the model

- Which parts of the input are the most important for making a given prediction?
- What are the recurring patterns in the input?

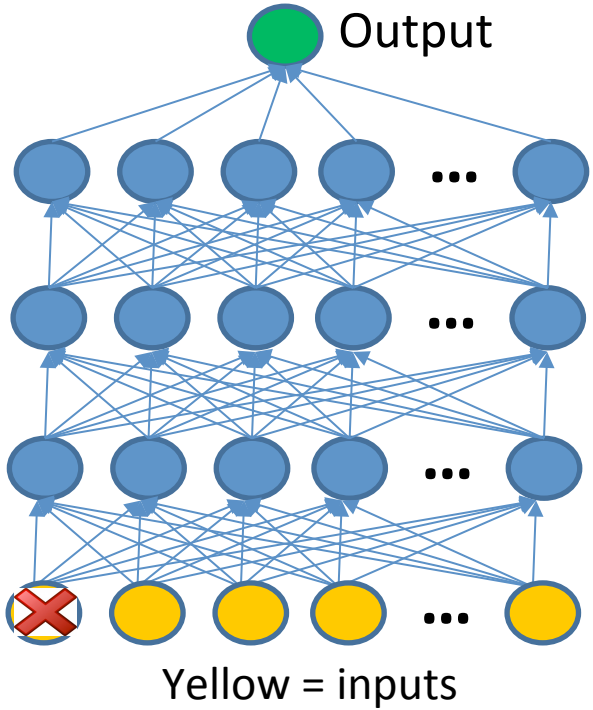
How can we find the important parts of the input for a given prediction?

Idea 1: perturbation



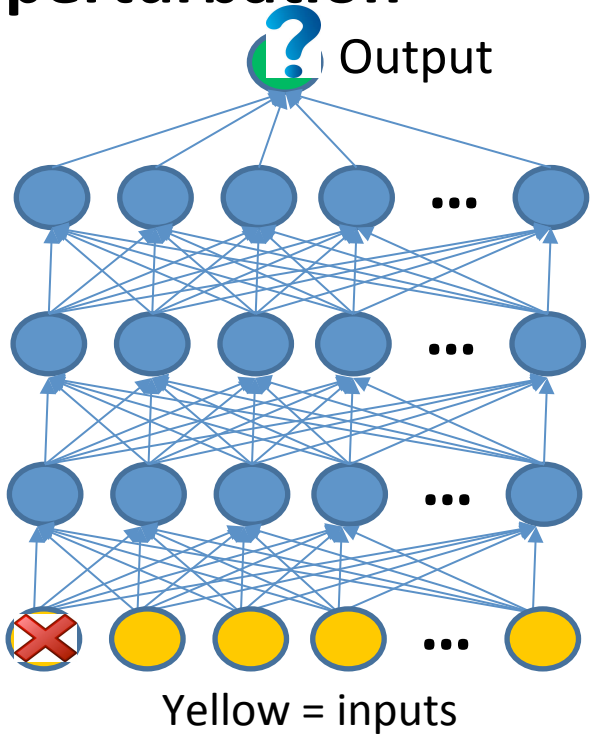
How can we find the important parts of the input for a given prediction?

Idea 1: perturbation



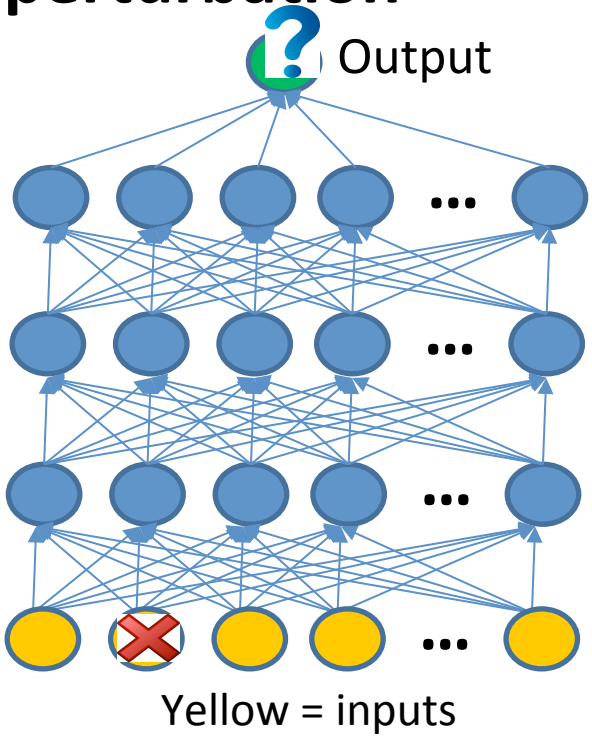
How can we find the important parts of the input for a given prediction?

Idea 1: perturbation



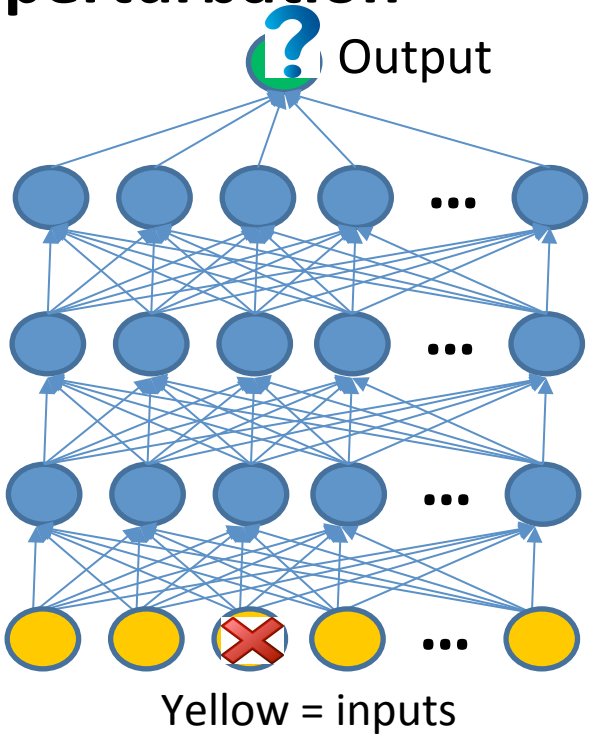
How can we find the important parts of the input for a given prediction?

Idea 1: perturbation



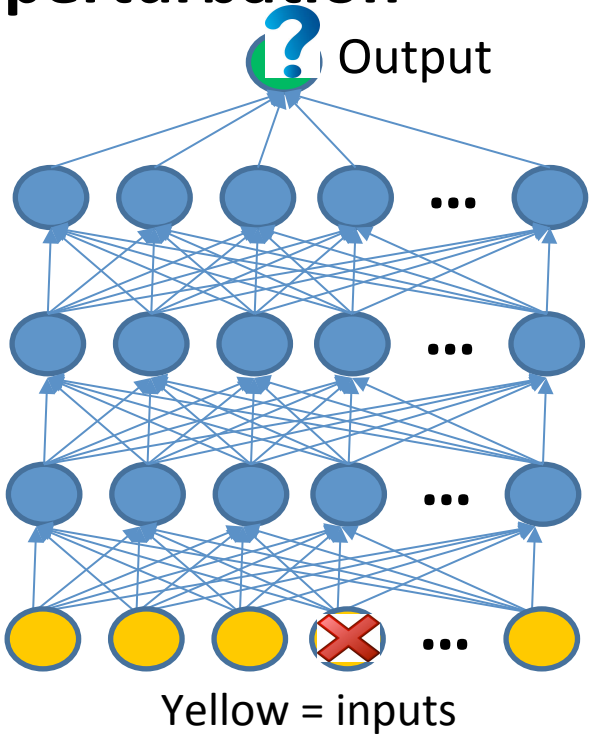
How can we find the important parts of the input for a given prediction?

Idea 1: perturbation



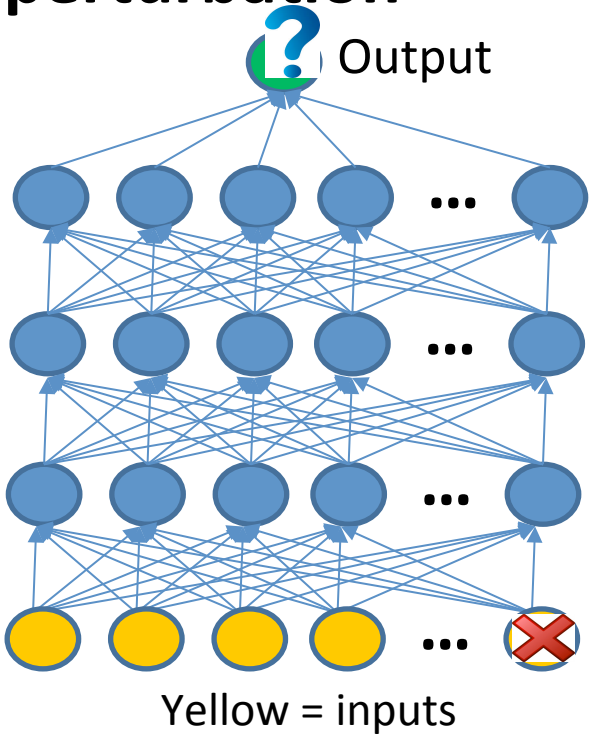
How can we find the important parts of the input for a given prediction?

Idea 1: perturbation



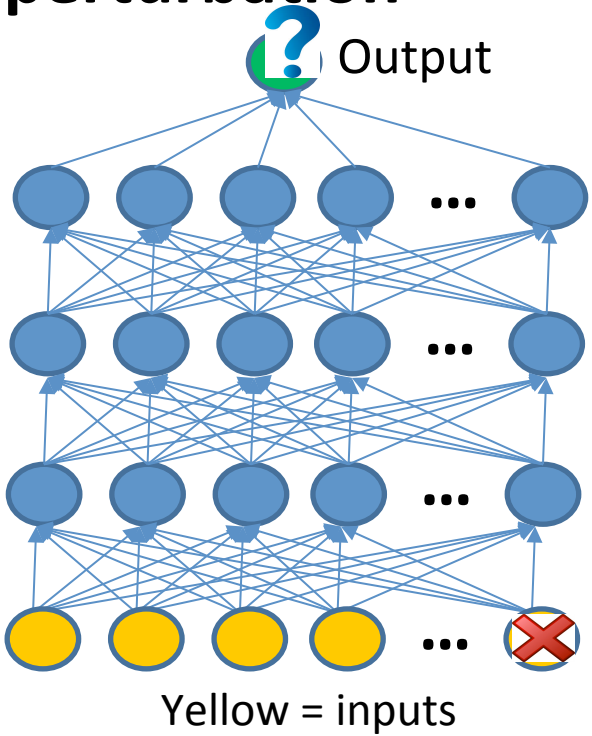
How can we find the important parts of the input for a given prediction?

Idea 1: perturbation



How can we find the important parts of the input for a given prediction?

Idea 1: perturbation

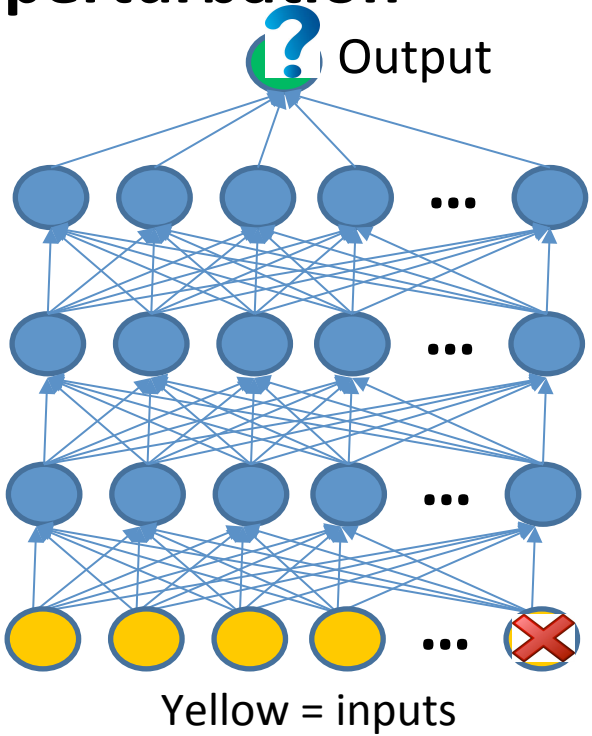


Drawbacks

1) Computational efficiency - requires one forward prop for each perturbation

How can we find the important parts of the input for a given prediction?

Idea 1: perturbation

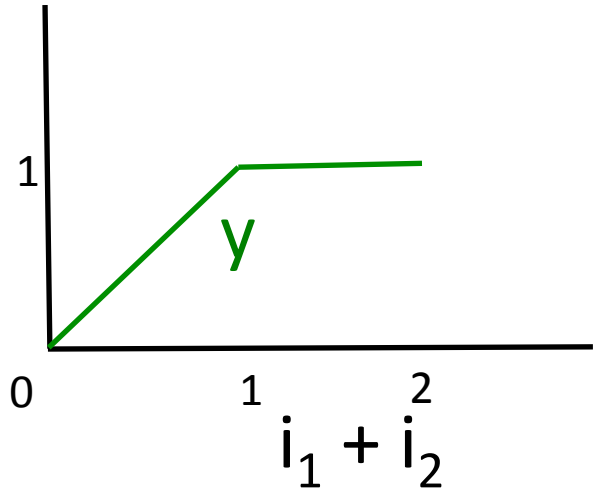


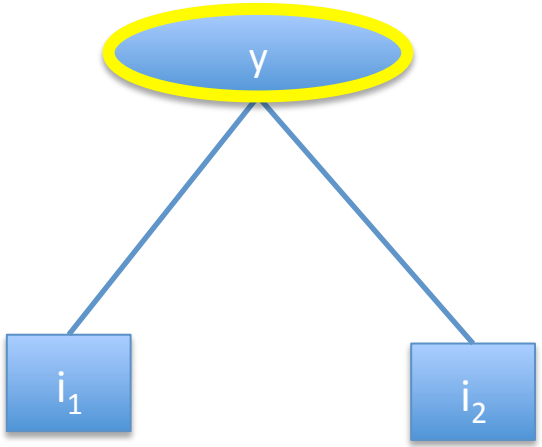
Drawbacks

1) Computational efficiency - requires one forward prop for each perturbation

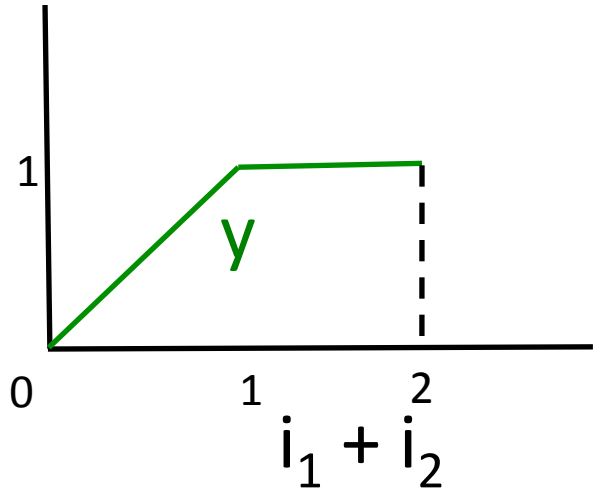
2) Saturation

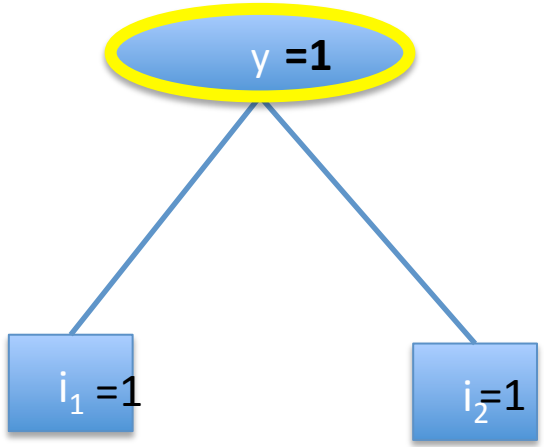
Saturation problem illustrated



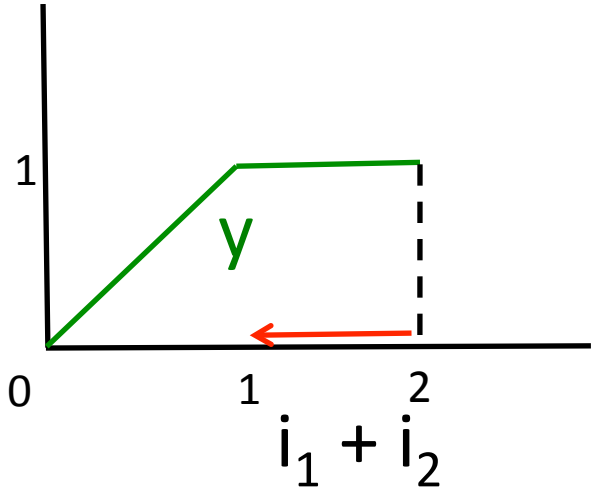


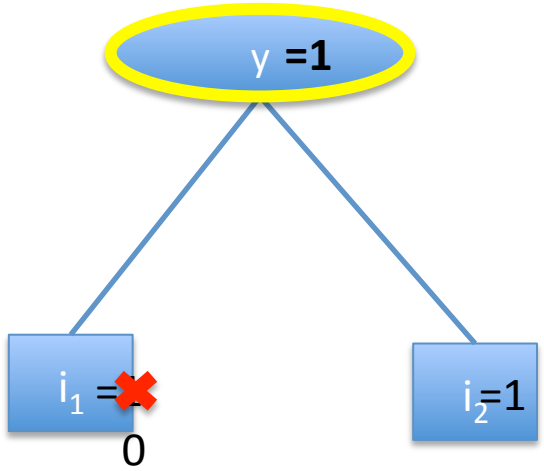
Saturation problem illustrated





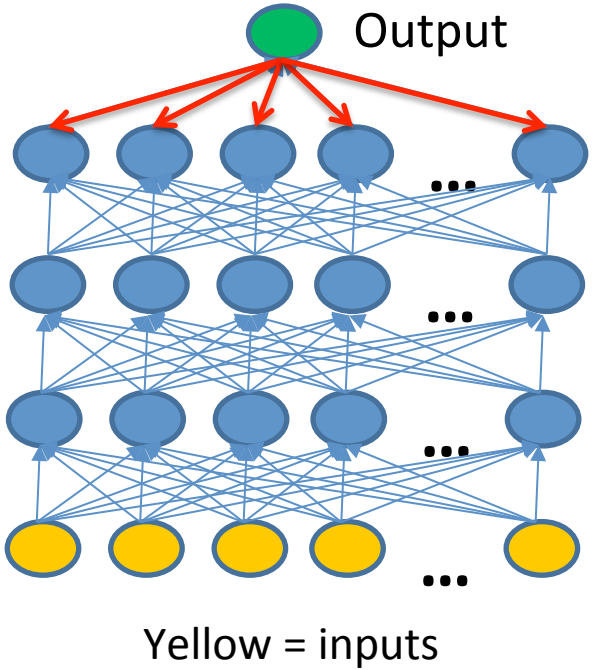
Saturation problem illustrated





How can we find the important parts of the input for a given prediction?

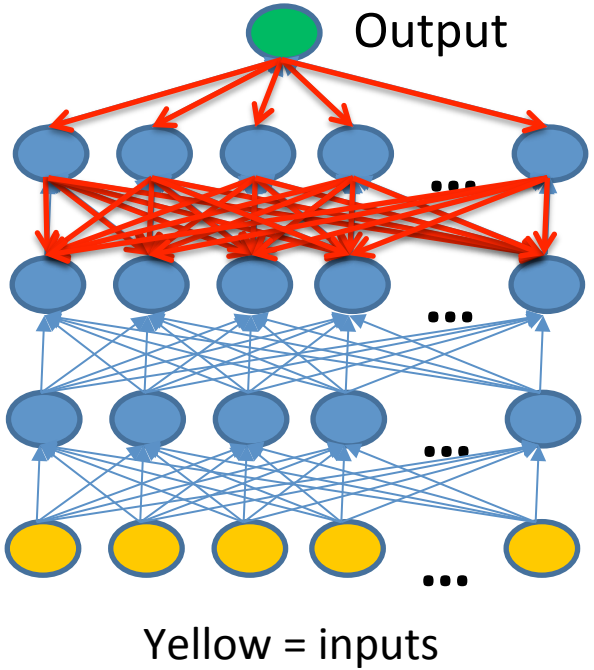
Idea 2: backpropagate importance



Yellow = inputs

How can we find the important parts of the input for a given prediction?

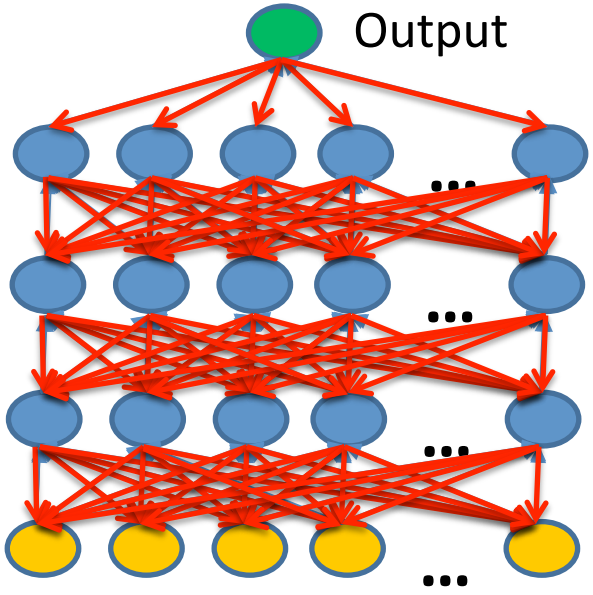
Idea 2: backpropagate importance



Yellow = inputs

How can we find the important parts of the input for a given prediction?

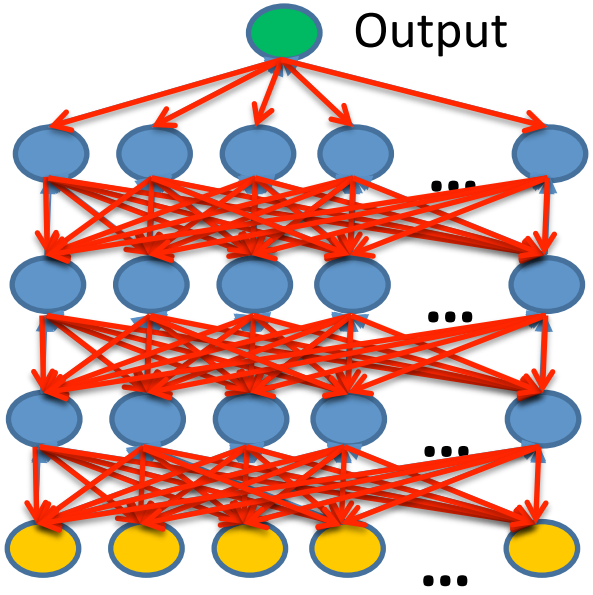
Idea 2: backpropagate importance



Yellow = inputs

How can we find the important parts of the input for a given prediction?

Idea 2: backpropagate importance



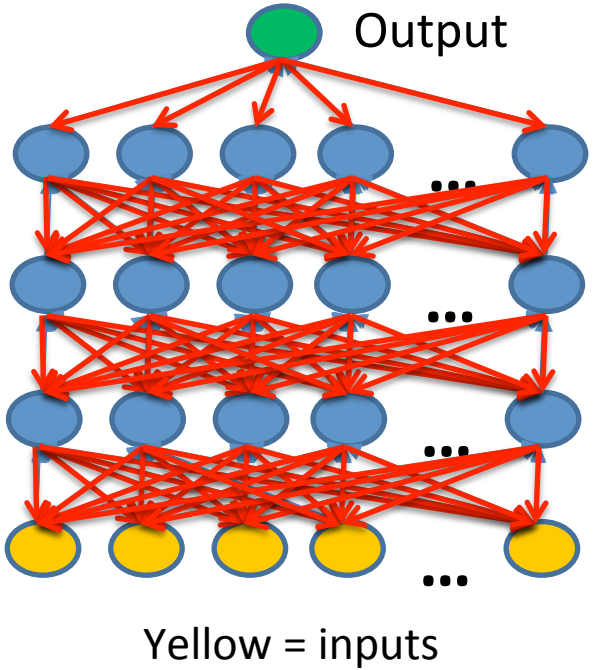
Yellow = inputs

Examples:

- Gradients (Simonyan et al.)
- Deconvolutional Networks (Zeiler & Fergus)
- Guided Backpropagation (Springenberg et al.)
- Layerwise Relevance Propagation (Bach et al.)
- Integrated Gradients (Sundararajan et al.)

How can we find the important parts of the input for a given prediction?

Idea 2: backpropagate importance

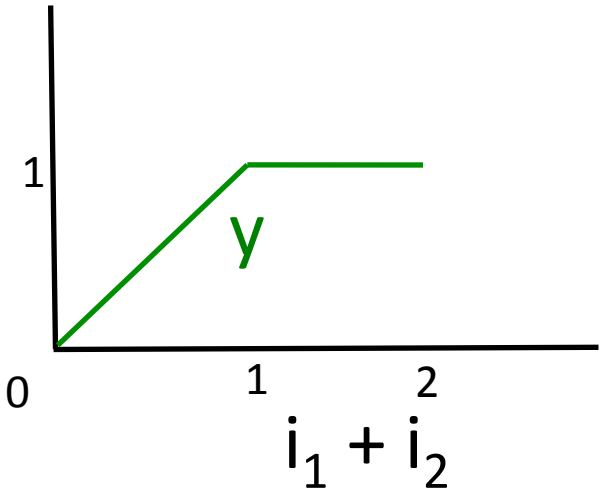


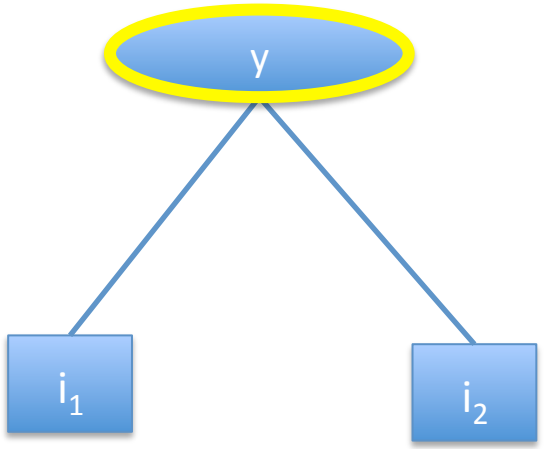
Yellow = inputs

Examples:

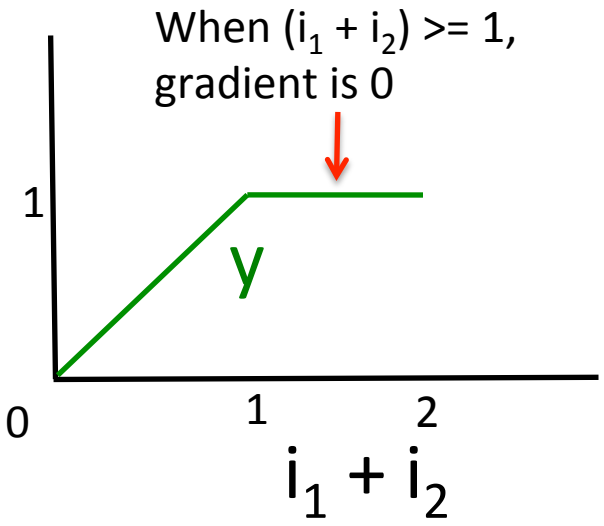
- Gradients (Simonyan et al.)
- Deconvolutional Networks (Zeiler & Fergus)
- Guided Backpropagation (Springenberg et al.)
- Layerwise Relevance Propagation (Bach et al.)
- Integrated Gradients (Sundararajan et al.)
- **DeepLIFT (Learning Important FeaTures)**
  - https://github.com/kundajelab/deeplift, ICML 2017
  - With Peyton Greenside and Anshul Kundaje

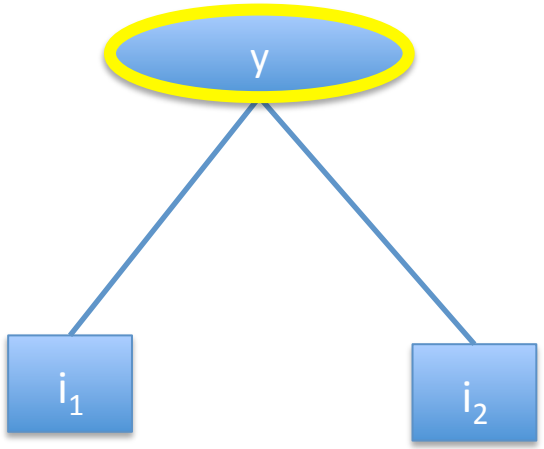
Saturation revisited





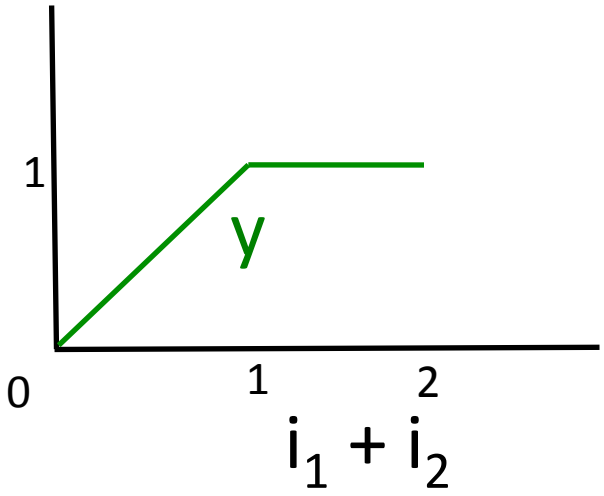
Saturation revisited

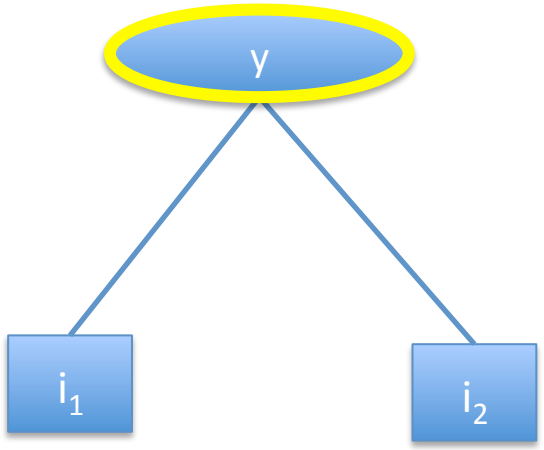




The DeepLIFT solution: difference from reference

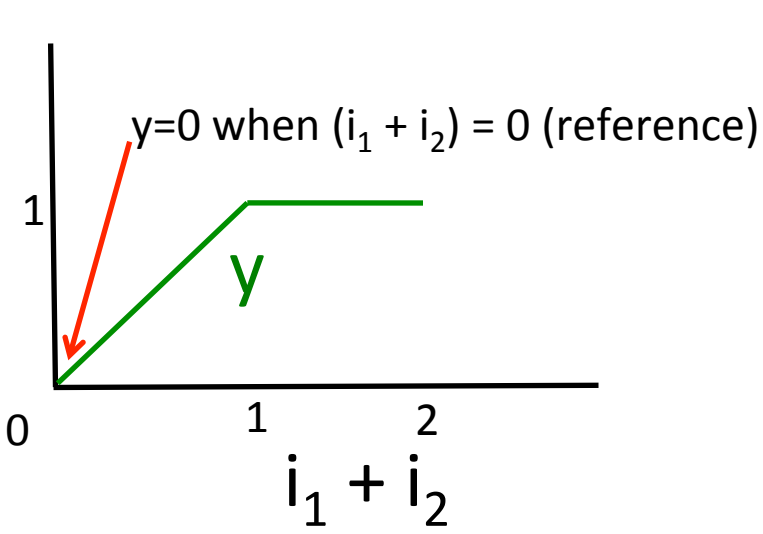
Reference: $i_1=0$ & $i_2=0$

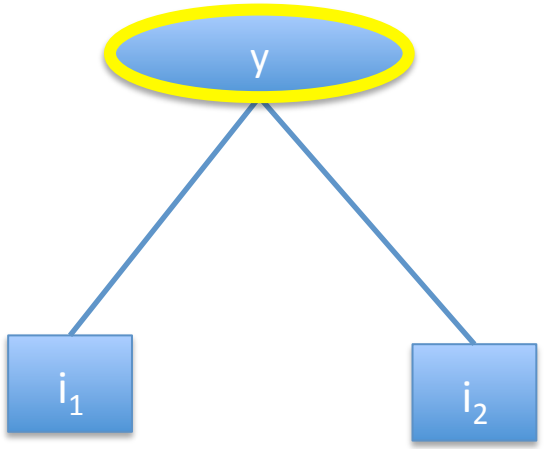




The DeepLIFT solution: difference from reference

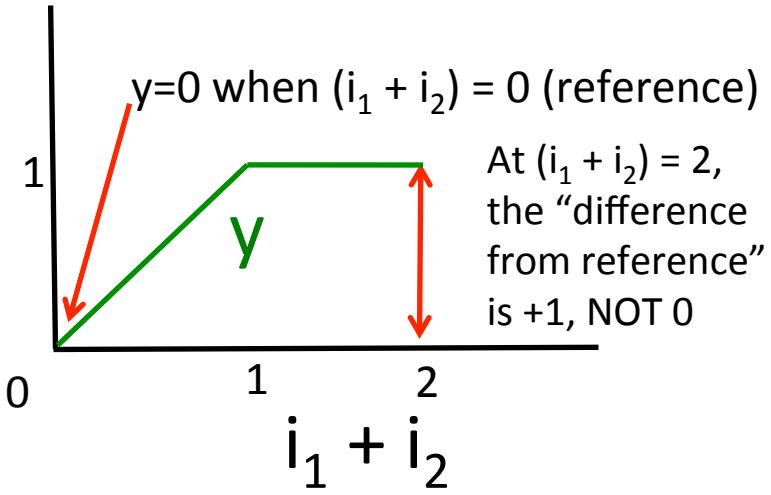
Reference: $i_1=0$ & $i_2=0$

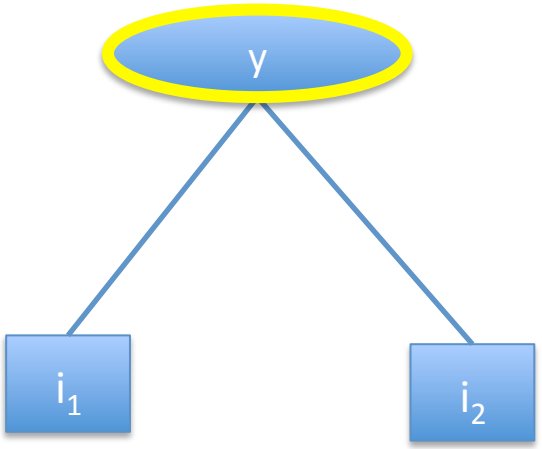




The DeepLIFT solution: difference from reference

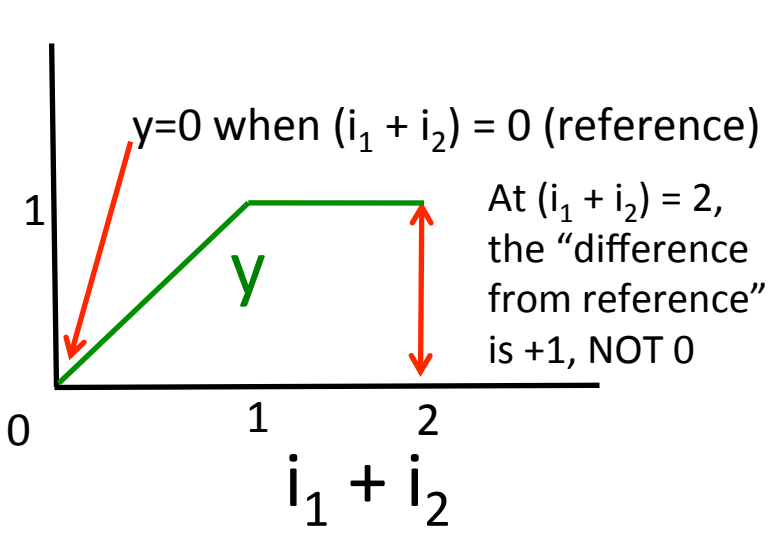
Reference: $i_1=0$ & $i_2=0$

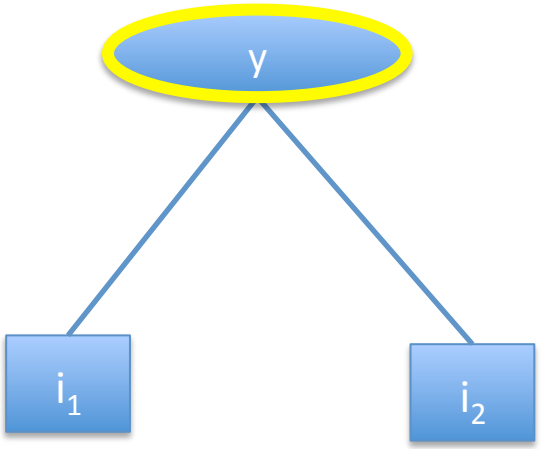




The DeepLIFT solution: difference from reference

Reference: $i_1=0$ & $i_2=0$



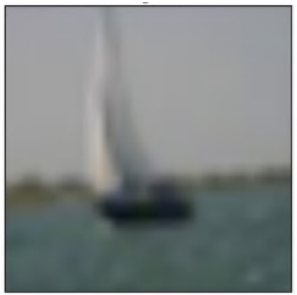


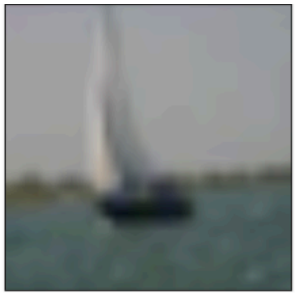
DeepLIFT addresses other failure modes besides saturation (see paper)

Reference matters!

CIFAR10 model, class = “ship”

Original





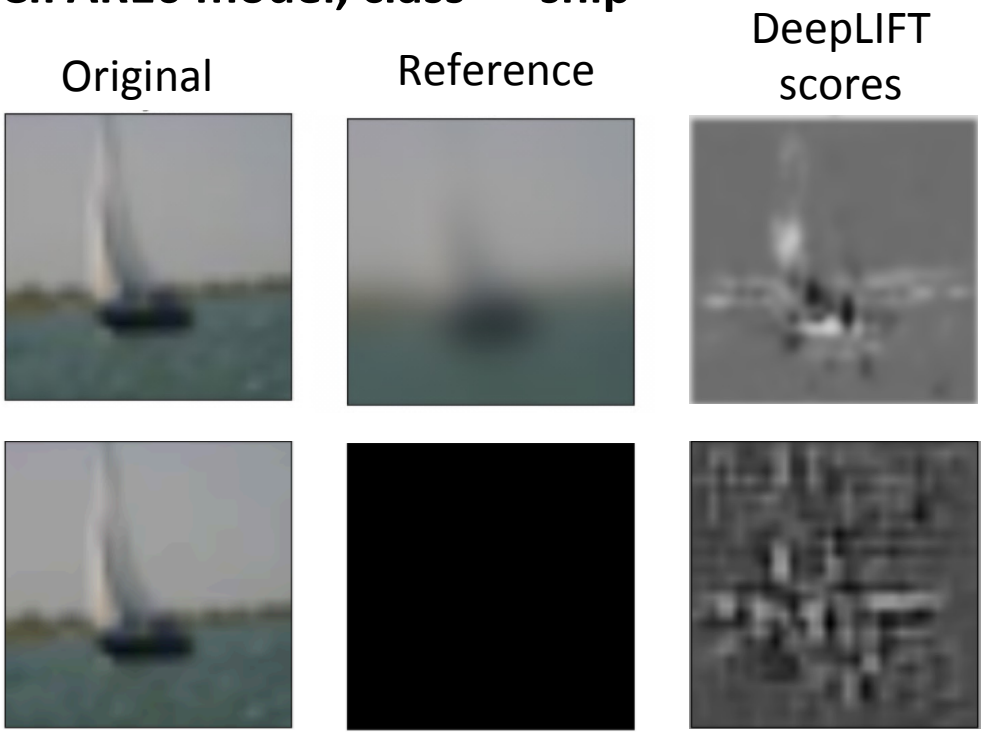
Reference matters!

CIFAR10 model, class = “ship”

Original | Reference | DeepLIFT scores

Reference matters!

CIFAR10 model, class = “ship”

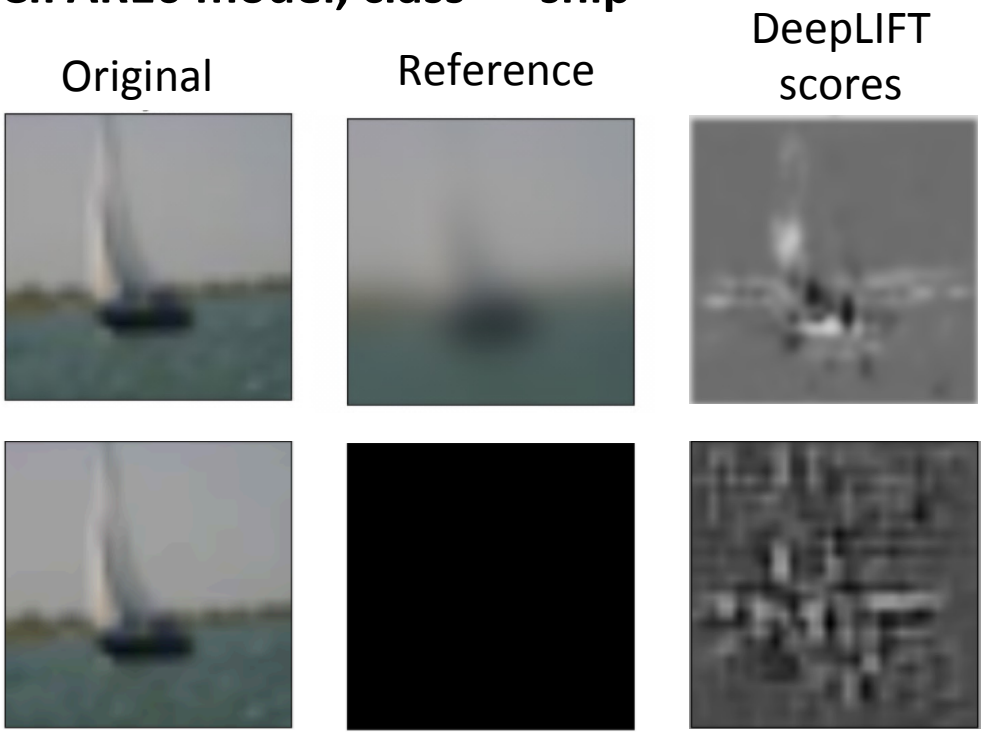


Suggestions on how to pick a reference:

- MNIST: all zeros (background)

Reference matters!

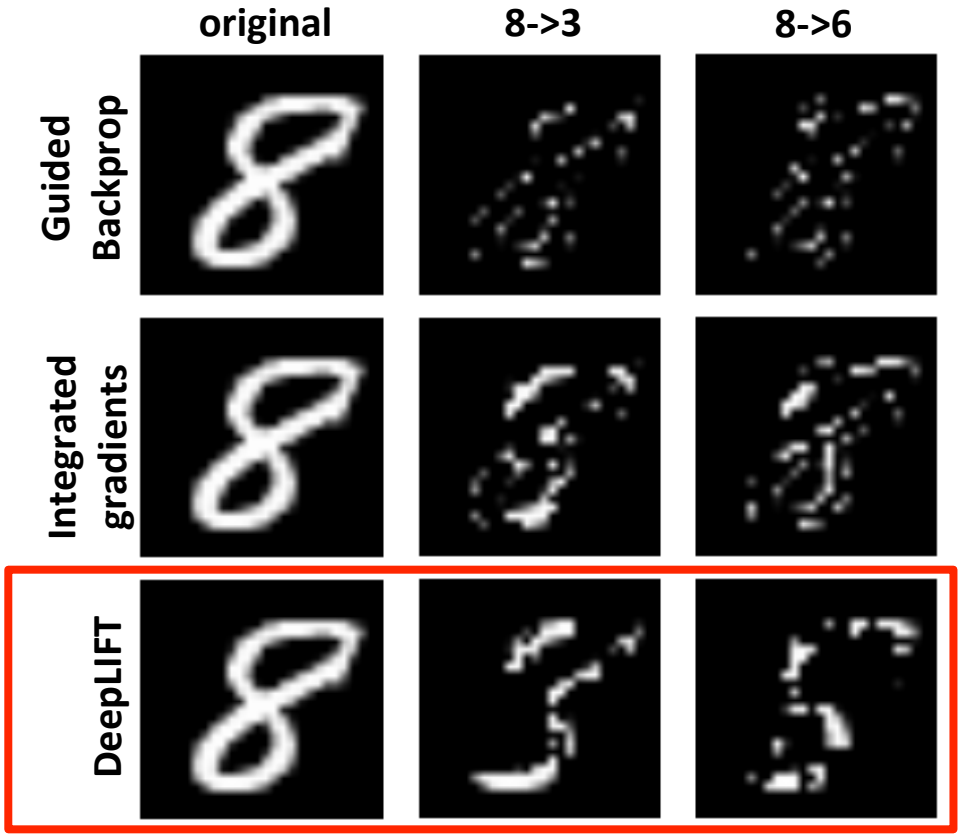
CIFAR10 model, class = "ship"



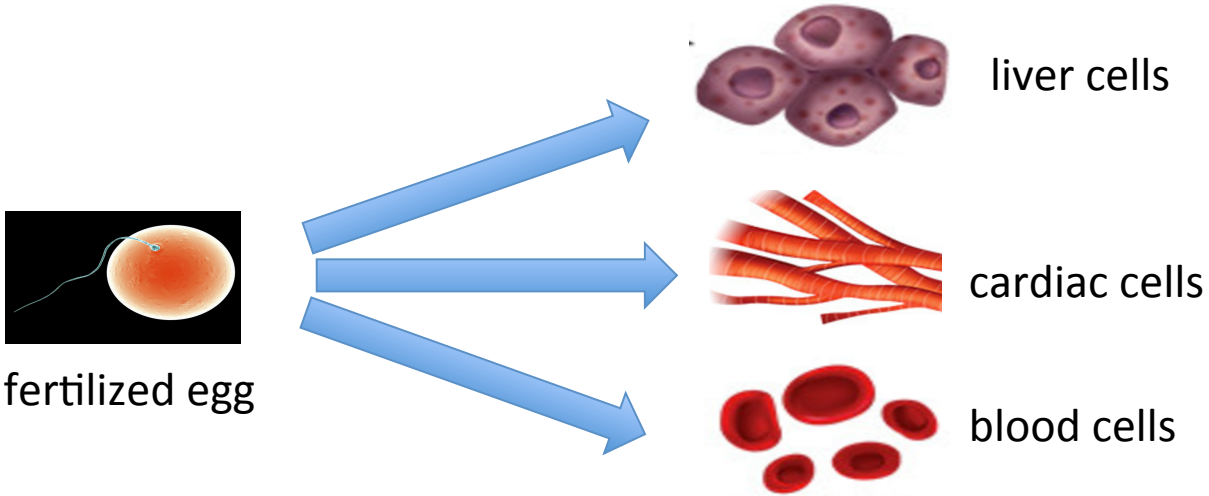
Suggestions on how to pick a reference:

- MNIST: all zeros (background)
- **Consider using a distribution of references**
  - E.g. multiple references generated by shuffling a genomic sequence

Eg: morphing 8 to a 3 or a 6

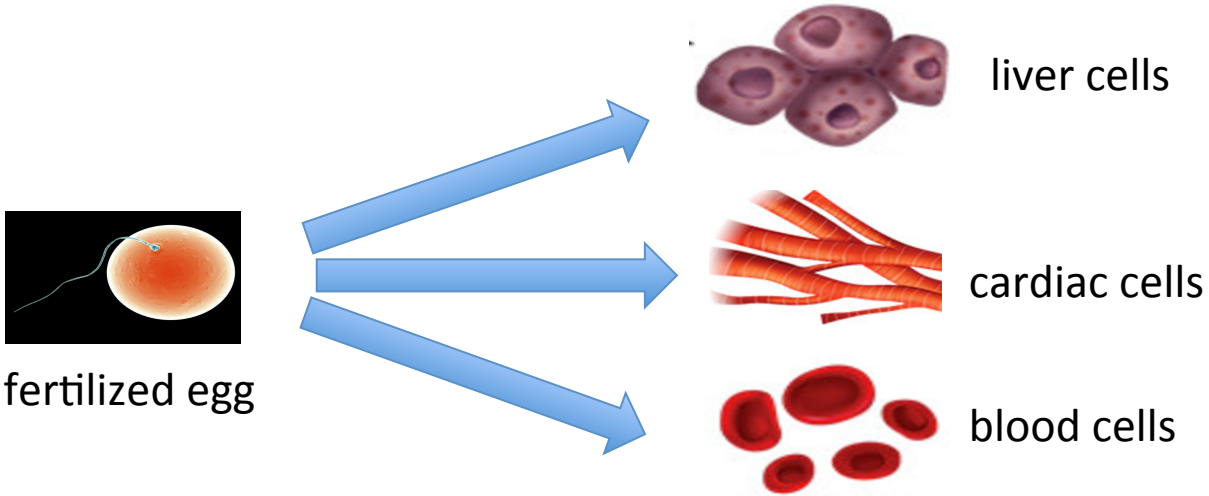


Example biological problem: understanding stem cell differentiation



Cell-types are different because different genes are turned on

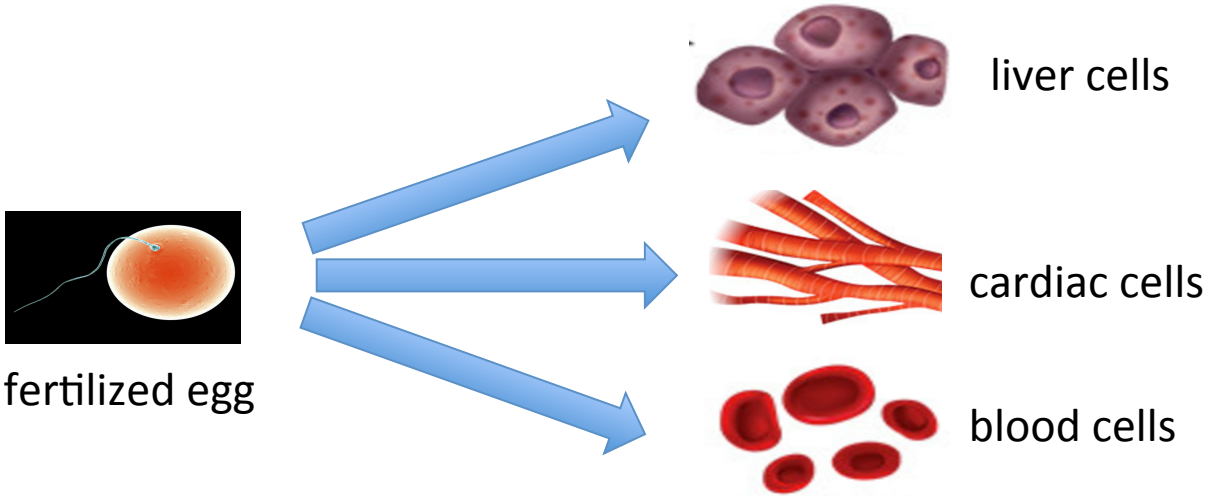
Example biological problem: understanding stem cell differentiation



Cell-types are different because different genes are turned on

How is cell-type-specific gene expression controlled?

Example biological problem: understanding stem cell differentiation

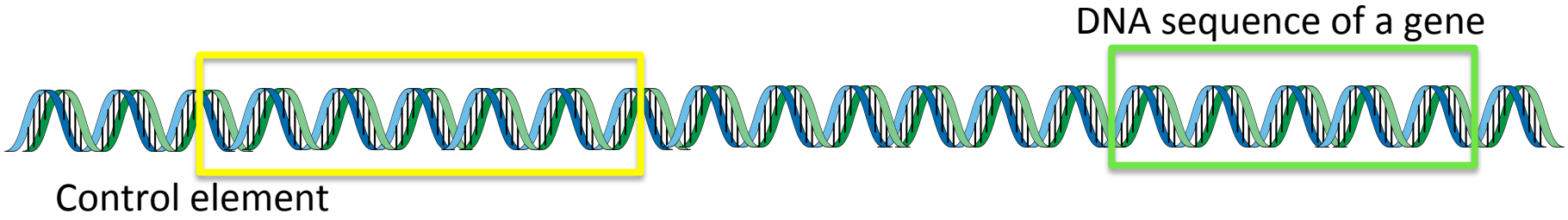


Cell-types are different because different genes are turned on

How is cell-type-specific gene expression controlled?

Ans: “control elements” act like switches to turn genes on

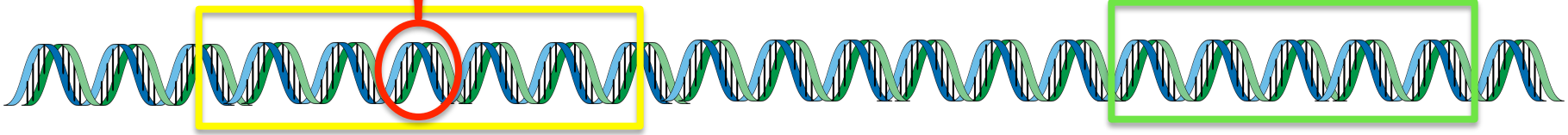
“Control Elements” are switches that turn genes



“Control Elements” are switches that turn genes

Sequence contain “DNA words”
that controller proteins bind to

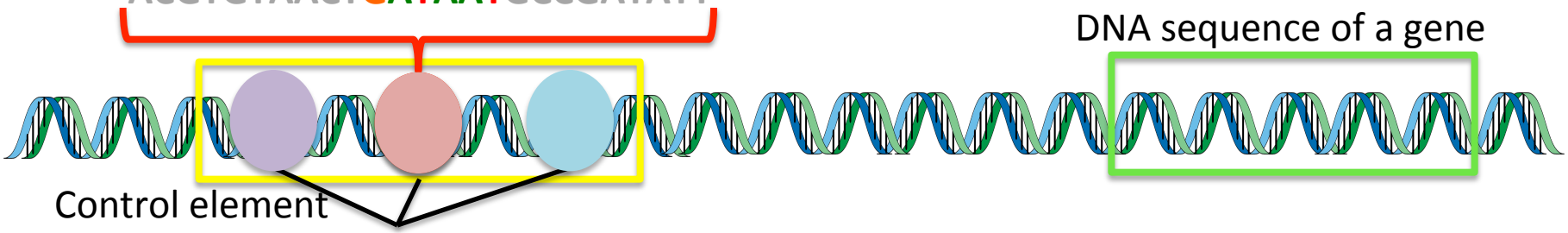
ACGTGTAACTGATAATGCCGATATT



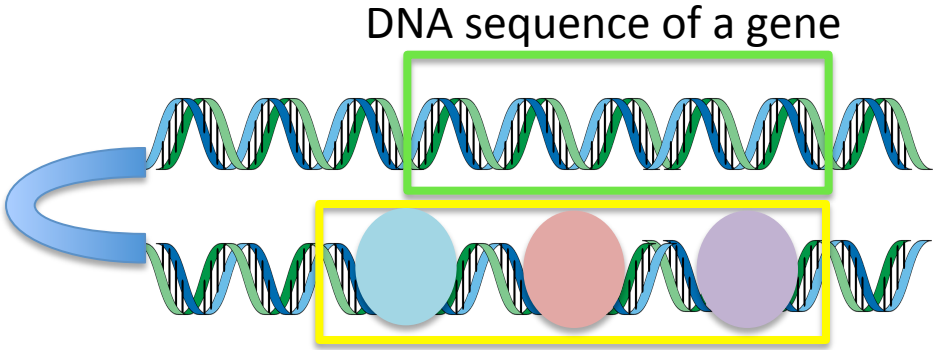
“Control Elements” are switches that turn genes

Sequence contain “DNA words”
that controller proteins bind to

ACGTGTAACTGATAATGCCGATATT

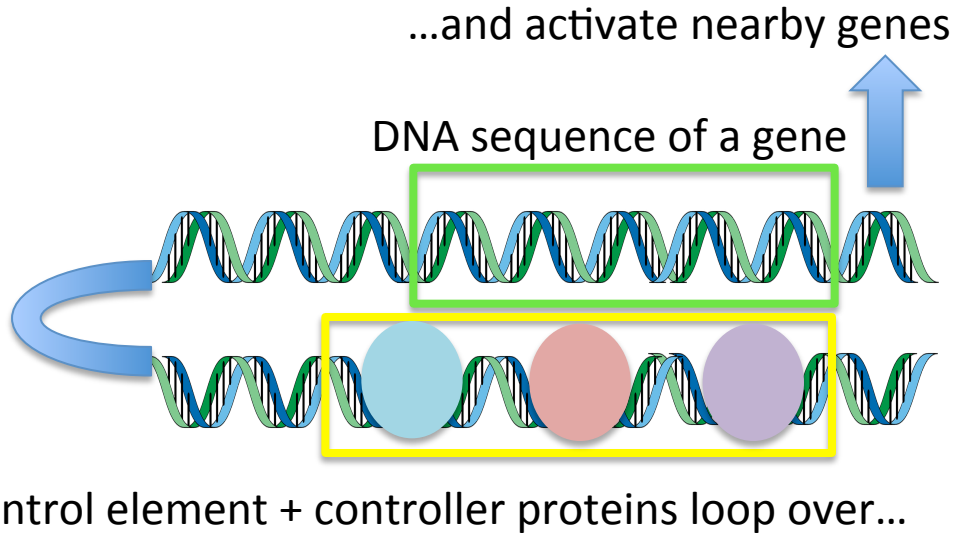


“Control Elements” are switches that turn genes

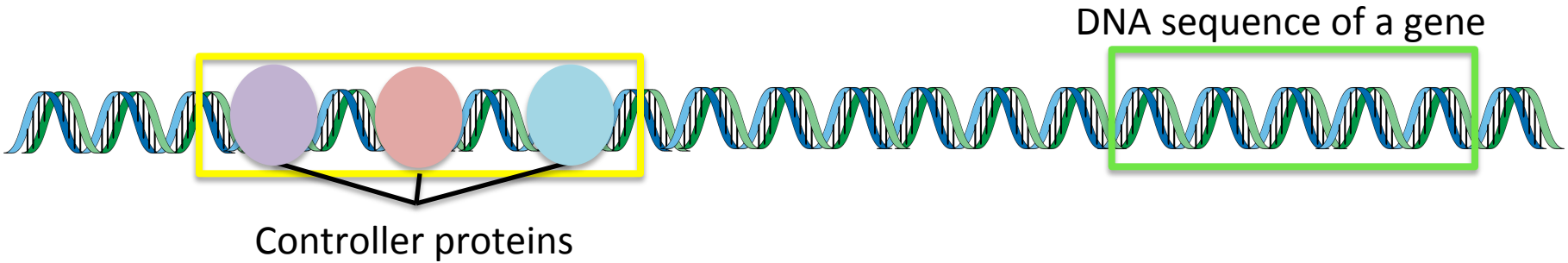


Control element + controller proteins loop over…

“Control Elements” are switches that turn genes



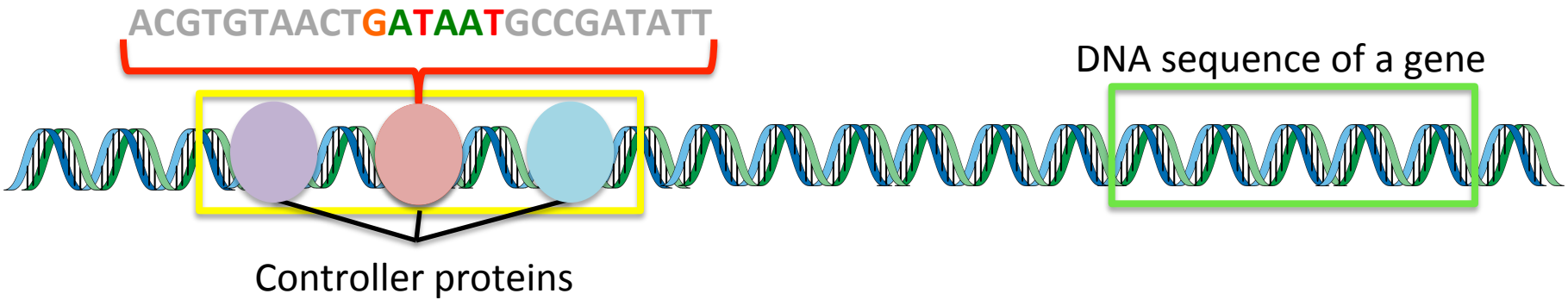
89%* of disease-associated mutations are outside genes!



*Stranger *et al*., *Genet.*, 2011

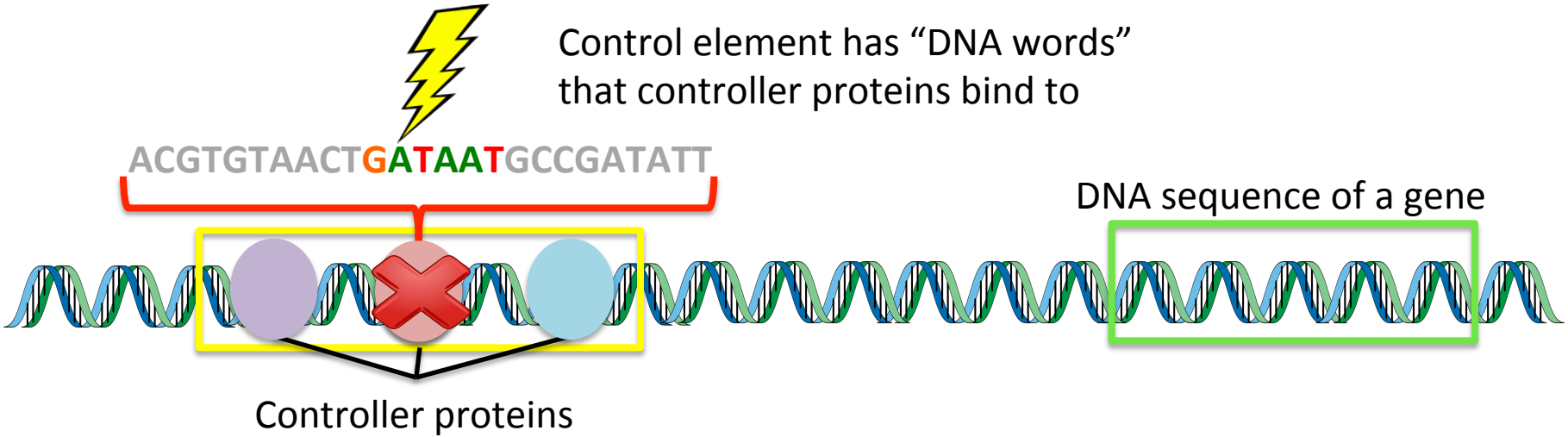
89%* of disease-associated mutations are outside genes!

Control element has “DNA words”
that controller proteins bind to



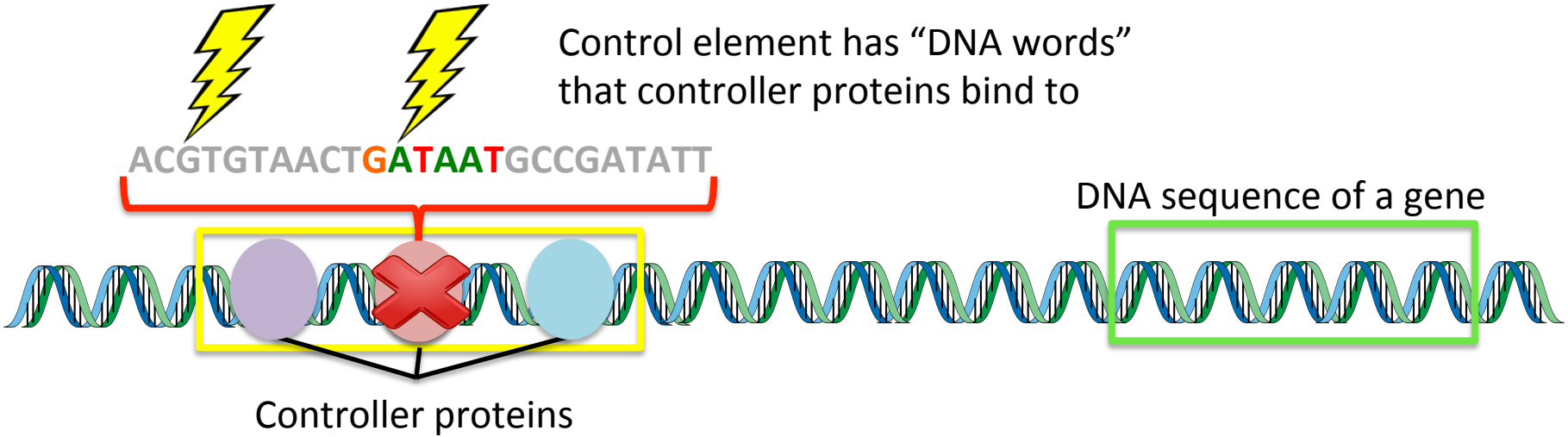
*Stranger *et al*., *Genet.*, 2011

89%* of disease-associated mutations are outside genes!



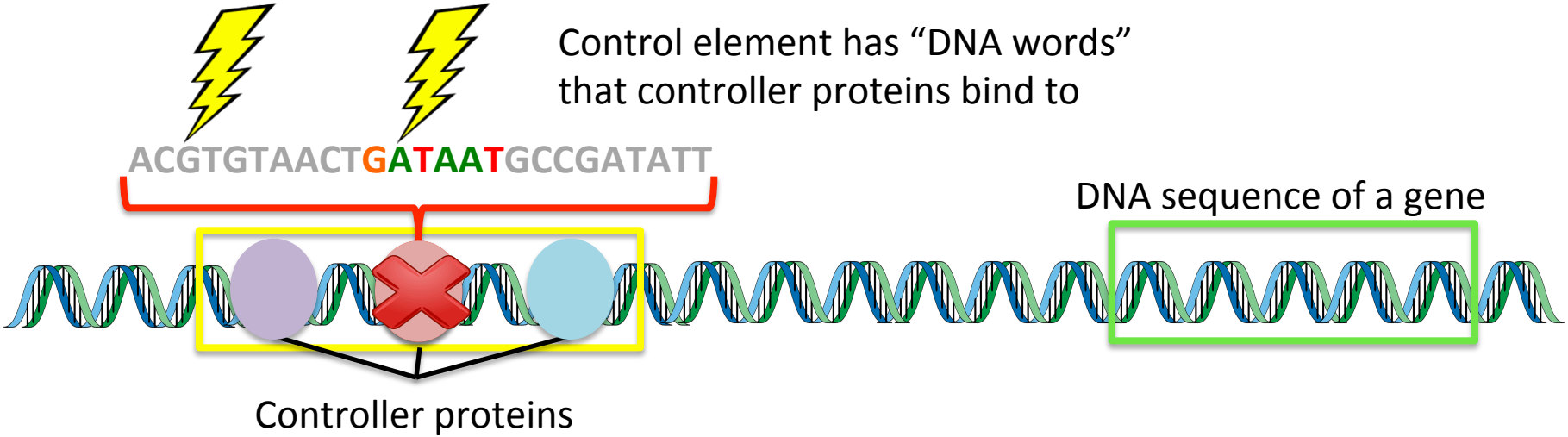
*Stranger *et al*., *Genet.*, 2011

89%* of disease-associated mutations are outside genes!



*Stranger *et al*., *Genet.*, 2011

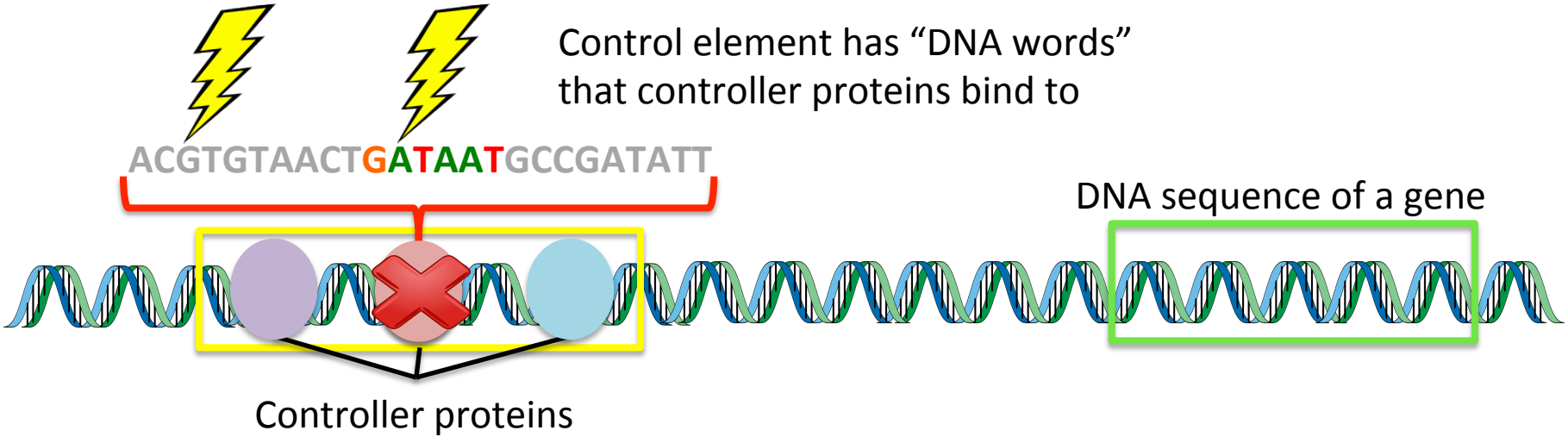
89%* of disease-associated mutations are outside genes!



Many positions in a control element are not essential for its function!

*Stranger *et al*., *Genet.*, 2011

89%* of disease-associated mutations are outside genes!



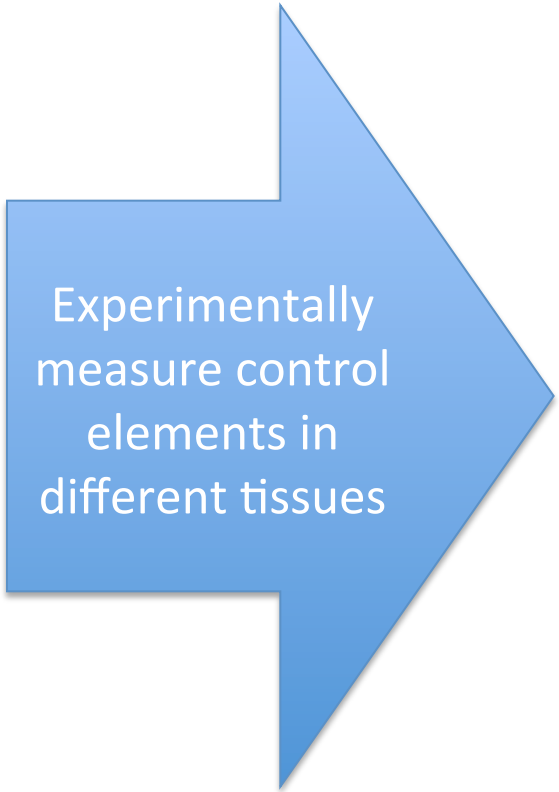
Many positions in a control element are not essential for its function!

→ Which positions in controller elements matter?

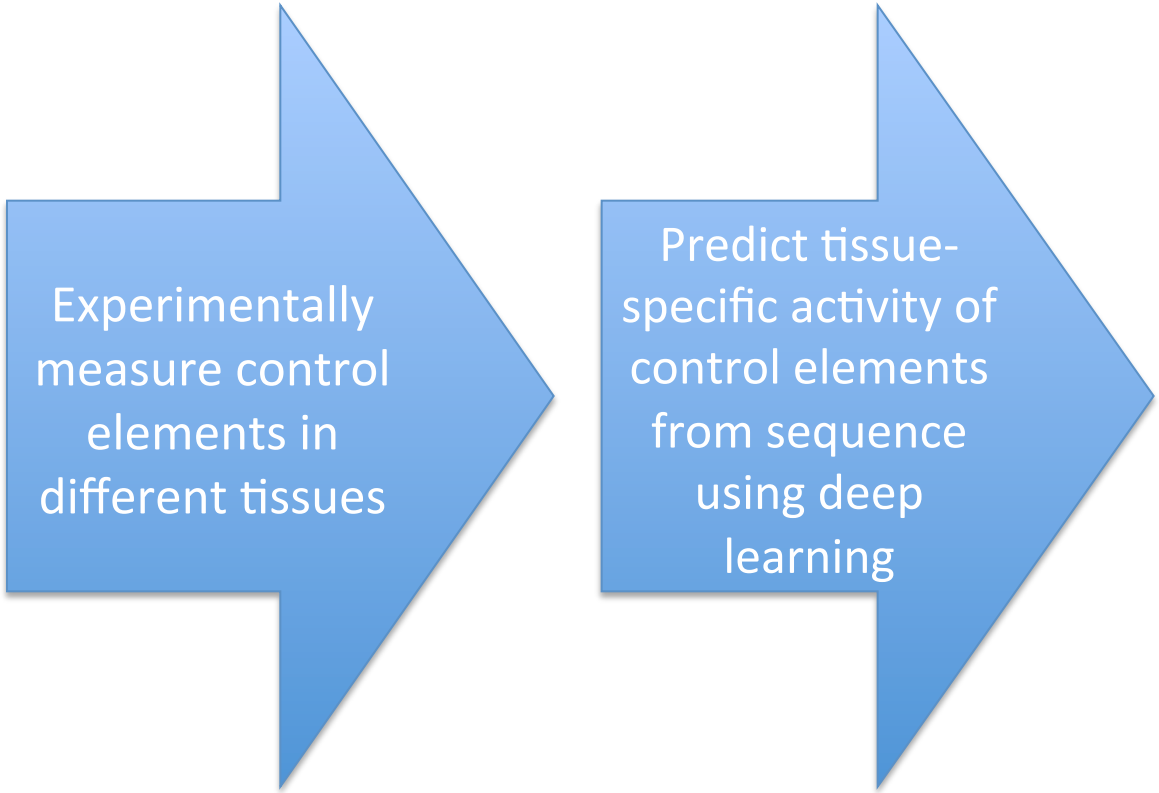
*Stranger *et al*., *Genet.*, 2011

Q: Which positions in control elements matter?

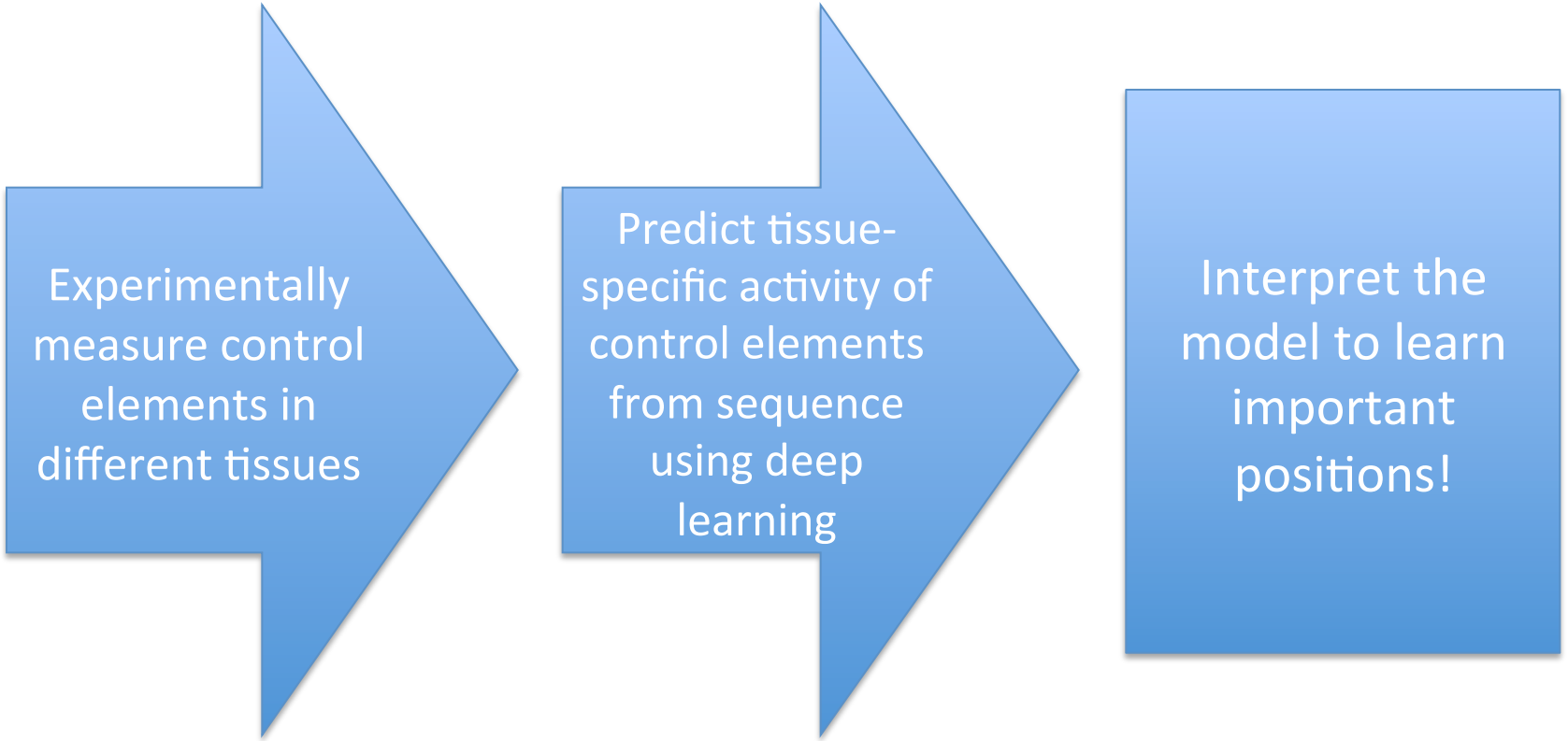
Q: Which positions in control elements matter?



Q: Which positions in control elements matter?

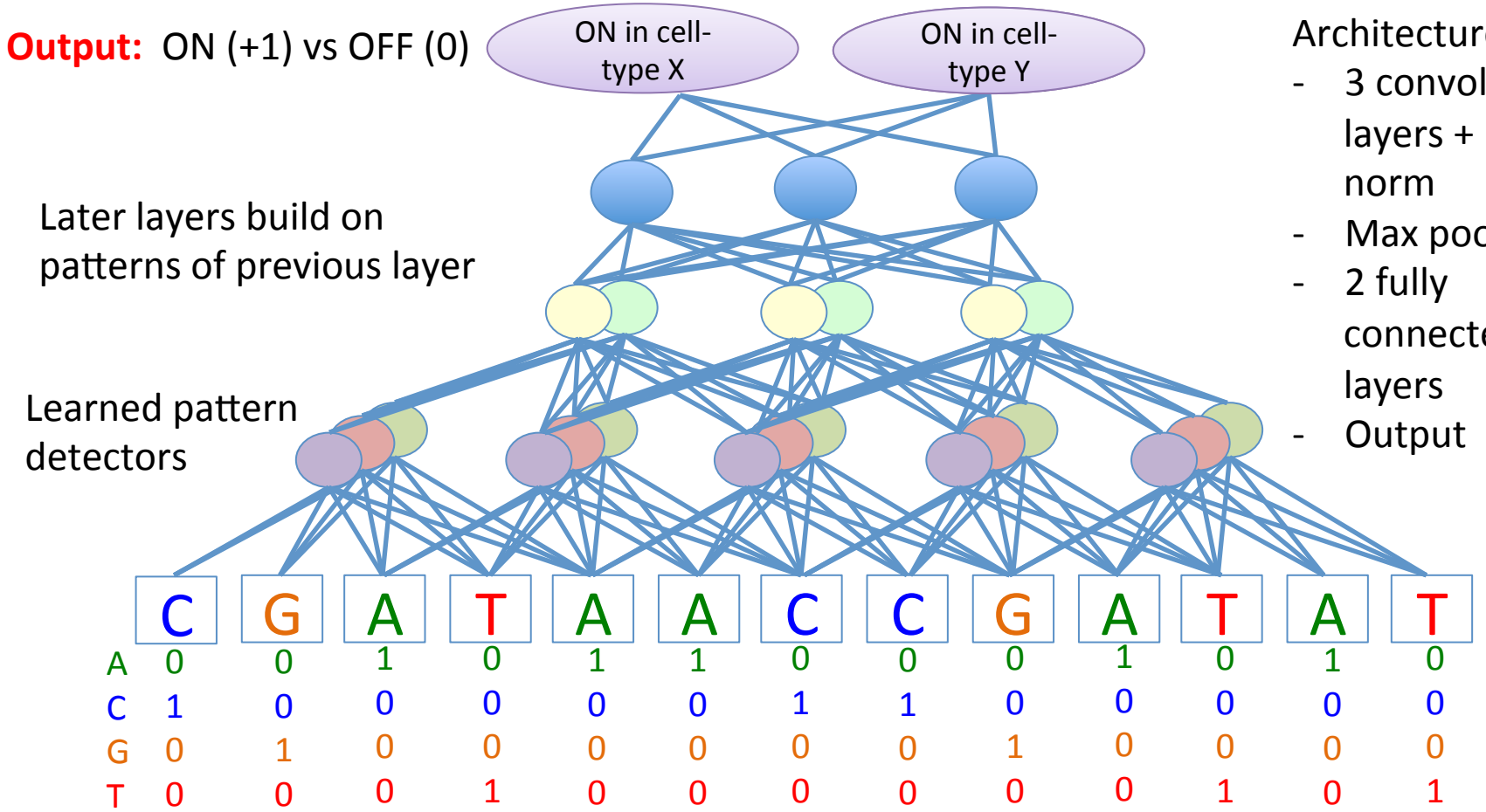


Q: Which positions in control elements matter?



Overview of deep learning model

Output: ON (+1) vs OFF (0)



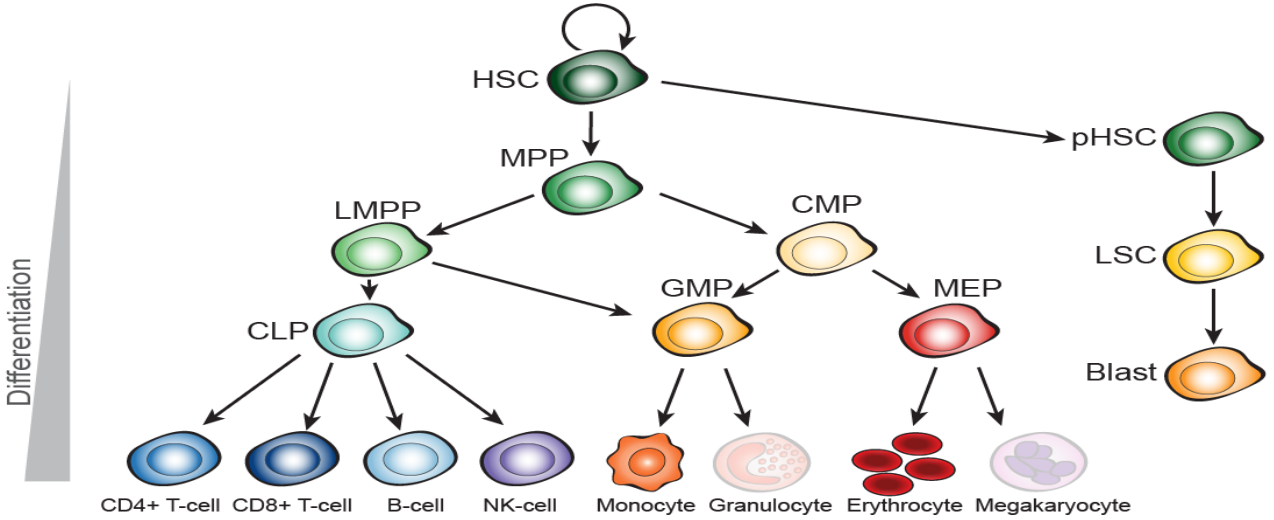
Architecture:

- 3 convolutional layers + batch norm
- Max pooling
- 2 fully connected layers
- Output

Input: DNA sequence represented as ones and zeros

Case study: understanding "control elements" of blood cell types

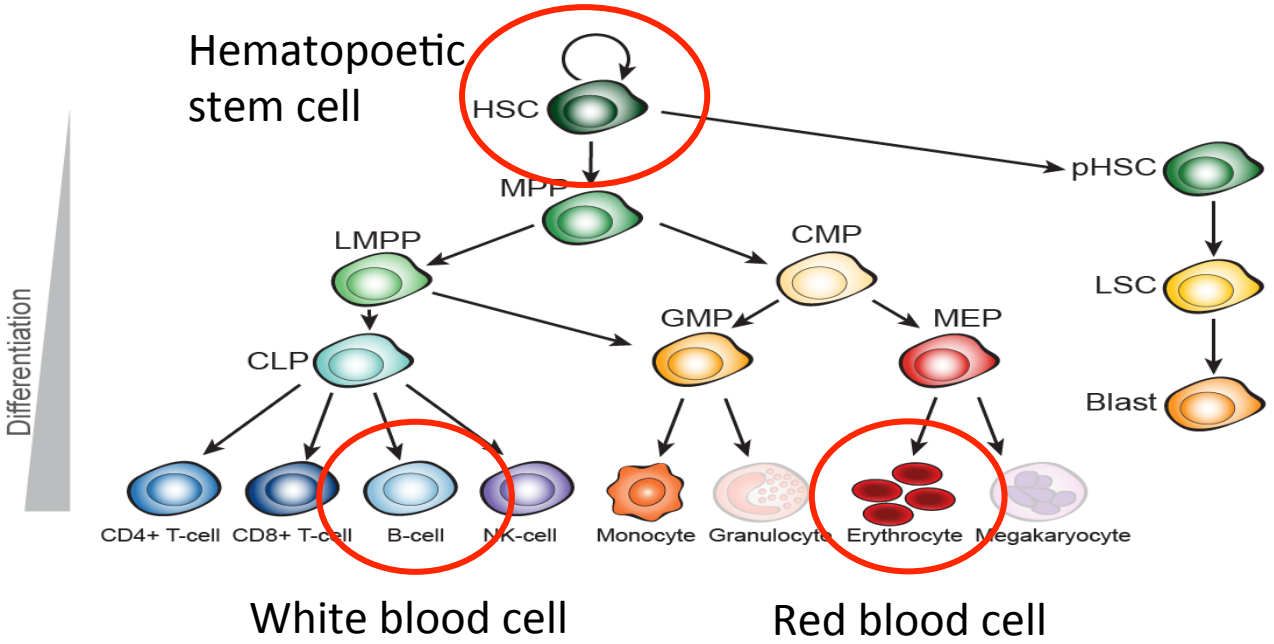
Publicly available dataset profiling control element activity (Corces & Buenrostro et al., 2016)



Peyton Greenside

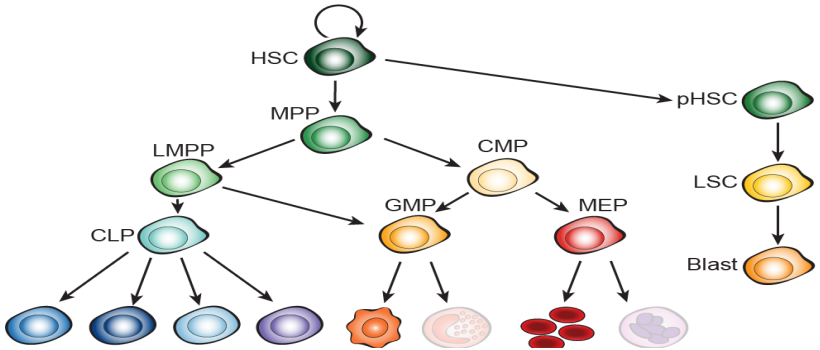
Case study: understanding "control elements" of blood cell types

Publicly available dataset profiling control element activity (Corces & Buenrostro et al., 2016)



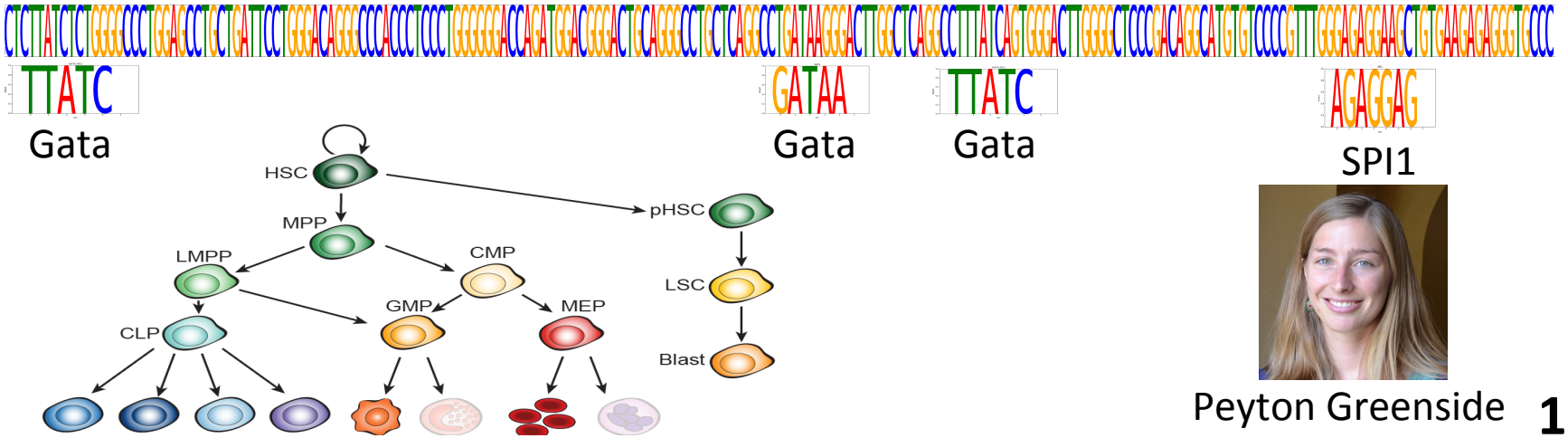
Peyton Greenside

Cell-type-specific use of “controller” sequence in HSC, B-cells and Erythroid



Peyton Greenside

Cell-type-specific use of “controller” sequence in HSC, B-cells and Erythroid



Peyton Greenside

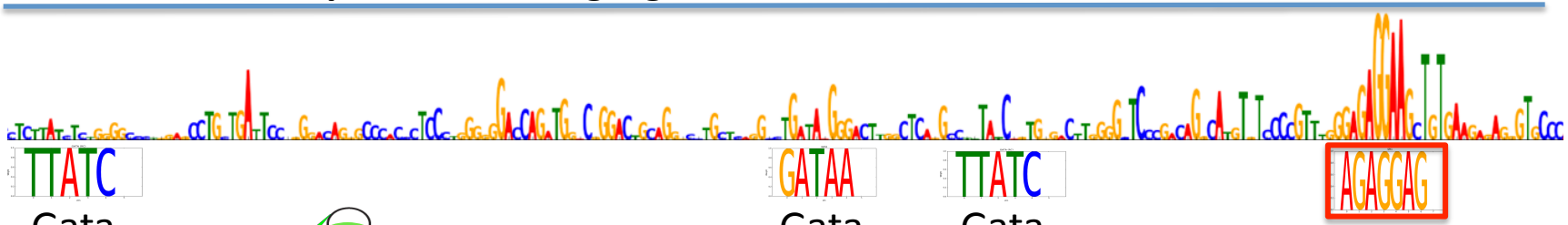
Cell-type-specific use of “controller” sequence in HSC, B-cells and Erythroid

Importance in **HSC’s**

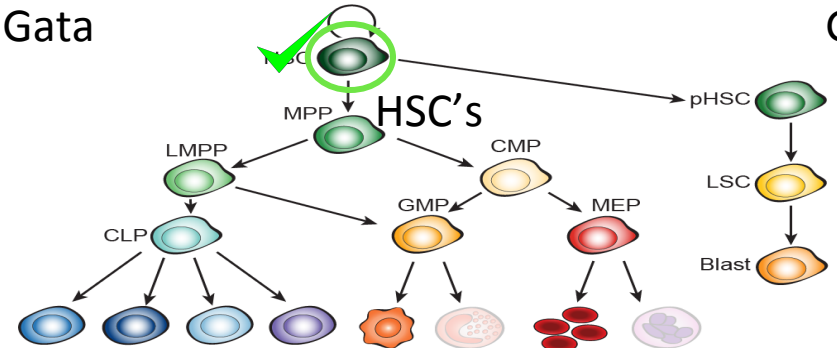
“Is an active control element” signal

SPI1 controller protein binding signal (Data unavailable)

GATA1 controller protein binding signal (Data unavailable)



Gata Gata Gata SPI1



Peyton Greenside

Cell-type-specific use of “controller” sequence in HSC, B-cells and Erythroid

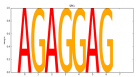
Importance in **B-cells**

“Is an active control element” signal	No peak
SPI1 controller protein binding signal	No peak
GATA1 controller protein binding signal	No peak

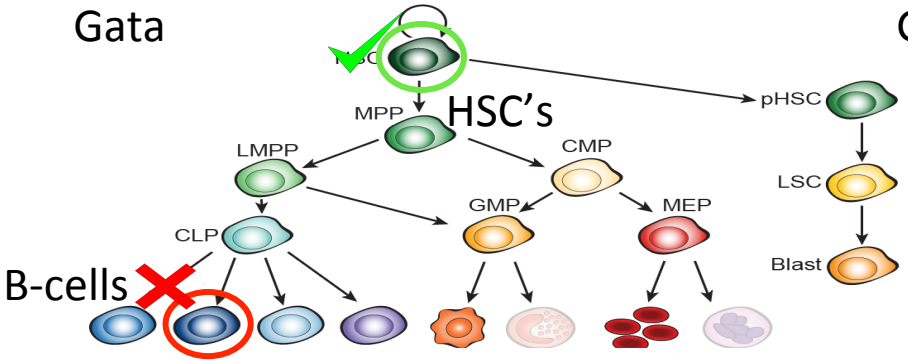
Gata

Gata

Gata



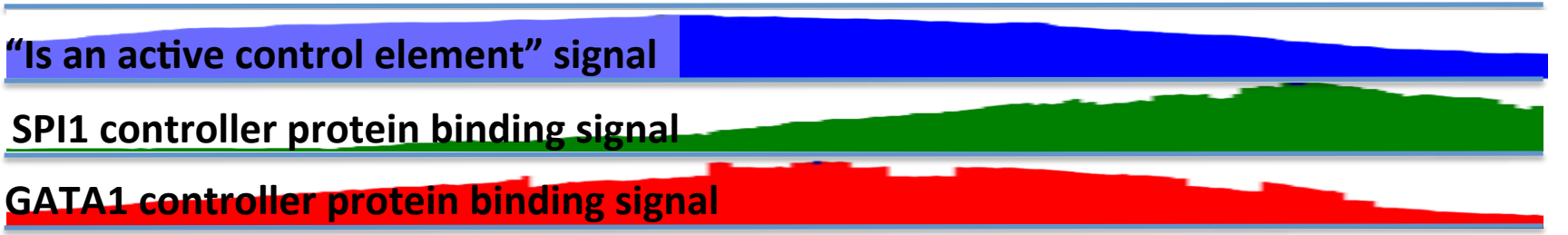
SPI1

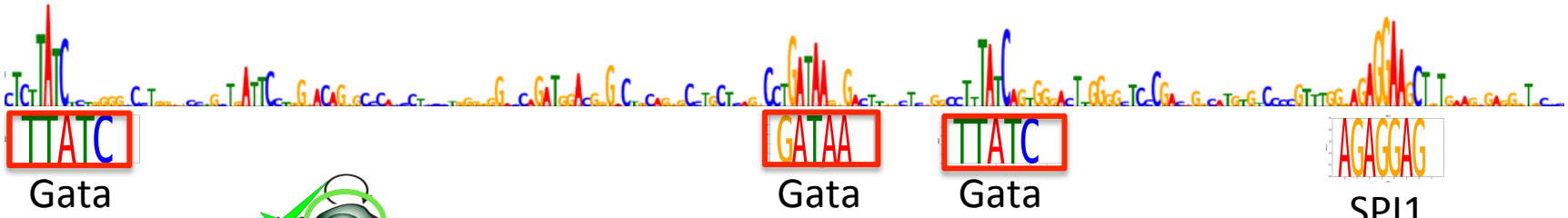


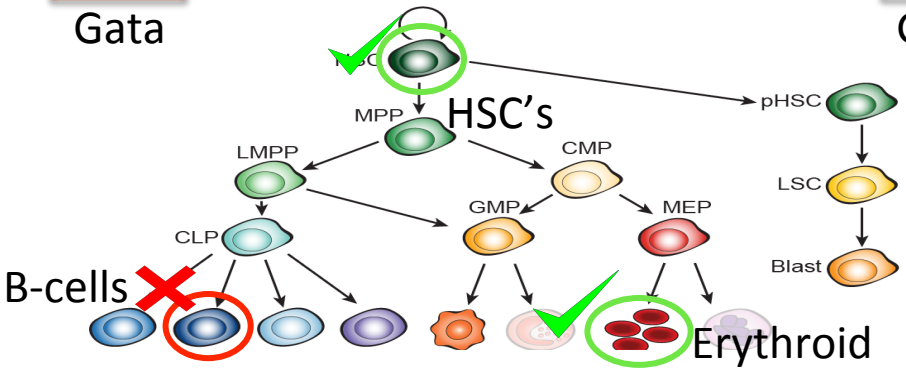
Peyton Greenside

Cell-type-specific use of “controller” sequence in HSC, B-cells and Erythroid

Importance in Erythroid

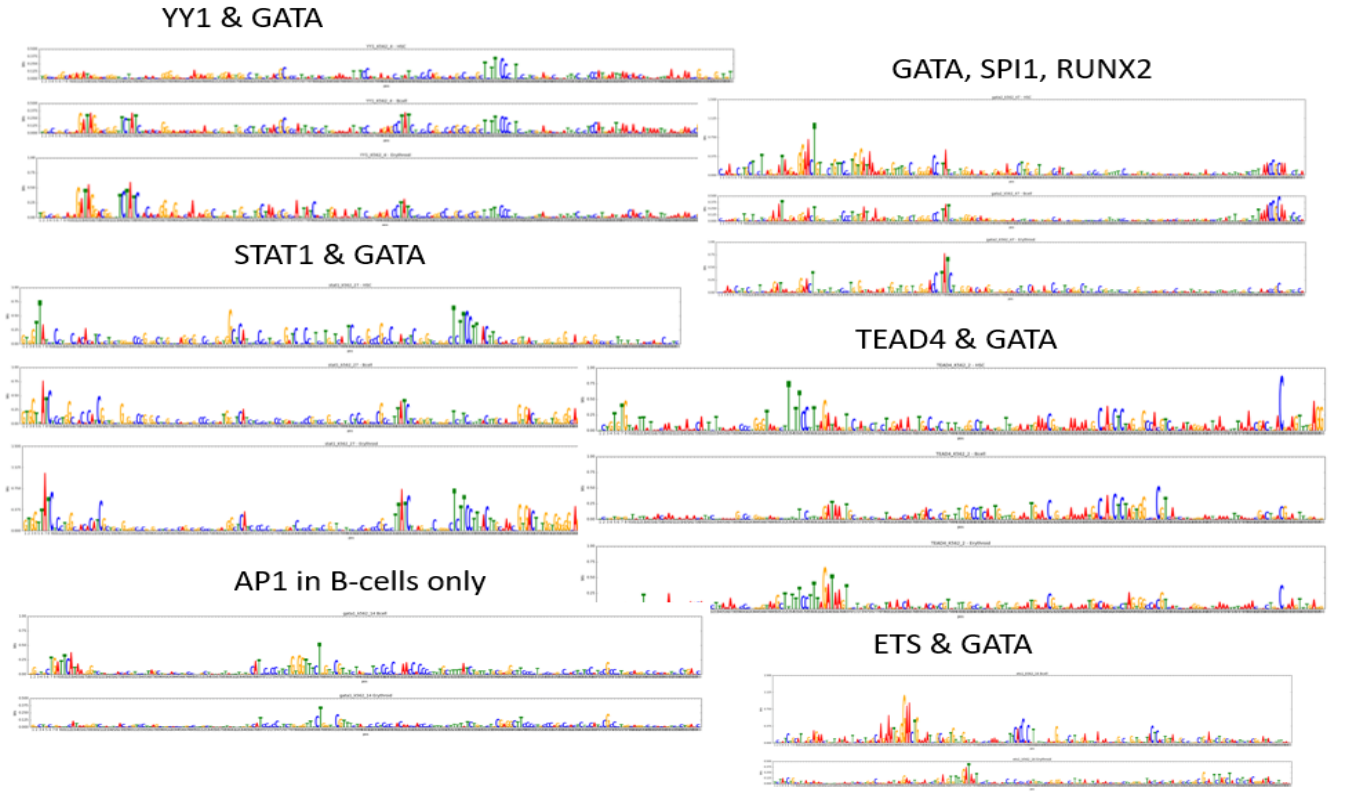






Peyton Greenside

Can study the regulatory code in millions of control elements!



Peyton Greenside

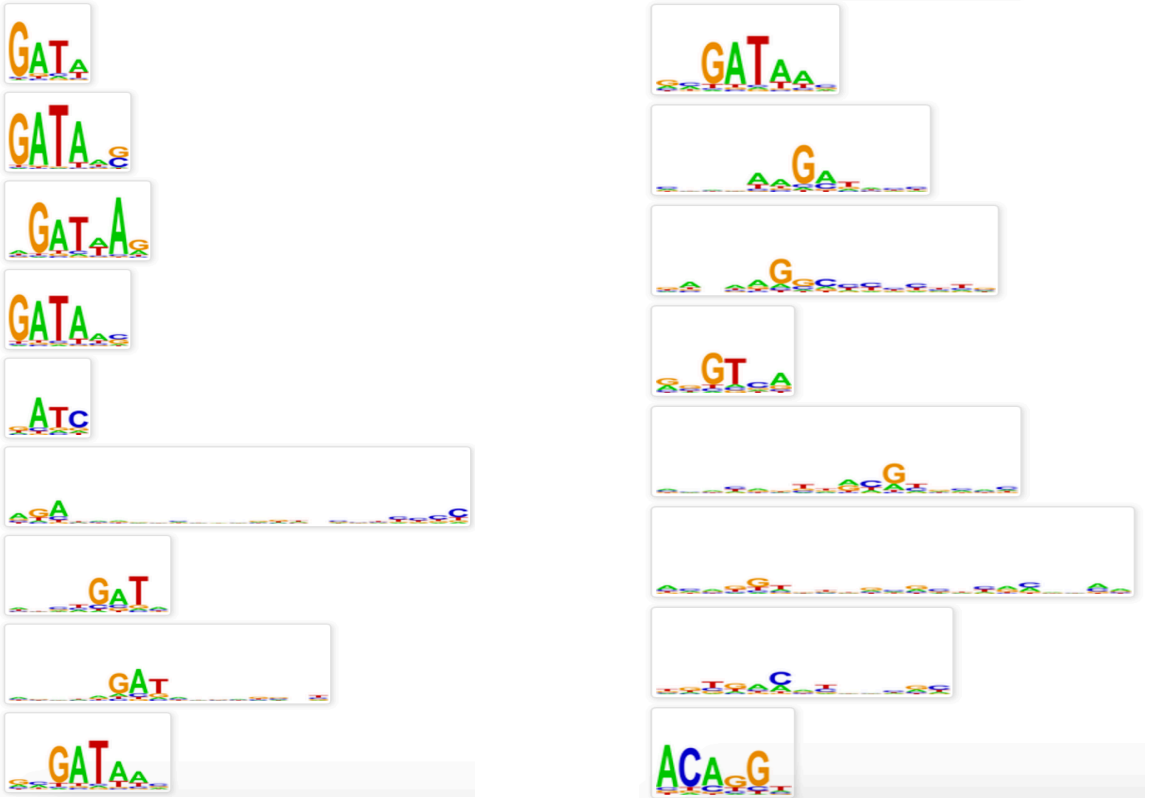
Questions for the model

- Which parts of the input are the most important for making a given prediction?
- What are the recurring patterns in the input?

Question in biology: What are the DNA "words" ("motifs") driving controller protein binding?

Naïve idea: look at individual pattern detectors

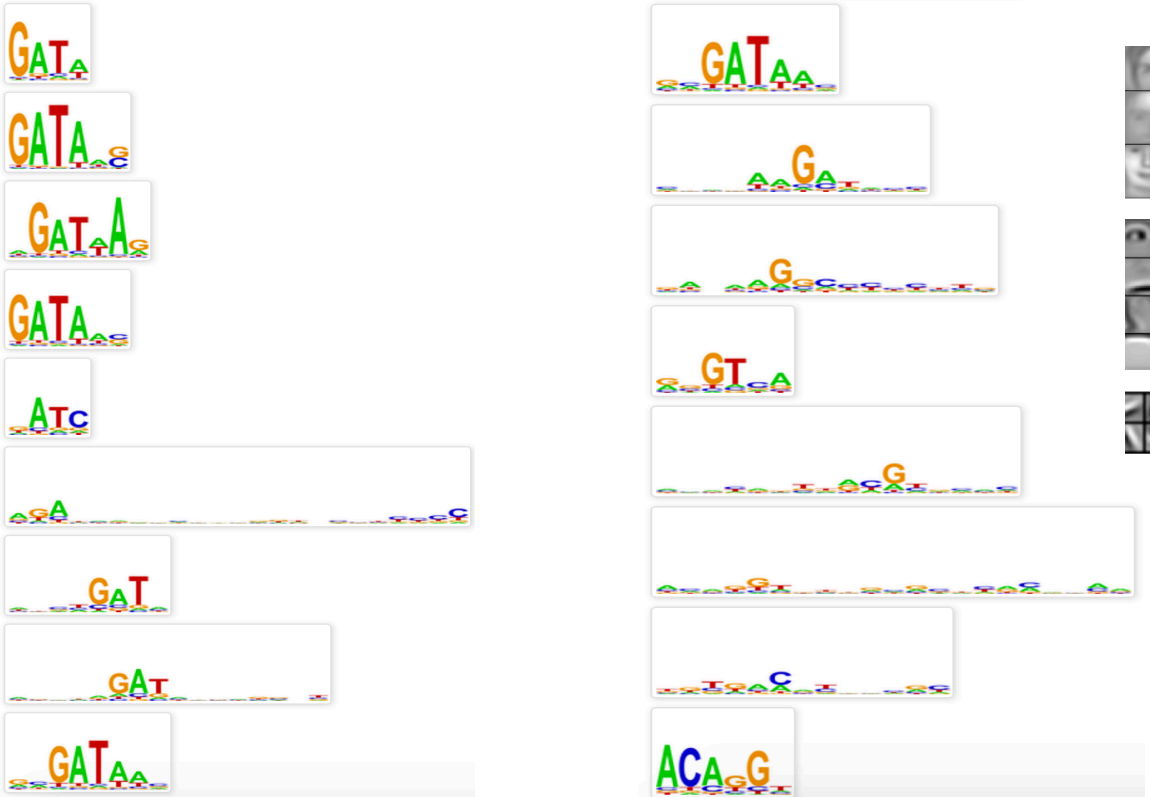
Individual GATA pattern detectors motifs found by DeepBind (Alipanahi et al.)

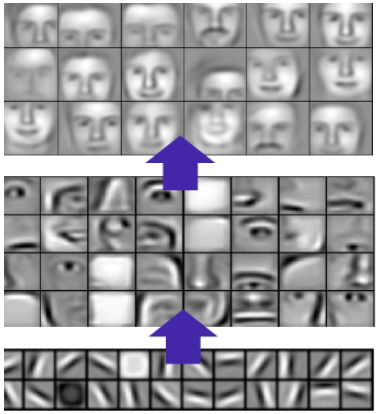


Problem: **High levels of redundancy, because multiple neurons cooperate with each other**

Naïve idea: look at individual pattern detectors

Individual GATA pattern detectors motifs found by DeepBind (Alipanahi et al.)





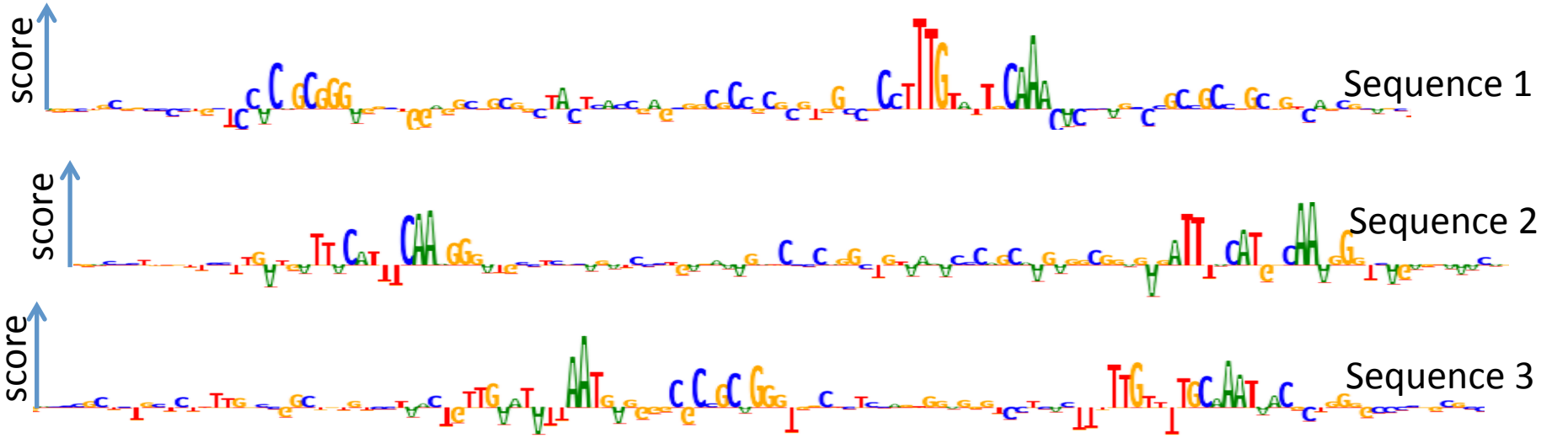
Computer vision

Problem: **High levels of redundancy, because multiple neurons cooperate with each other**

How do we combine the contributions of multiple pattern detectors to find consolidated patterns?

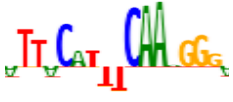
How do we combine the contributions of multiple pattern detectors to find consolidated patterns?

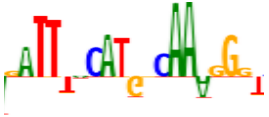
Insight: input-level importance scores reveal combined contributions



How do we combine the contributions of multiple pattern detectors to find consolidated patterns?

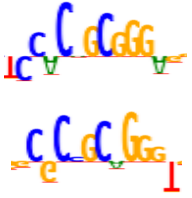
Insight: input-level importance scores reveal combined contributions

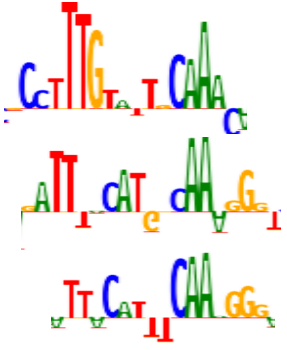




How do we combine the contributions of multiple pattern detectors to find consolidated patterns?

Insight: input-level importance scores reveal combined contributions





How do we combine the contributions of multiple pattern detectors to find consolidated patterns?

Insight: input-level importance scores reveal combined contributions

How do we combine the contributions of multiple pattern detectors to find consolidated patterns?

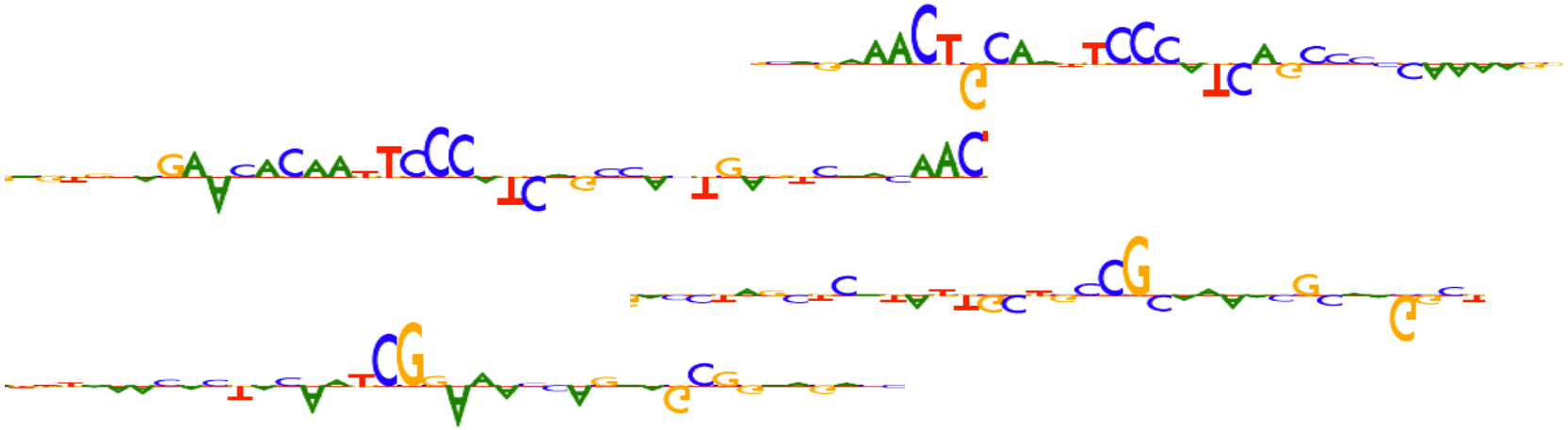
Insight: input-level importance scores reveal combined contributions

TF-MoDISco: TF **Mo**tif **D**iscovery from **I**mportance **Sco**res

TF-MoDISco: More details on the clustering

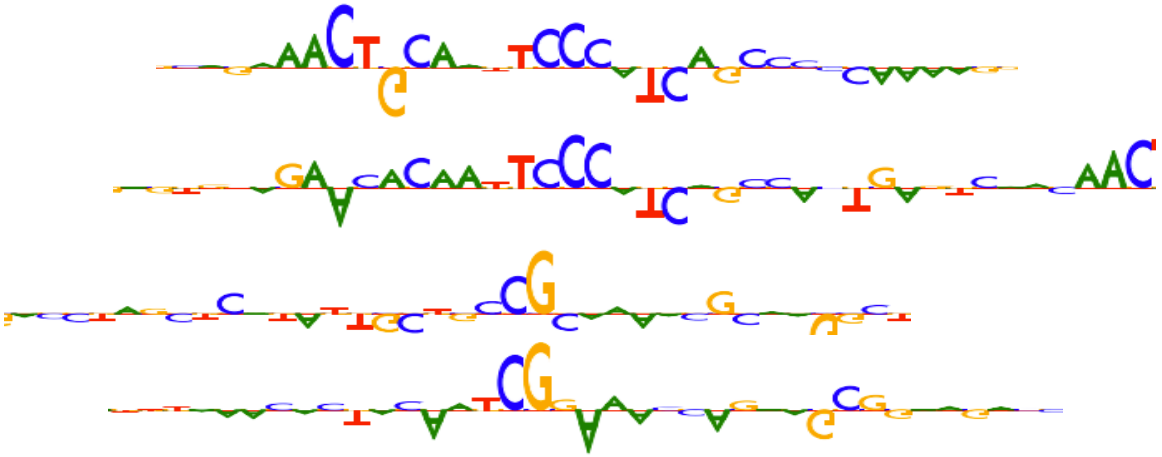
TF-MoDISco: More details on the clustering

(1) Compute affinities between pairs of important segments using a cross-correlation-like metric



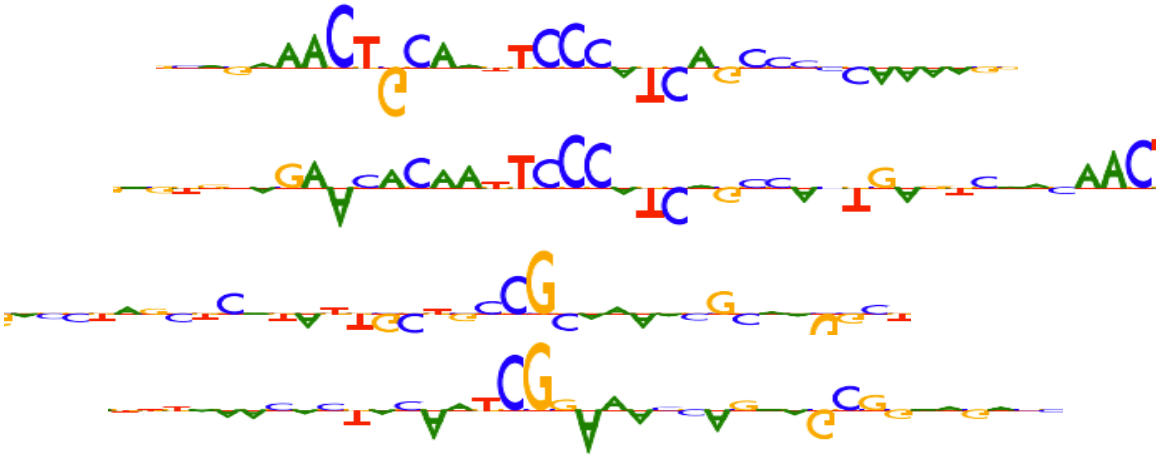
TF-MoDISco: More details on the clustering

(1) Compute affinities between pairs of important segments using a cross-correlation-like metric



TF-MoDISco: More details on the clustering

(1) Compute affinities between pairs of important segments using a cross-correlation-like metric



(2) Cluster affinity matrix with community detection

Key idea: Density-Adaptive Distances (1)

- Problem: notion of “far away” varies with the cluster
  - Weak patterns clusters: instances of pattern may be farther away on average
  - Notion of “far” needs to take this into account

Key idea: Density-Adaptive Distance (2)

- Soln: Adapt notion of distance to the local density of the data!

Key idea: Density-Adaptive Distance (2)

- Soln: Adapt notion of distance to the local density of the data!
  - First step of t-sne: compute conditional probs

$$p_{j|i} \propto \exp(-\beta_i \times \text{distance}(i, j))$$

  - β_i is tuned to attain a desired perplexity!

Key idea: Density-Adaptive Distance (2)

- Soln: Adapt notion of distance to the local density of the data!
  - First step of t-sne: compute conditional probs

$$p_{j|i} \propto \exp(-\beta_i \times \text{distance}(i, j))$$

  - β_i is tuned to attain a desired perplexity!
    - Larger β_i will be used in denser region of the space

Key idea: Density-Adaptive Distance (2)

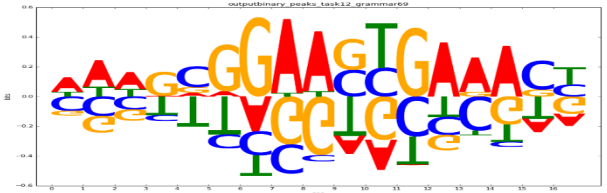
- Soln: Adapt notion of distance to the local density of the data!
  - First step of t-sne: compute conditional probs

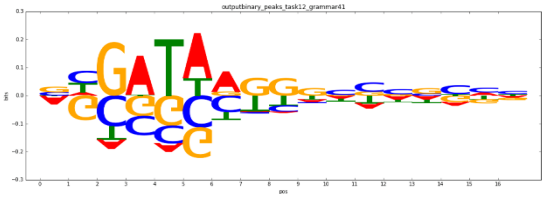
$$p_{j|i} \propto \exp(-\beta_i \times \text{distance}(i, j))$$

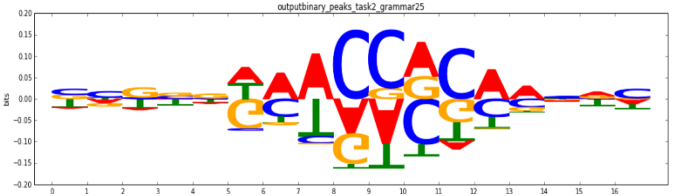
  - β_i is tuned to attain a desired perplexity!
    - Larger β_i will be used in denser region of the space
  - Use density-adapted probabilities with clustering based on Louvain community detection

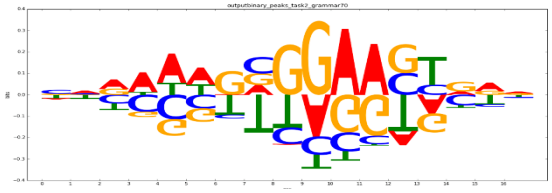
We can learn 1000s of known and novel DNA words defining tissue-specific control elements!

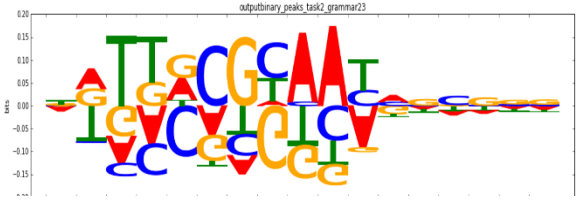
Peyton Greenside

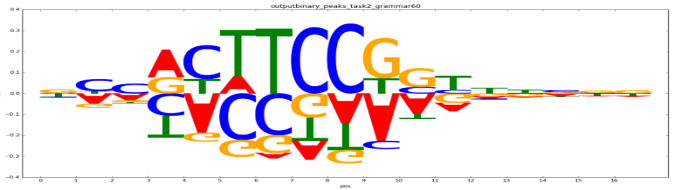


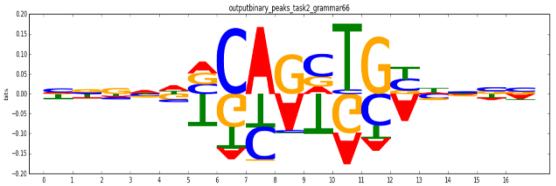


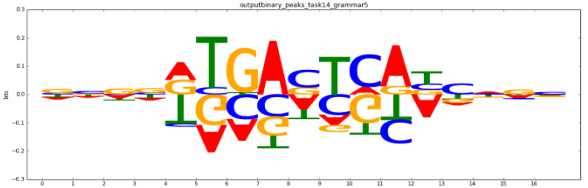


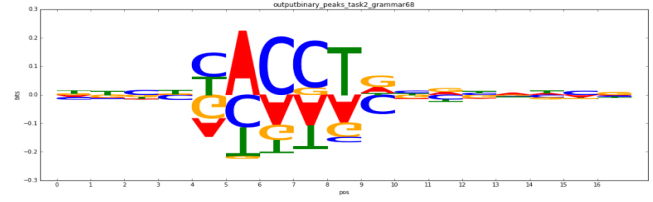


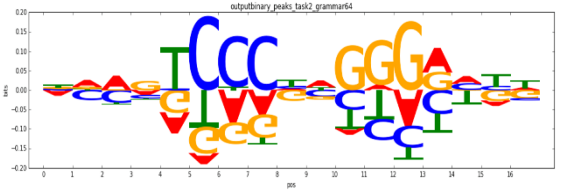


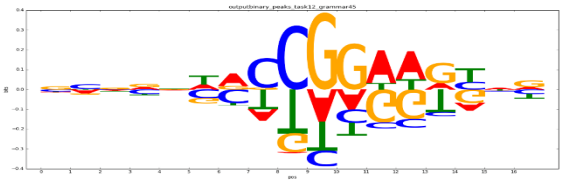












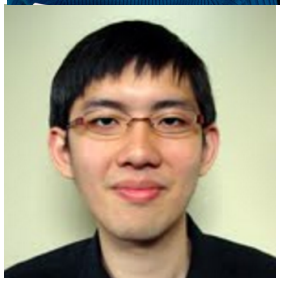
Summary

- DeepLIFT: can efficiently reveal important parts of the input for a given prediction
  - With advantages over other methods
  - https://github.com/kundajelab/deeplift
- TF-MoDISco: Motif Discovery from Importance Scores
  - More details in talk at NIPS comp bio: https://www.youtube.com/watch?v=fXPGVJg956E
- Can be used to understand the regulatory sequence controlling tissue-specific control elements

Thank you to the Kundaje lab!

Anshul Kundaje

Anna Shcherbina



Chuan Sheng Foo

Peyton Greenside

Johnny Israeli

Irene Kaplow

Oana Ursu
Amr Alexandari
Daniel Kim
Michael Wainberg
Maryna Taranova
Chris Probert
Jin-Wook Lee

Funding

HHMI International Student Research Fellowship
Bio-X Fellowship
Microsoft Women's Fellowship
NIH R01ES02500902