



Deep learning for genomic discovery

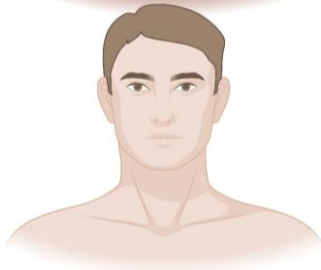
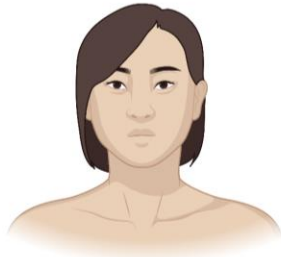
Anshul Kundaje

Twitter: @anshulkundaje

Website: <http://anshul.kundaje.net>

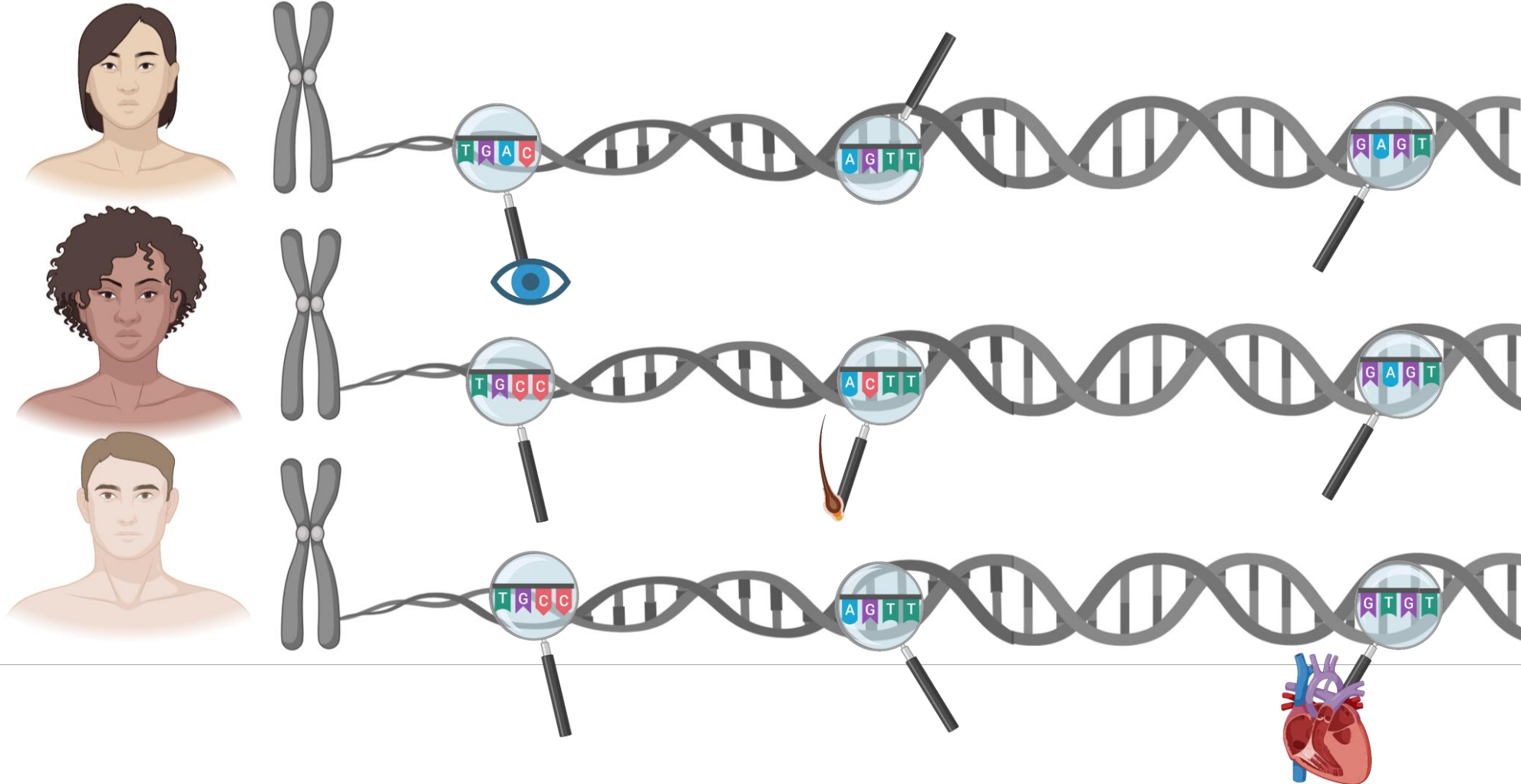
Genetic variants associated with traits and diseases

TGCCAAGCAGCAAAGTTTTGCTGCTGTTTATTTTTGTAGCTCTTACTATAGAGTTTTAA



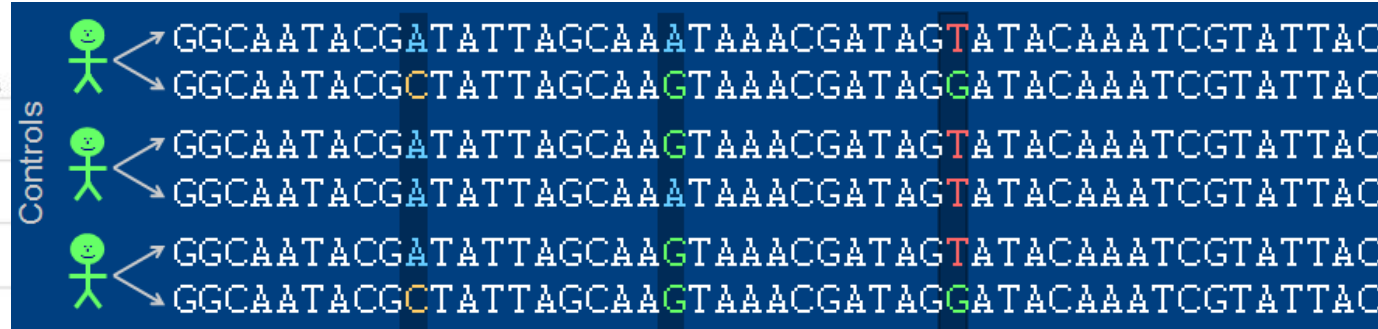
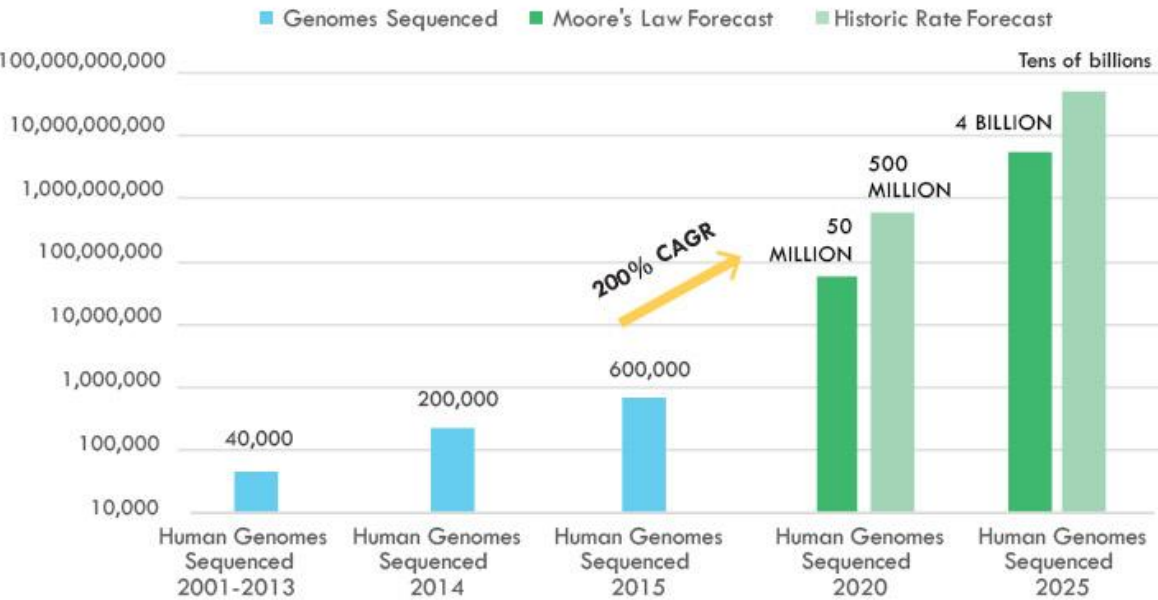
Genetic variants associated with traits and diseases

TGCCAAGCAGCAAAGTTTTGCTGCTGTTTATTTTTGTAGCTCTTACTATAGAGTTTTAA



Population sequencing to identify disease-associated genetic variants

The Number of Human Genomes Sequenced (log scale)



Source: National Human Genome Research Institute (NHGRI), ARK Investment Management LLC



GA II
1.6 billion bp per day
(2008)



GA IIX
5 billion bp per day
(2009)



HiSeq 2500
60 billion bp per day
(2012)



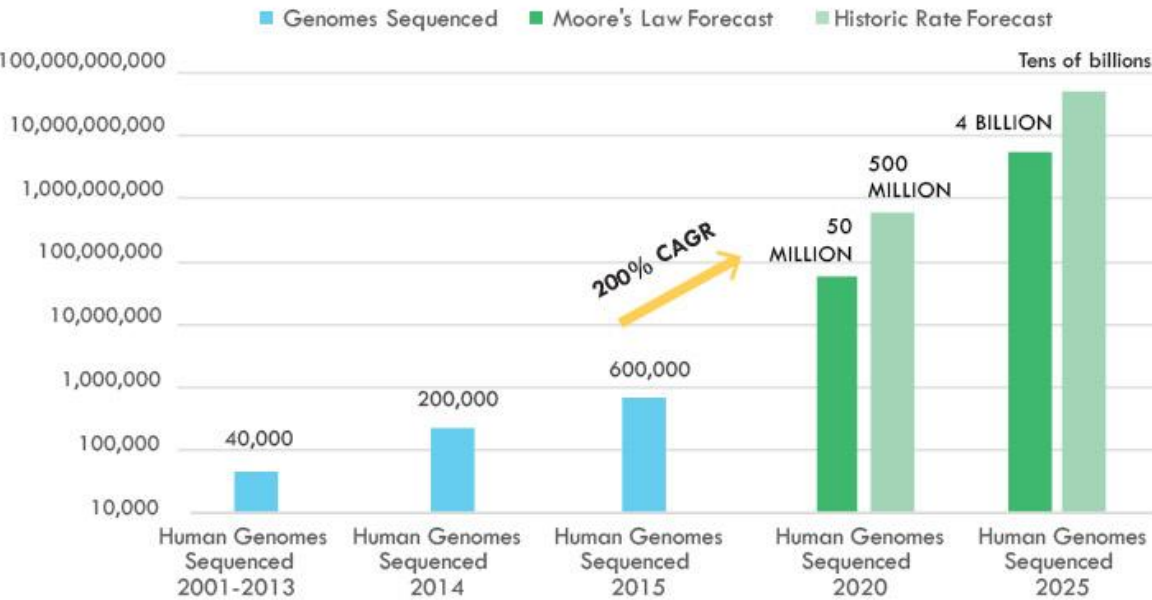
Oxford Nanopore
technology

Images: www.illumina.com/systems
Numbers: www.politgenomics.com/next-generation-sequencing-informatics
Dates: Illumina press releases

Millions of common and rare genetic variants found in human population

Population sequencing to identify disease-associated genetic variants

The Number of Human Genomes Sequenced (log scale)



Source: National Human Genome Research Institute (NHGRI), ARK Investment Management LLC



GA II
1.6 billion bp per day (2008)



GA IIX
5 billion bp per day (2009)

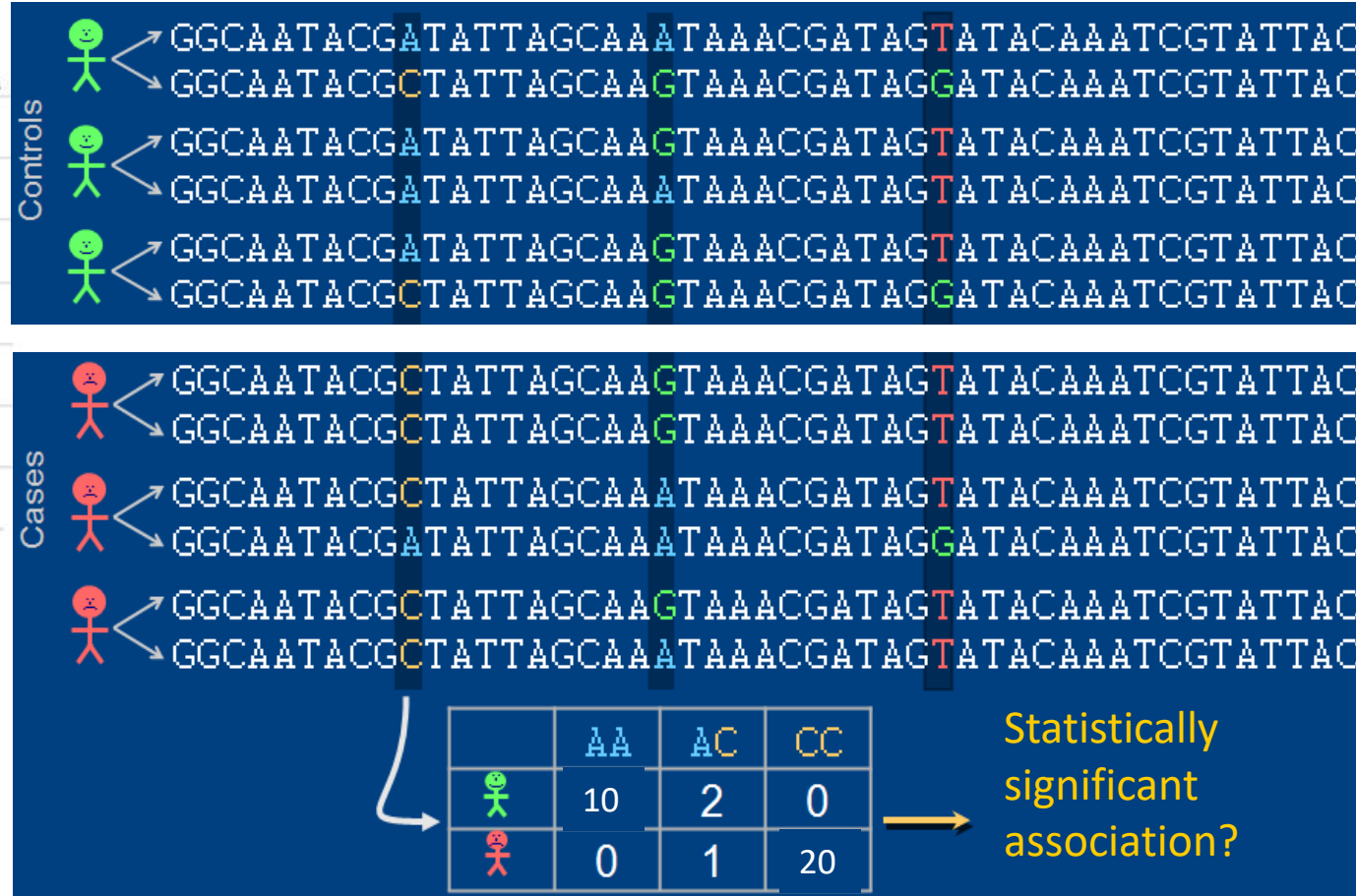


HiSeq 2500
60 billion bp per day (2012)



Oxford Nanopore technology

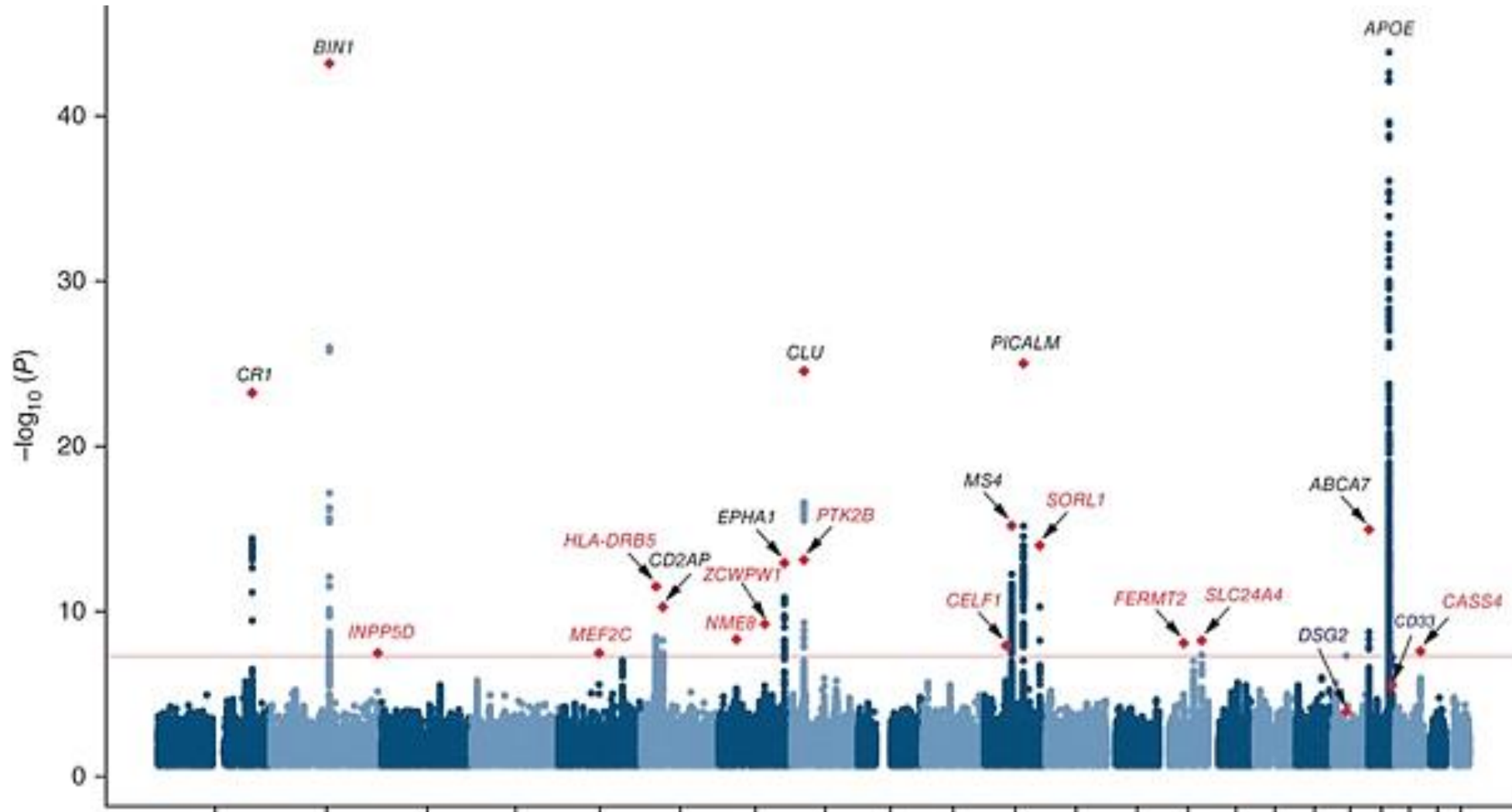
Images: www.illumina.com/systems
Numbers: www.politgenomics.com/next-generation-sequencing-informatics
Dates: Illumina press releases



Millions of common and rare genetic variants found in human population

Genetic variants associated with Alzheimer's disease

Statistical significance of association



(Lambert et al., Nat. Genet., 2013)

Genomic position →

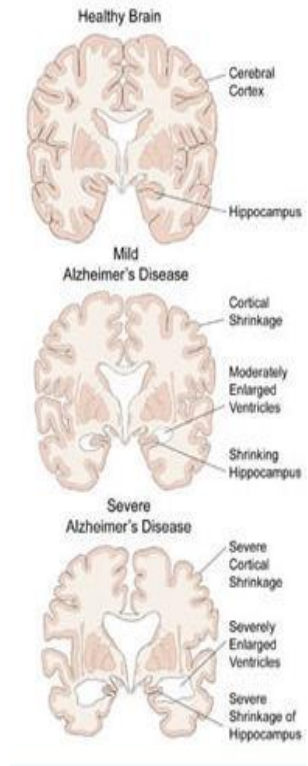


Illustration by Bob Morreale, American Health Assistance Foundation

Decoding genome function

TGCCAAGCAGCAAAGTTTTGCTGCTGTTTATTTTTGTAGCTCTTACTATATTCT
ACTTTTACCATTGAAAATATTGAGGAAGTTATTTATATTTCTATTTTTTATATAT
TATATATTTTATGTATTTAATATTACTATTACACATAATTATTTTTTATATATATGA
AGTACCAATGACTTCCTTTTCCAGAGCAATAATGAAATTTACAGTATGAAA
ATGGAAGAAATCAATAAAATTATACGTGACCTGTGGCGAAGTACCTATCGTG
GACAAGGTGAGTACCATGGTGTATCACAATGCTCTTTCCAAAGCCCTCTCC
GCAGCTCTTCCCCTTATGACCTCTCATCATGCCAGCATTACCTCCCTGGACCC
CTTTCTAAGCATGTCTTTGAGATTTTCTAAGAATTCTTATCTTGGCAACATCTT
GTAGCAAGAAAATGTAAAGTTTTCTGTTCCAGAGCCTAACAGGACTTACATA
TTTGACTGCAGTAGGCATTATATTTAGCTGATGACATAATAGGTTCTGTCATA
GTGTAGATAGGGATAAGCCAAAATGCAATAAGAAAAACCATCCAGAGGAA
ACTCTTTTTTTTTTCTTTTTCTTTTTTTTTTTTCCAGATGGAGTCTCGCACTTC
TCTGTCACCCGGGCTGGAGCGCAGTGGTGCAATCTTGGCTCACTGCAACCT
CCACCTCCTGGGTTTCAGGTGATTCTCCACCTCAGCCTCCCGAGTAGTAGCT
GGAATTACAGGTGCGCGCTCCACACCTGGCTAATTTTTTTGTATTCTTAGTA
GAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACCTCCTGCCCTCA
GGTGATCTGCCACCTTGGCCTCCAGTGTTGGGTTTACAGGCGTGAGCCA
CCGCGCCTGGCCTGGAGGAACTCTAACAGGGGAACTAAGAAAGAGTTG
AGGCTGAGGAACTGGGGCATCTGGGTTGCTTCTGGCCAGACCACCAGGCT
CTTGAATCCTCCAGCCAGAGAAAGAGTTTCCACACCAGCCATTGTTTTCT
CTGGTAATGTCAGCCTCATCTGTTGTTCTAGGCTTACTTGATATGTTTGTA
ATGACAAAAGGCTACAGAGCATAGGTTCTCTAAAATATTCTTCTTCTGTGT
CAGATATTGAATACATAGAAATACGGTCTGATGCCGATGAAAATGTATCAGCT
TCTGATAAAAGGCGGAATTATAACTACCGAGTGGTGATGCTGAAGGGAGAC
ACAGCCTTGGATATGCGAGGACGATGCAGTGCTGGACAAAAGGCAGGTAT
CTCAAAGCCTGGGGAGCCAACTCACCCAAGTAACTGAAAGAGAGAAACA
AACATCAGTGCAGTGGAAGCACCCAAGGCTACACCTGAATGGTGGGAAGC
TCTTTGCTGCTATATAAAATGAATCAGGCTCAGCTACTATTATT

~ 3 billion nucleotides

Decoding genome function

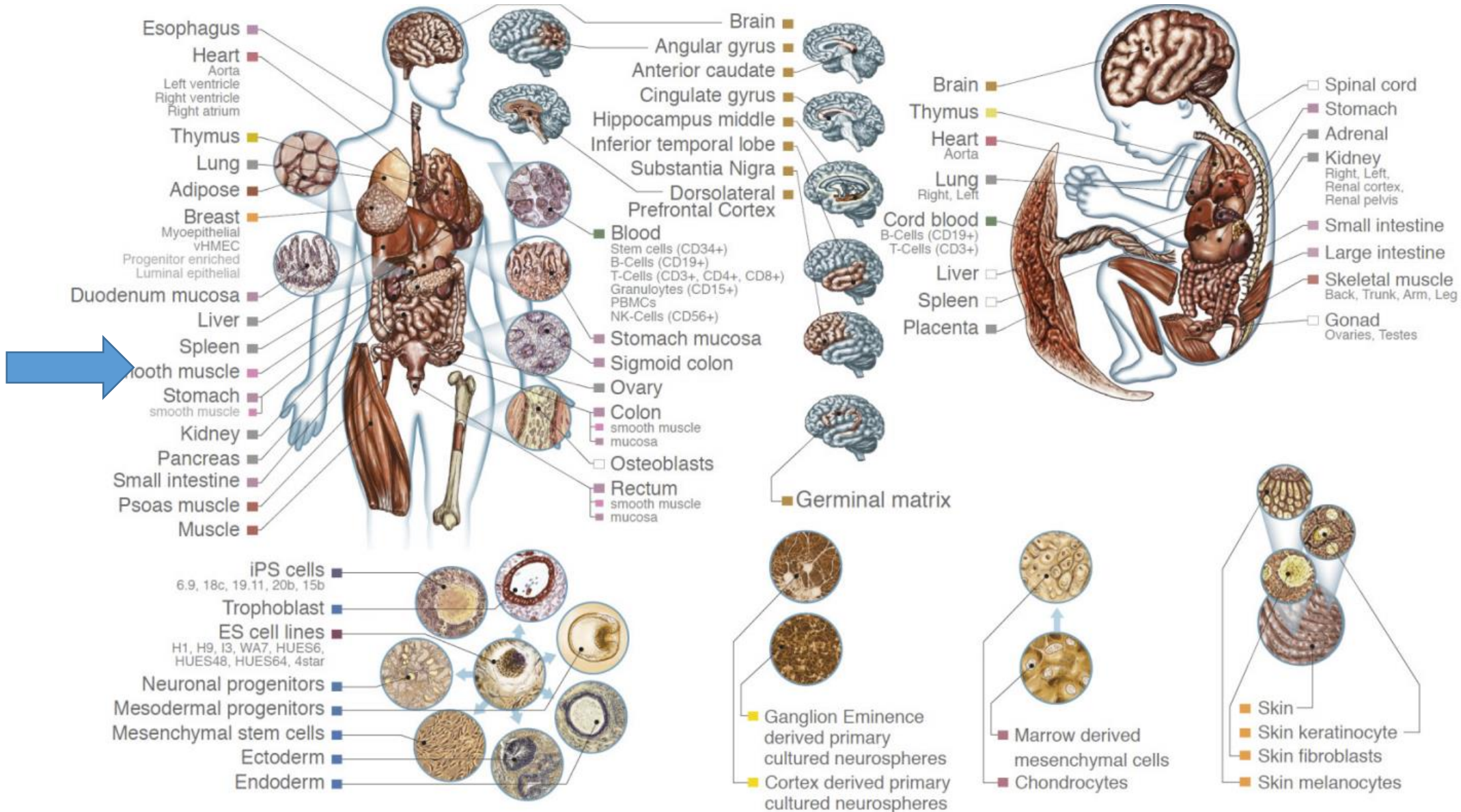
TGCCAAGCAGCAAAGTTTTGCTGCTGTTTATTTTTGTAGCTCTTACTATATTCT
ACTTTTACCATTGAAAATATTGAGGAAGTTATTTATATTTCTATTTTTTATATAT
TATATATTTTATGTATTTTAATATACTATTACACATAATTATTTTTTATATATATGA
AGTACCAATGACTTCCTTTTCCAGAGCAATAATGAAATTTACAGTATGAAA
ATGGAAGAAATCAATAAAATTATACGTGACCTGTGGCGAAGTACCTATCGTG
GACAAGGTGAGTACCATGGTGTATCAQAAATGCTCTTTCCAAAGCCCTCTCC
GCAGCTCTTCCCCTTATGACCTCTCATCATGCCAGCATTACCTCCCTGGACCC
CTTTCTAAGCATGTCTTTGAGATTTTCTAAGAATTCTTATCTTGGCAACATCTT
GTAGCAAGAAAATGTAAAGTTTTCTGTTCCAGAGCCTAACAGGACTTACATA
TTTGACTGCAGTAGGCATTATATTTAGCTGATGACATAATAGGTTCTGTCATA
GTGTAGATAGGGATAAGCCAAAATGCAATAAGAAAAACCATCCAGAGGAA
ACTCTTTTTTTTTTCTTTTTCTTTTTTTTTTTTTTCCAGATGGAGTCTCGCACTTC
TCTGTCACCCGGGCTGGAGCGCAGTGGTGCAATCTTGGCTCACTGCAACCT
CCACCTCCTGGGTTTCAGGTGATTCTCCACCTCAGCCTCCCGAGTAGTAGCT
GGAATTACAGGTGCGCGCTCCACACCTGGCTAATTTTTTGTATTCTTAGTA
GAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACCTCCTGCCCTCA
GGTGATCTGCCACCTTGGCCTCCAGTGTTGGGTTTACAGGCGTGAGCCA
CCGCGCCTGGCCTGGAGGAACTCTAACAGGGGAACTAAGAAAGAGTTG
AGGCTGAGGAACTGGGGCATCTGGGTTGCTTCTGGCCAGACCACCAGGCT
CTTGAATCCTCCAGCCAGAGAAAGAGTTTCCACACCAGCCATTGTTTTCT
CTGGTAATGTCAGCCTCATCTGTTGTTCTAGGCTTACTTGATATGTTTGTA
ATGACAAAAGGCTACAGAGCATAGGTTCTCTAAAATATTCTTCTTCTGTGT
CAGATATTGAATACATAGAAATACGGTCTGATGCCGATGAAAATGTATCAGCT
TCTGATAAAAGGCGGAATTATAACTACCGAGTGGTGATGCTGAAGGGAGAC
ACAGCCTTGGATATGCGAGGACGATGCAGTGCTGGACAAAAGGCAGGTAT
CTCAAAGCCTGGGGAGCCAACTCACCCAAGTAACTGAAAGAGAGAAACA
AACATCAGTGCAGTGAAGCACCCAAGGCTACACCTGAATGGTGGGAAGC
TCTTTGCTGCTATATAAAATGAATCAGGCTCAGCTACTATTATT

Function?

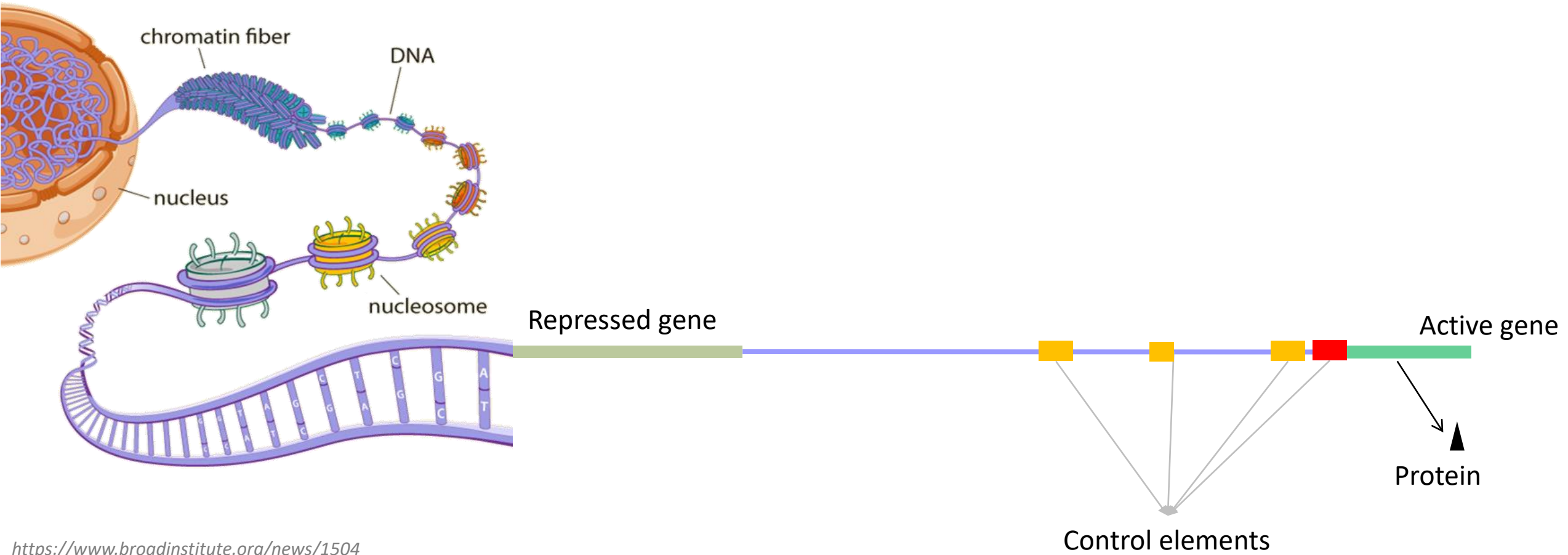
~ 3 billion nucleotides

One genome ⇔ many cell types

ACCAGTTACGACGG
 TCAGGGTACTGATA
 CCCCAAACCGTTGA
 CCGCATTTACAGAC
 GGGGTTTGGGTTTT
 GCCCCACACAGGTA
 CGTTAGCTACTGGT
 TTAGCAATTTACCG
 TTACAACGTTTACA
 GGGTTACGGTTGGG
 ATTTGAAAAAAGT
 TTGAGTTGGTTTTT
 TCACGGTAGAACGT
 ACCTTACAAA.....

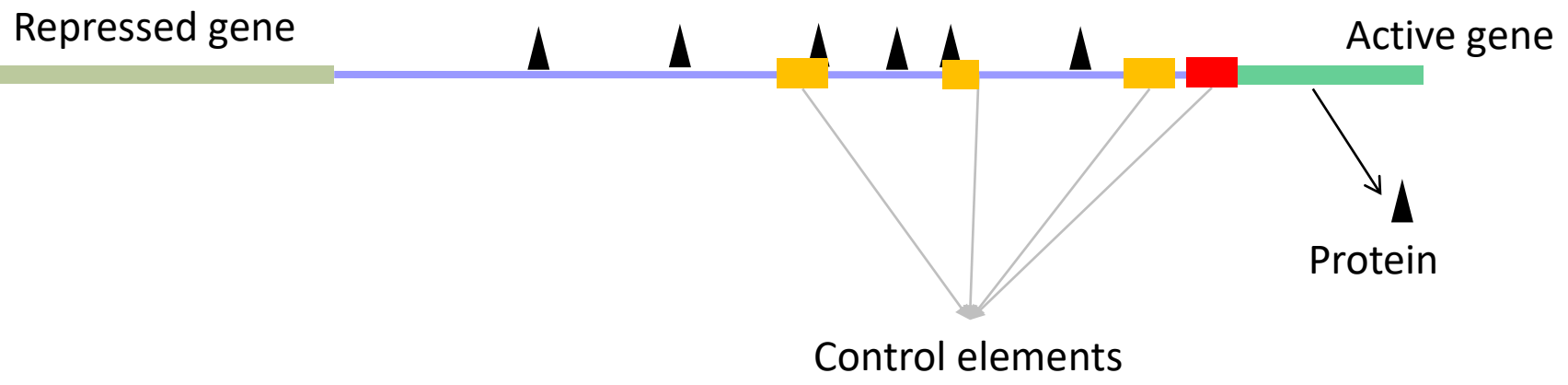
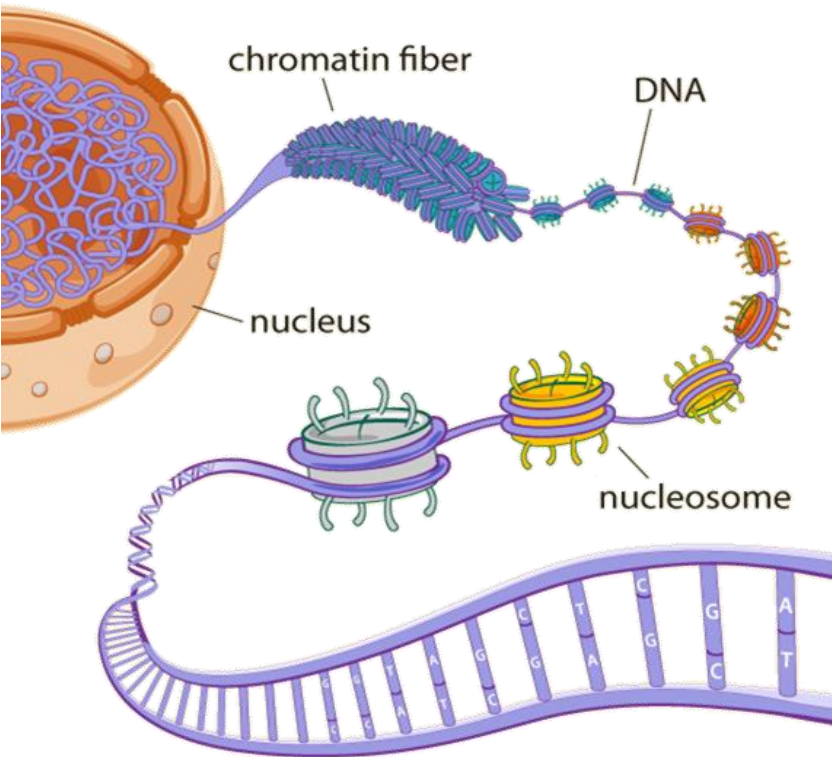


Molecular mapping of functional components of the genome



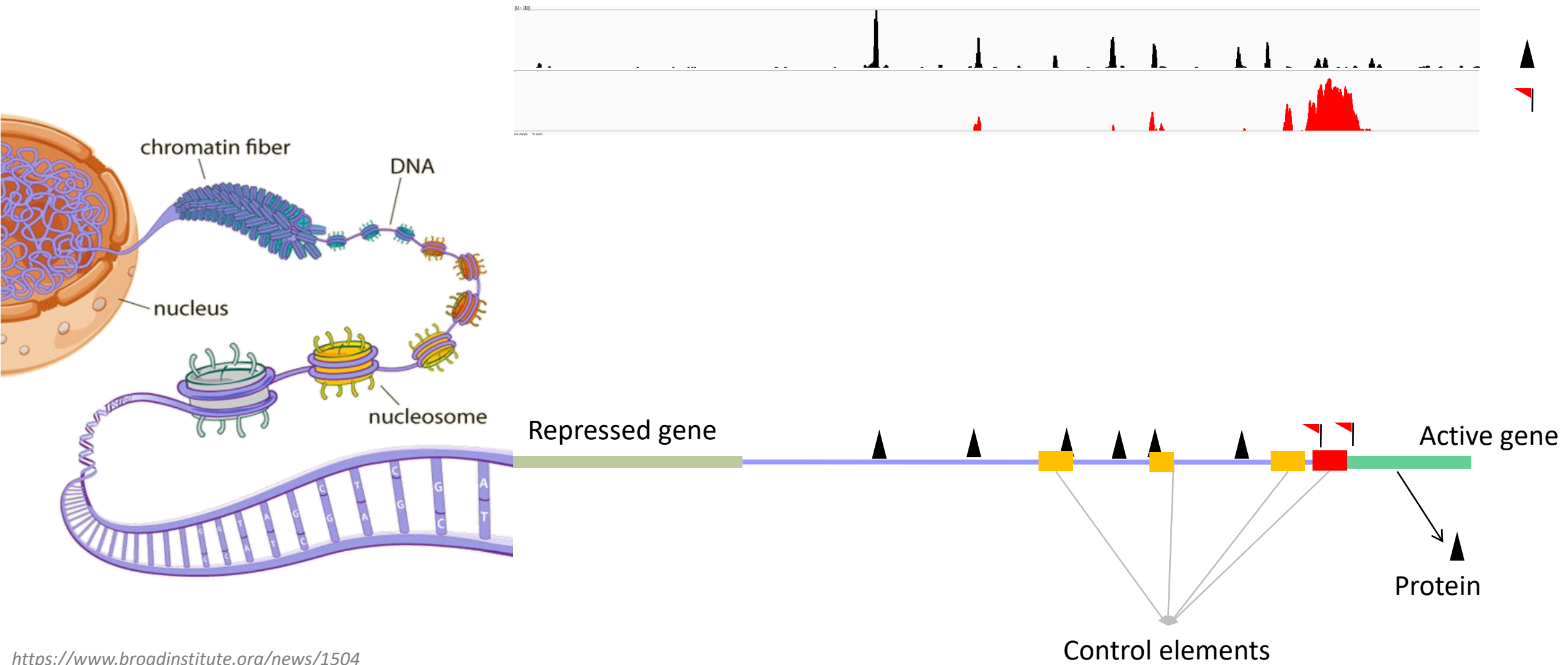
<https://www.broadinstitute.org/news/1504>

Molecular mapping of functional components of the genome

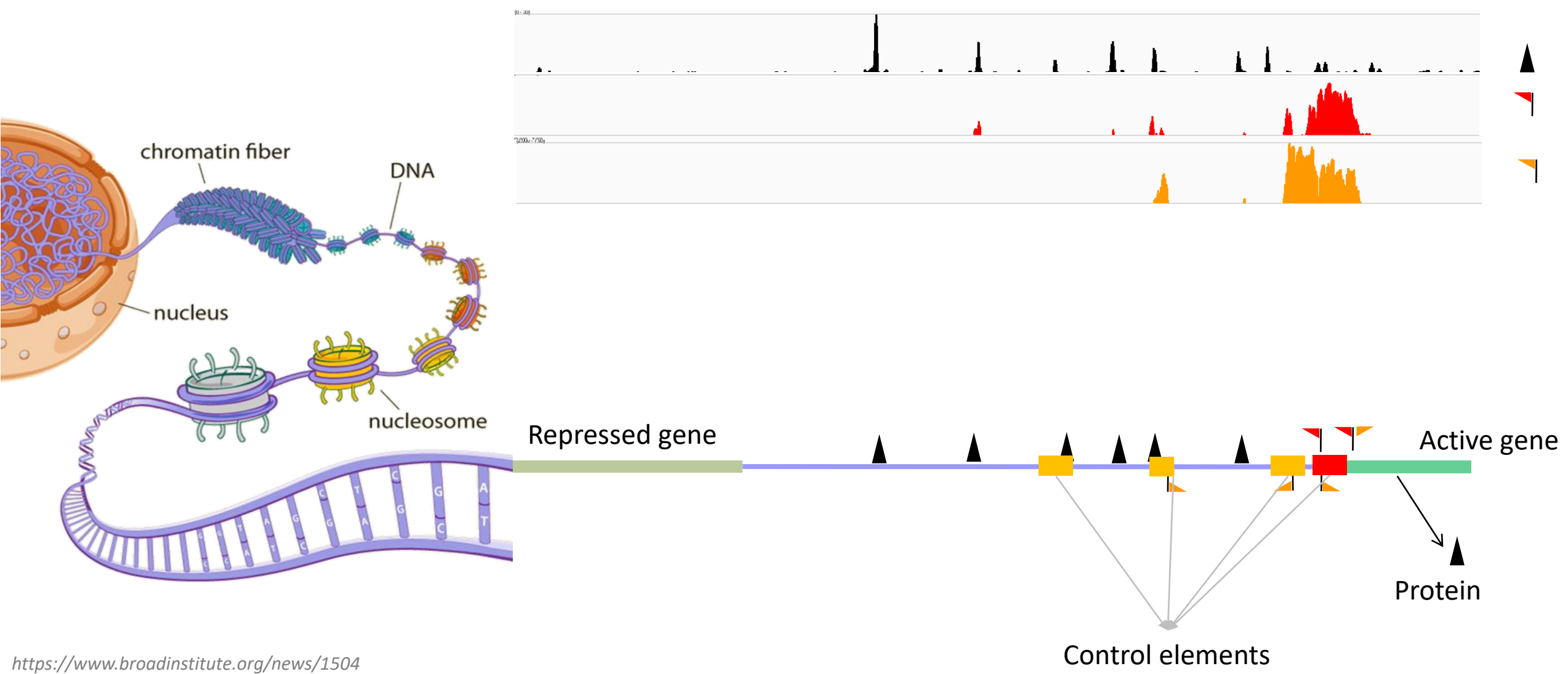


<https://www.broadinstitute.org/news/1504>

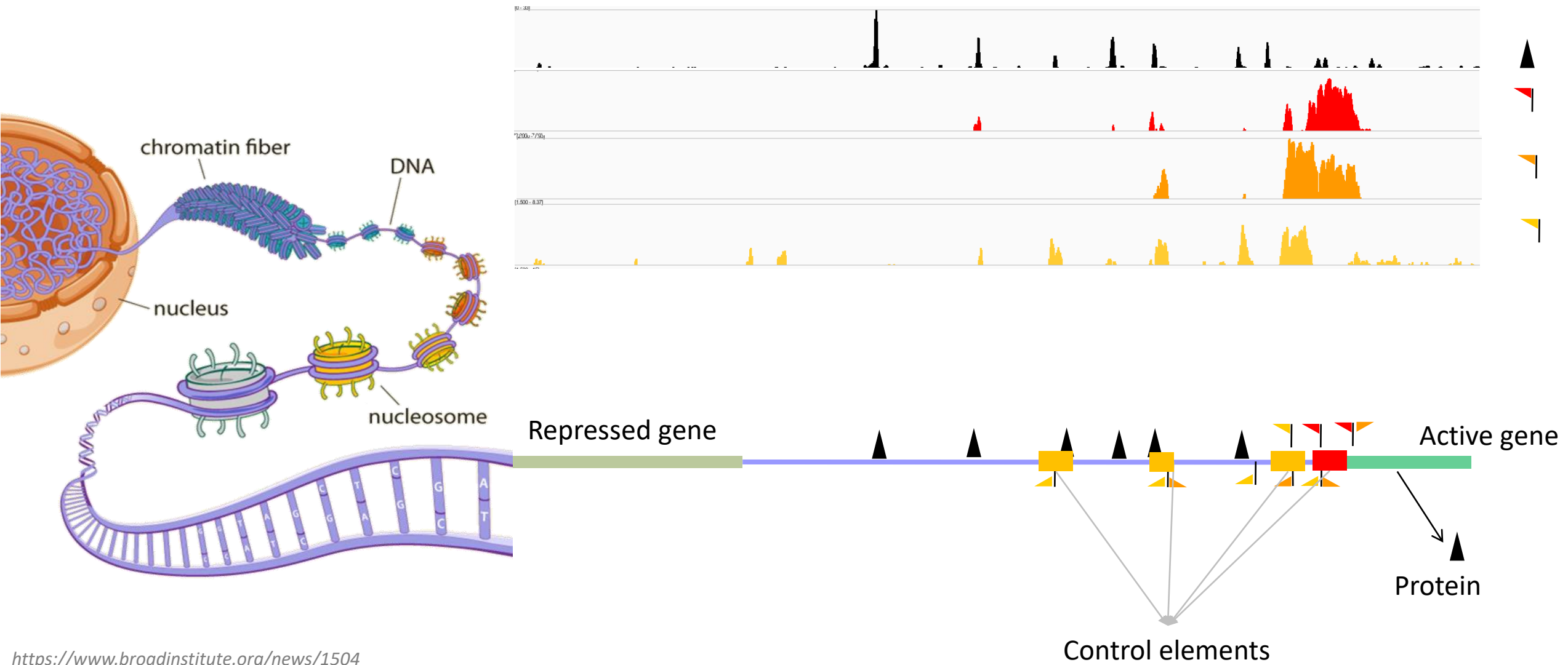
Molecular mapping of functional components of the genome



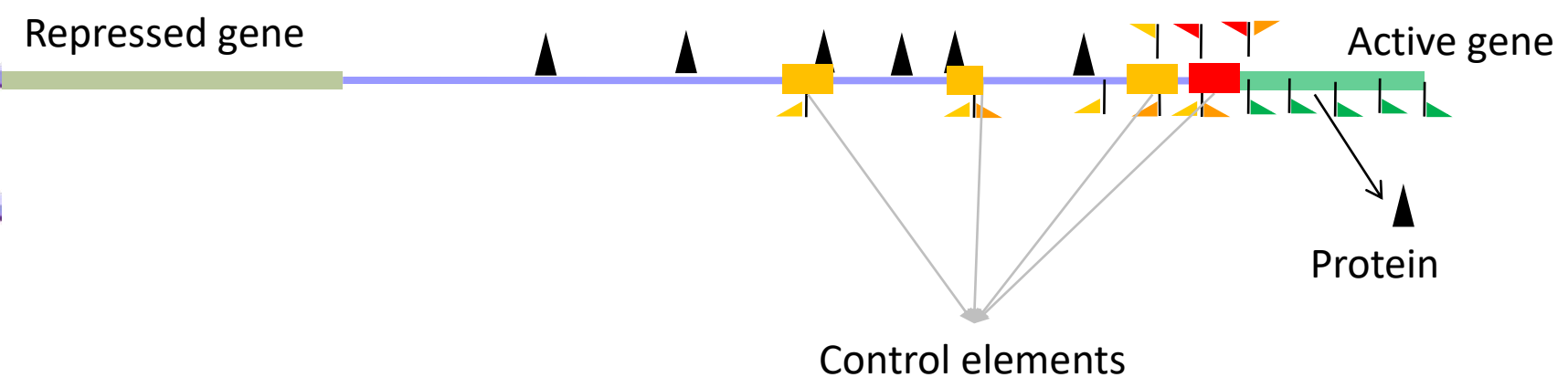
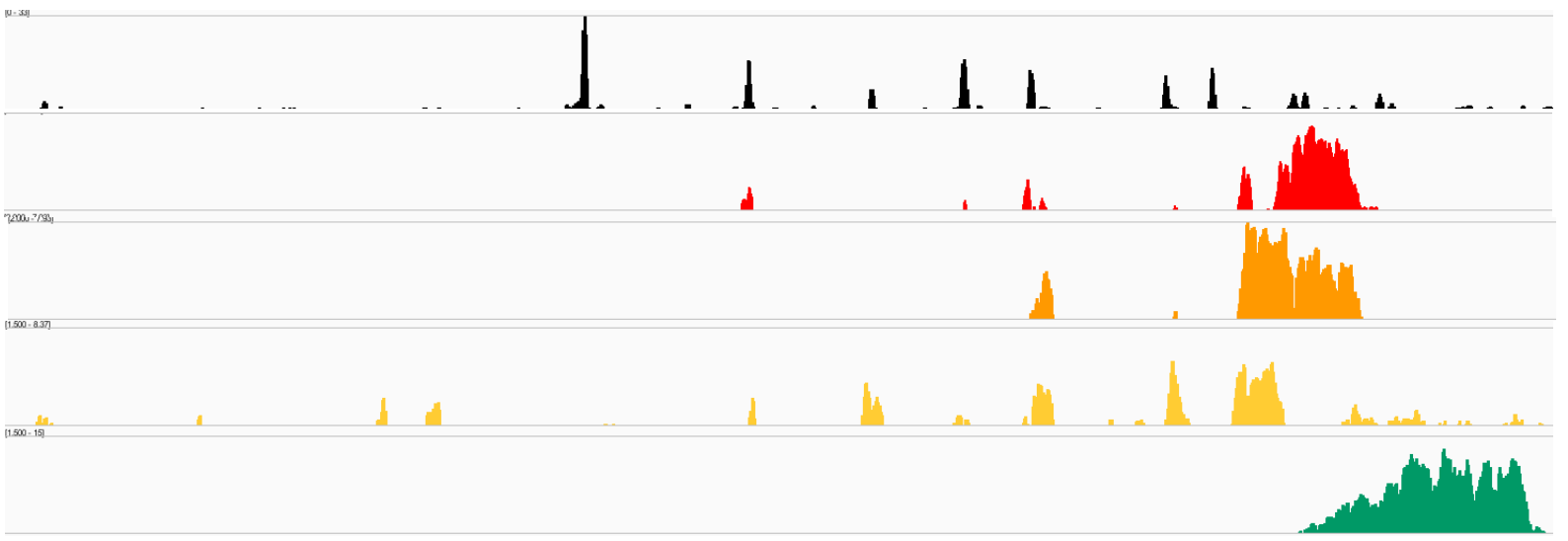
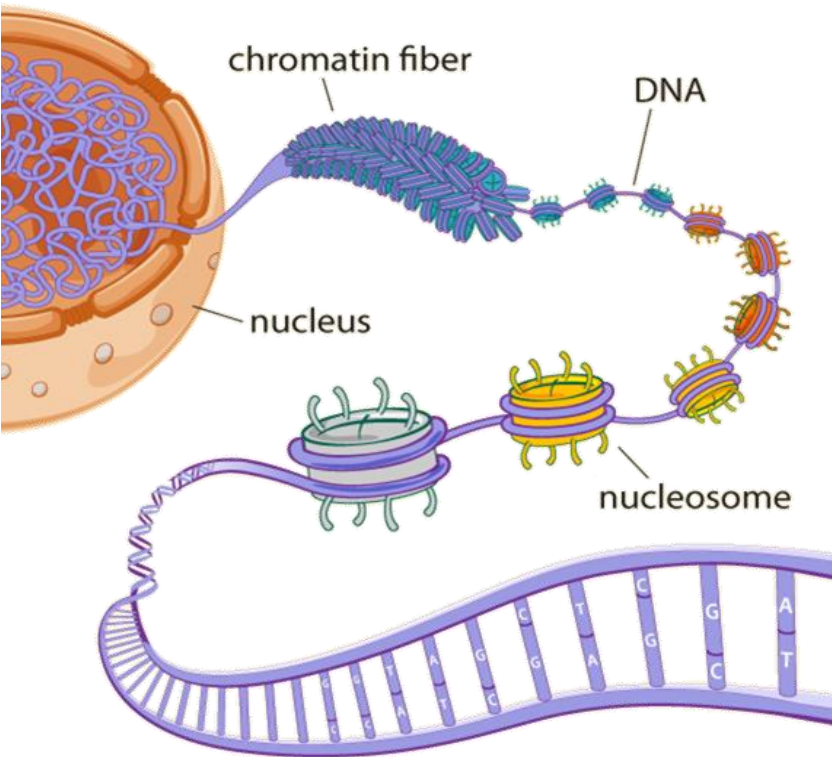
Molecular mapping of functional components of the genome



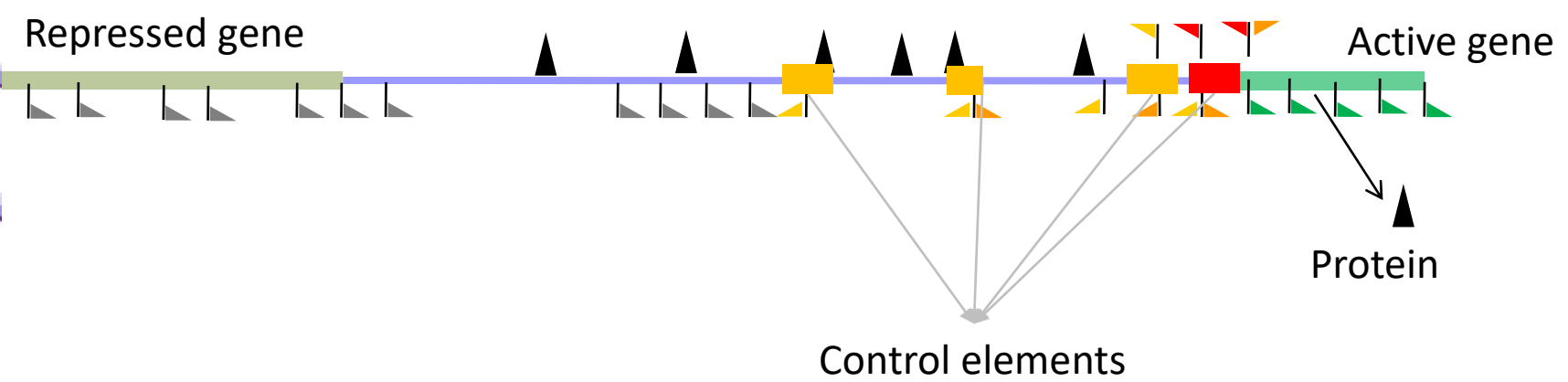
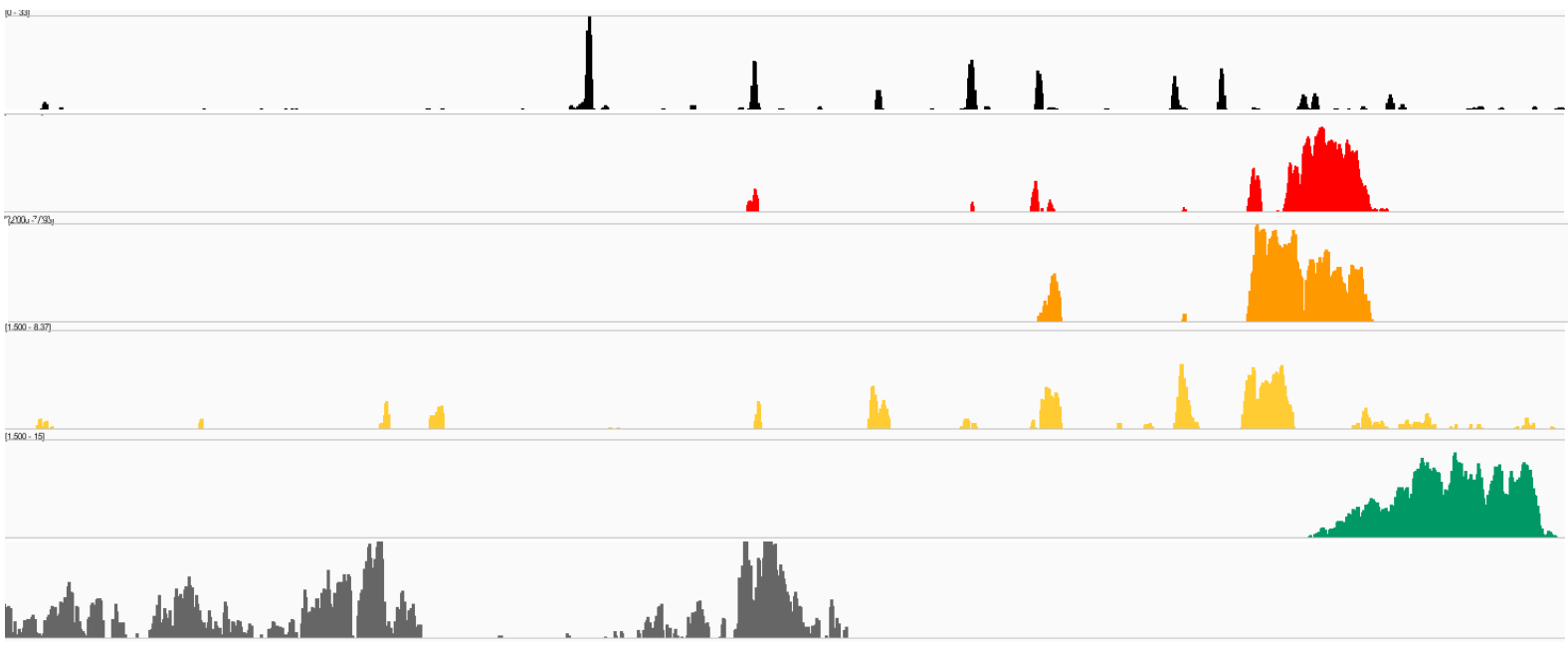
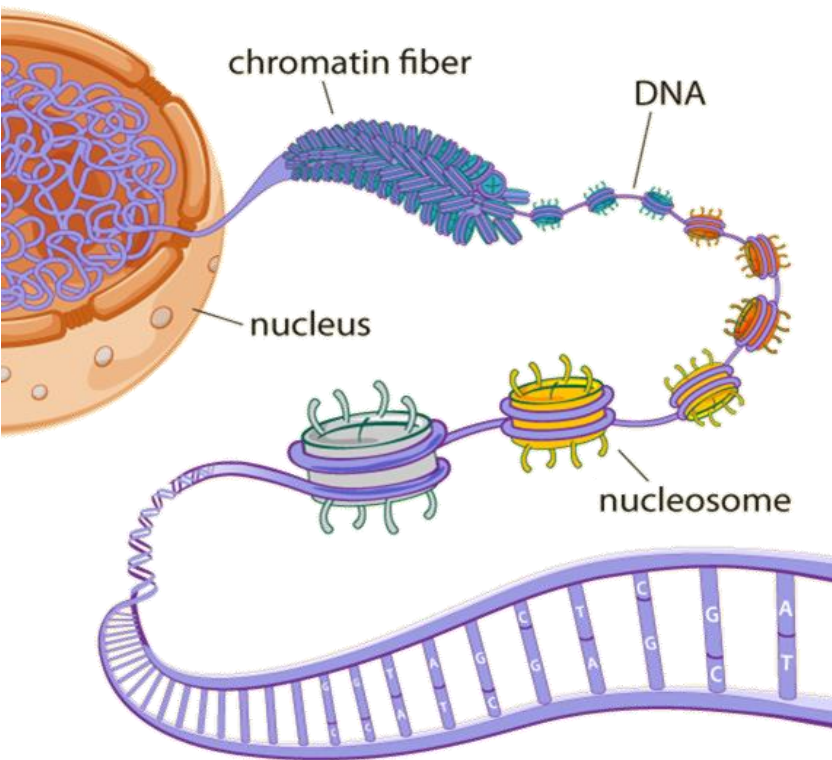
Molecular mapping of functional components of the genome



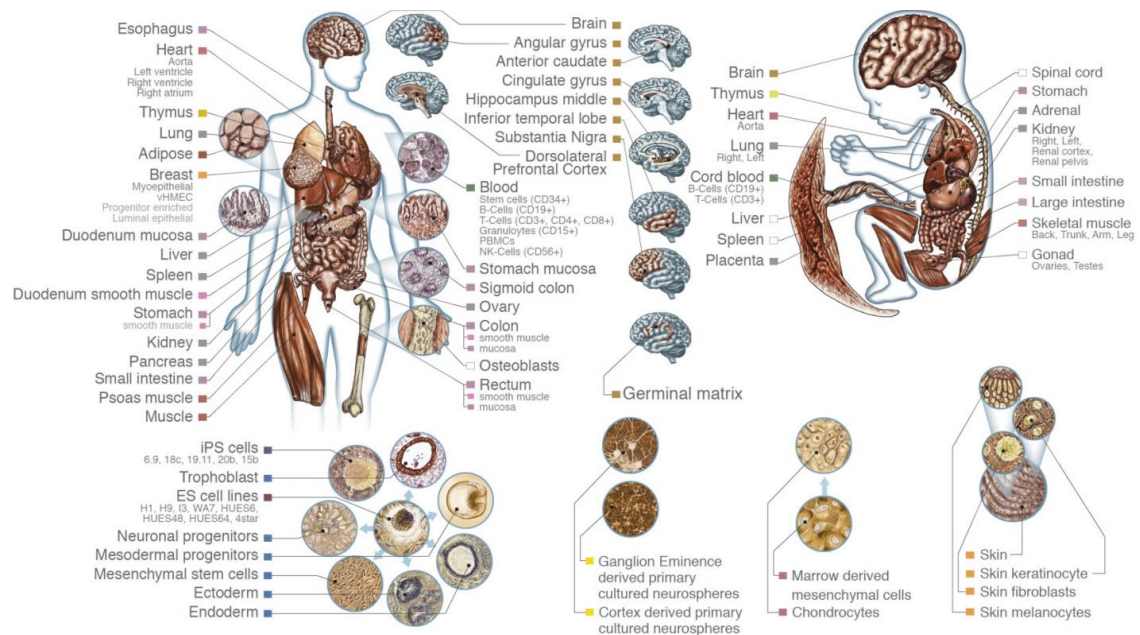
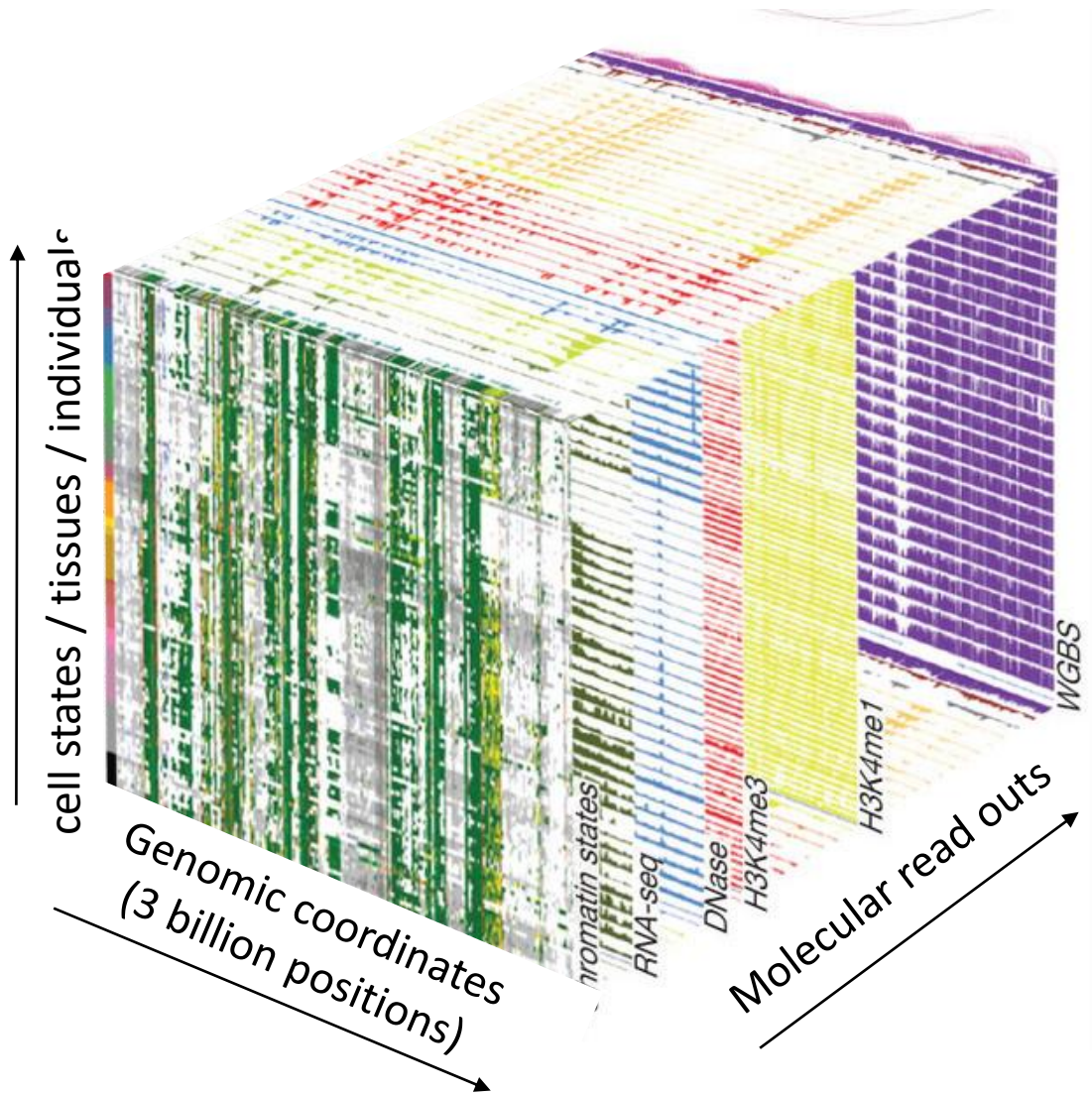
Molecular mapping of functional components of the genome



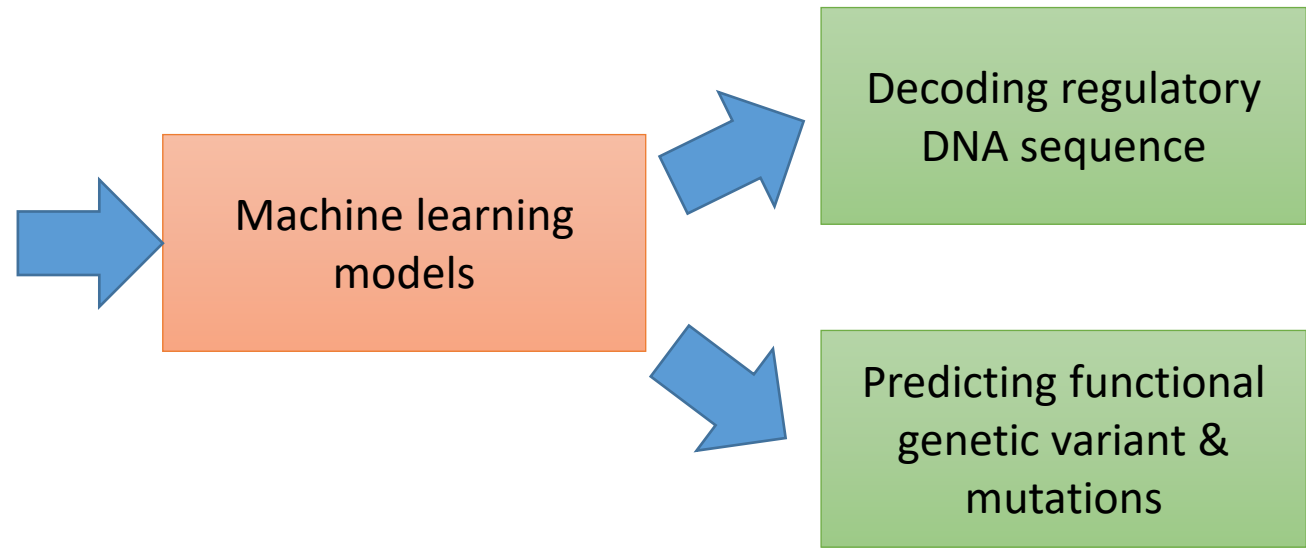
Molecular mapping of functional components of the genome



<https://www.broadinstitute.org/news/1504>



100s of Cell-Types/Tissues



Dunham, Kundaje et al. 2012 Nature
Kundaje et al. 2015 Nature

>95% of disease variants are not disrupting protein coding gene regions

Benign

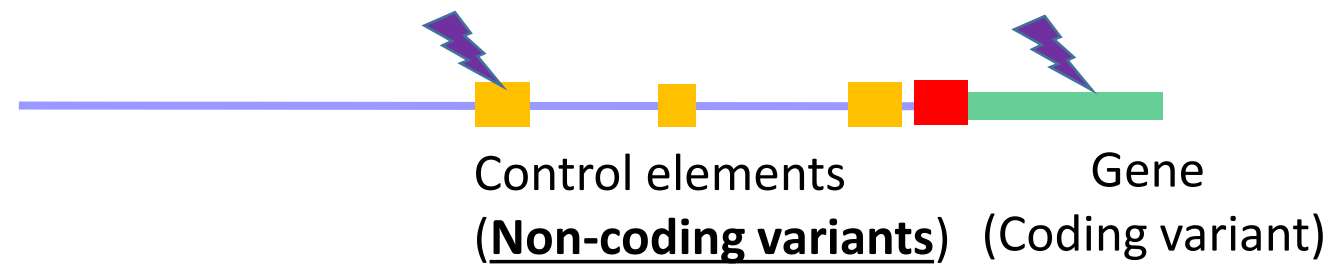
.....ACTGATCG**C**AATCG.....



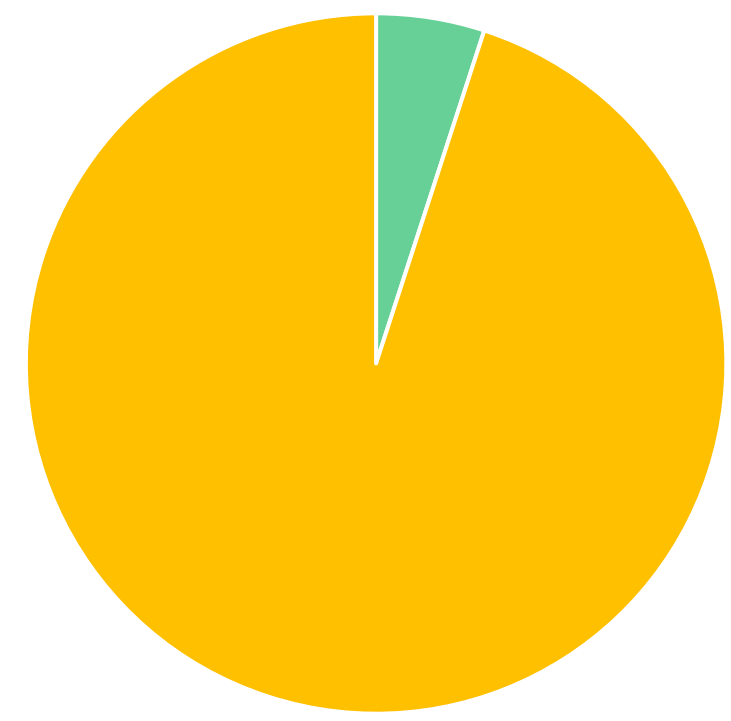
.....ACTGATCG**G**AATCG.....

Risk

>95% of disease variants are not disrupting protein coding gene regions



Benign
.....ACTGATCG**C**AATCG.....
.....ACTGATCG**G**AATCG.....
Risk



■ Coding ■ Non-coding

Decoding syntax of regulatory DNA



Ziga Avsec



Avanti Shrikumar



Melanie Weilert

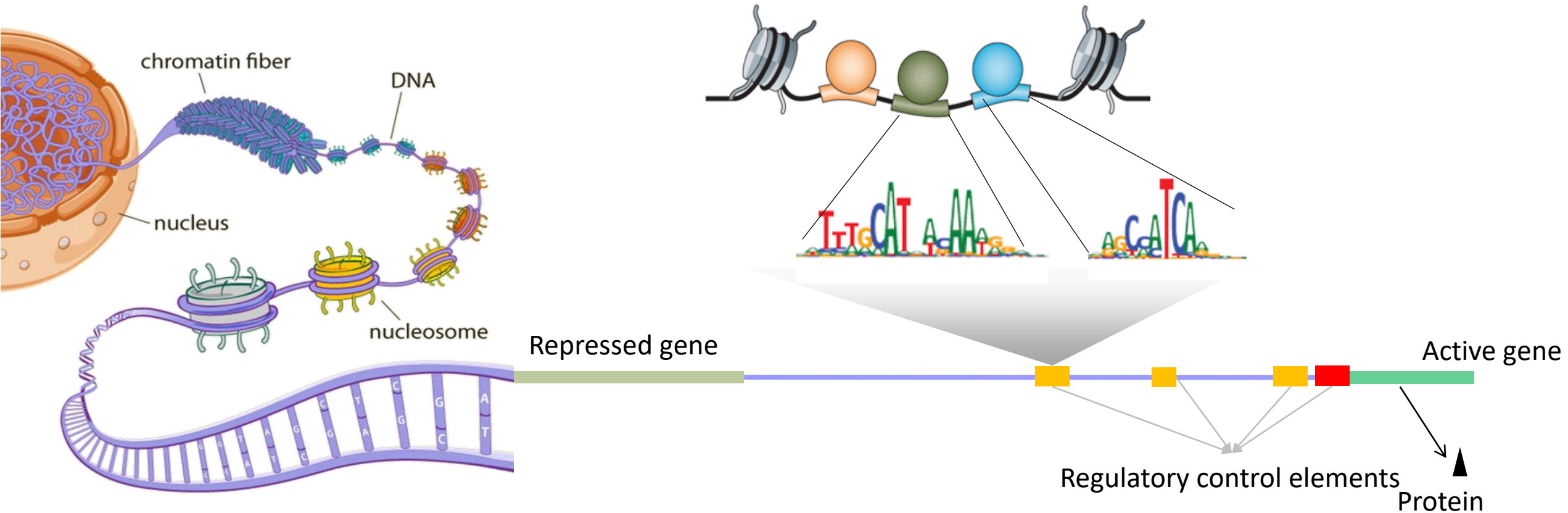


Amr Mohamed

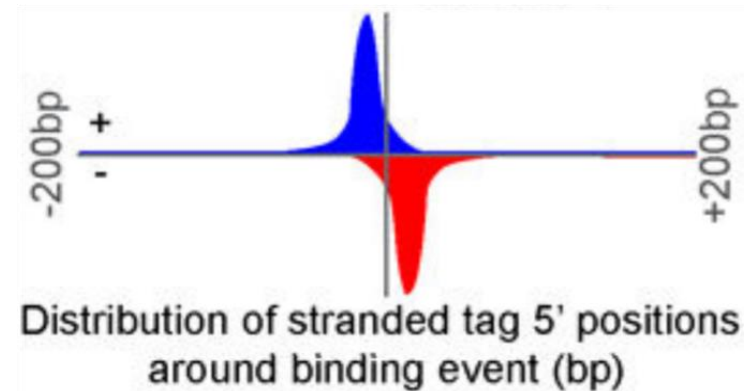
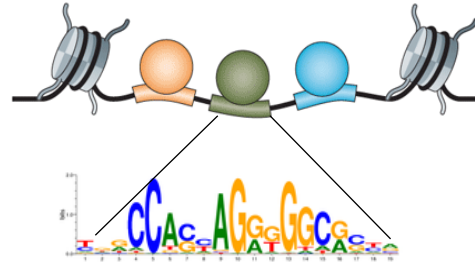


Julia Zeitlinger

Control elements encode syntax of DNA words recognized by regulatory protein complexes



High-resolution 'shapes' of experimental profiles capture exquisite information about protein-DNA contacts

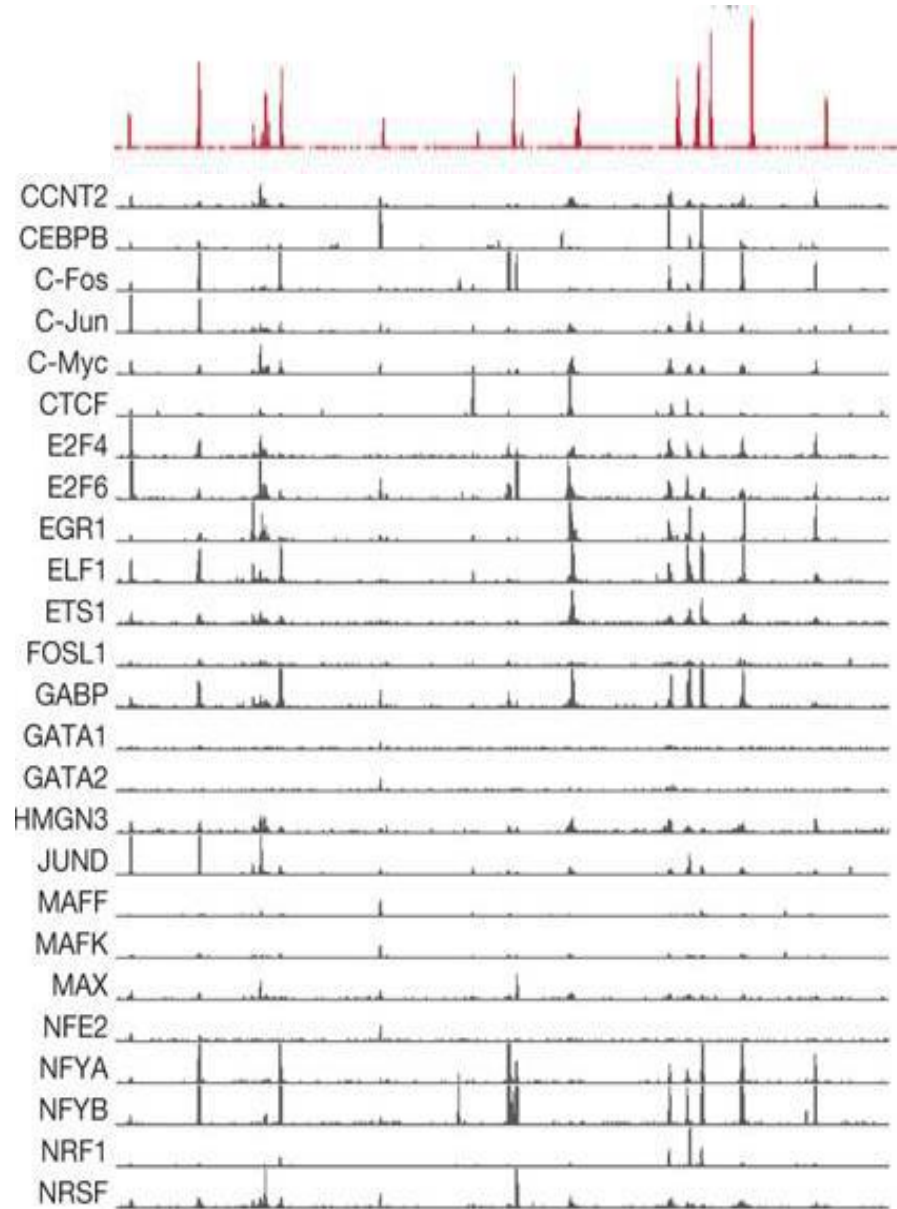


Protein-DNA binding expt.



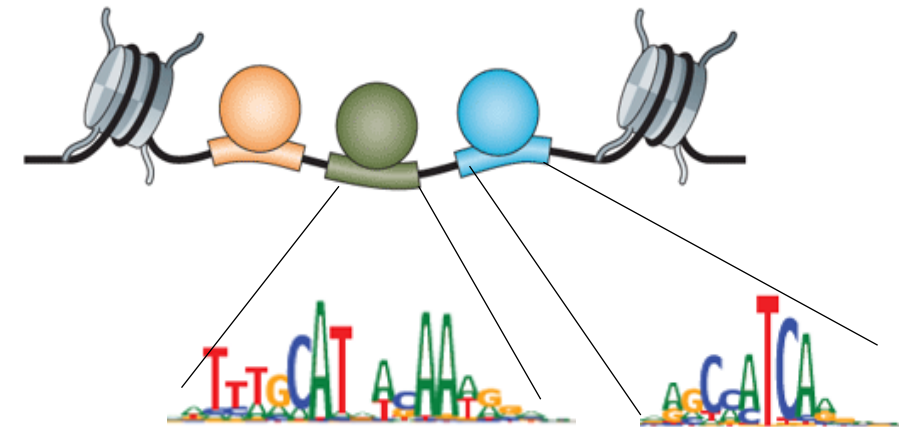
DNA accessibility experiments

Genome-wide maps of regulatory DNA (control elements)



DNA accessibility experiments

Protein-DNA binding experiments

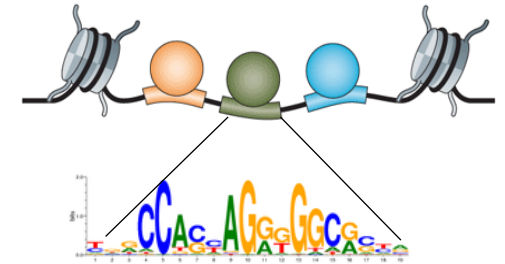
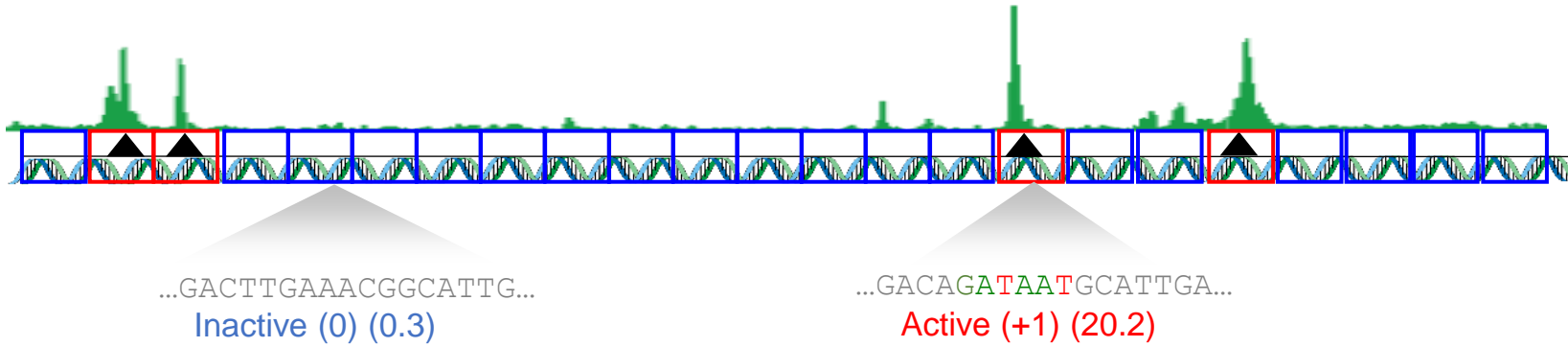


?

Motif syntax: rules of arrangement, preferred spacing, orientation => cooperativity

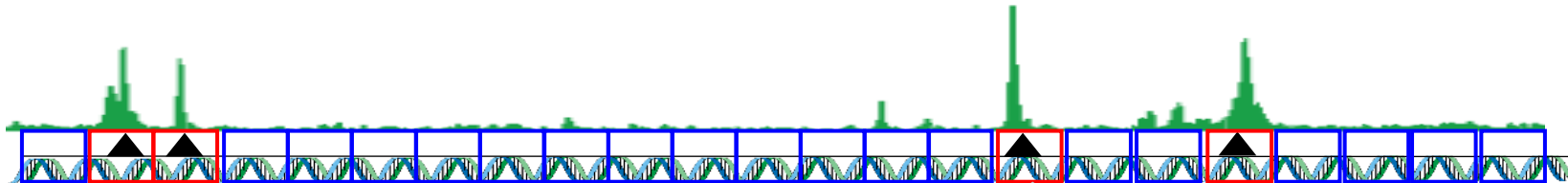
Predictive model of regulatory DNA

Transcription factor CHIP-seq data OR chromatin accessibility (DNase-seq / ATAC-seq data)



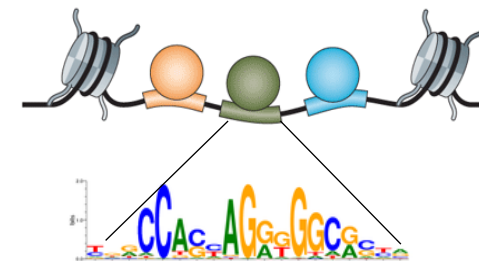
Predictive model of regulatory DNA

Transcription factor CHIP-seq data OR chromatin accessibility (DNase-seq / ATAC-seq data)



...GACTTGAAACGGCATTG...
Inactive (0) (0.3)

...GACAGATAATGCATTGA...
Active (+1) (20.2)



...GACAGATAATGCATTGA...

...ACTGTCATGGATATTCT...

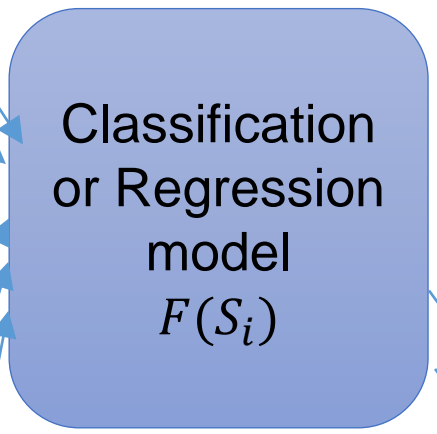
...GATATTCTACTGTAAG...

DNA sequences (S_i)

...CAACCTTGAACGGCATTG...

...GACTTGAAACGGCATTG...

...CAGTATGCATACGTGAA...



Class = +1 (20.2)

Class = +1 (10.6)

Class = +1 (15.8)

Measured Labels (Y_i)

Class = 0 (0.3)

Class = 0 (1.2)

Class = 0 (3.5)



Bound



Unbound

Arvey et al. 2012
Ghandi et al. 2014
Setty et al. 2015

Deep learning framework for decoding regulatory DNA



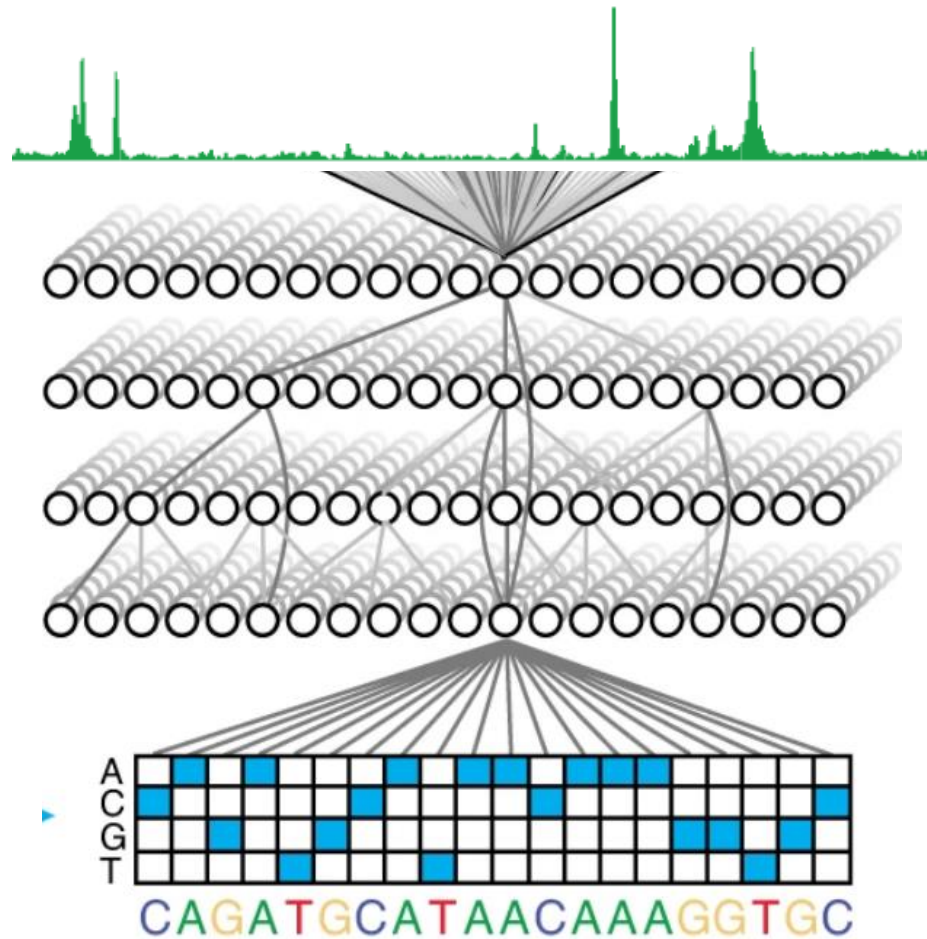
Ziga Avsec



Anusri Pampari



Anna Shcherbina



Avanti Shrikumar



Alex Tseng



Surag Nair



Jacob Schreiber

BPNet

(maps sequence to base-resolution profiles)

One model for every expt.

Avsec et al. 2021, Nature Genetics

Shrikumar et al. 2017, ICML

Tseng et al. 2020, NeurIPS

Nair et al, 2022, Bioinformatics

Schreiber et al. 2022, Biorxiv

Deep learning framework for decoding regulatory DNA



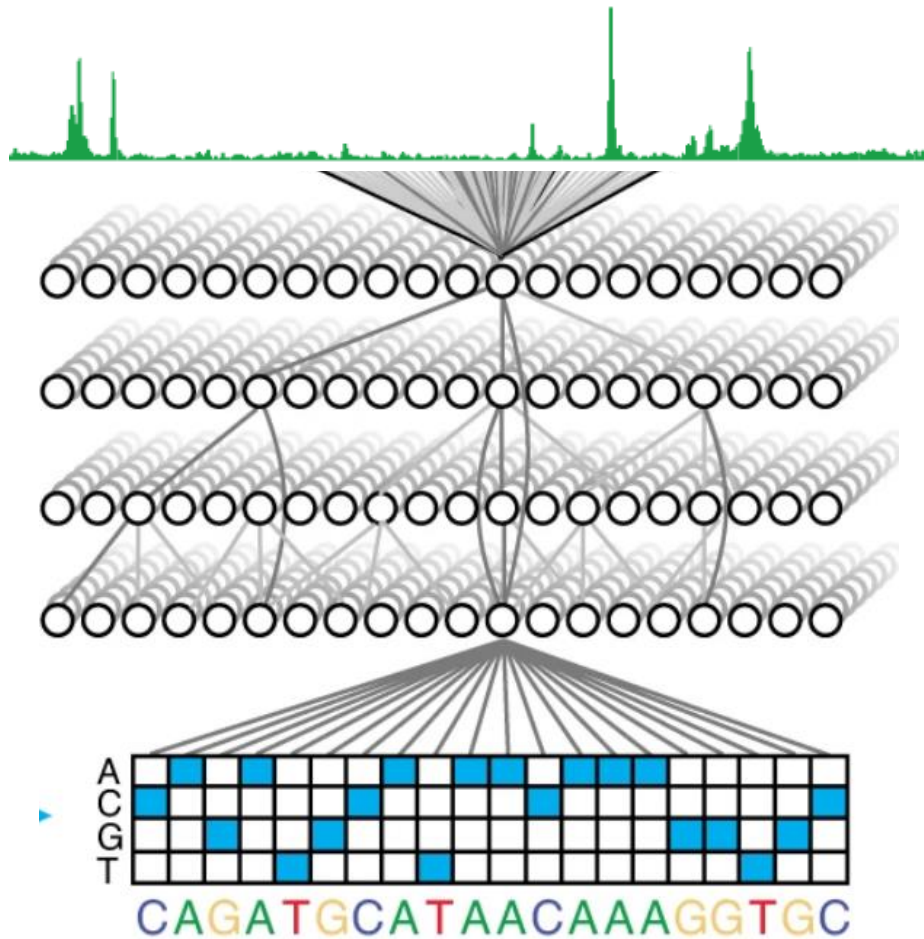
Ziga Avsec



Anusri Pampari



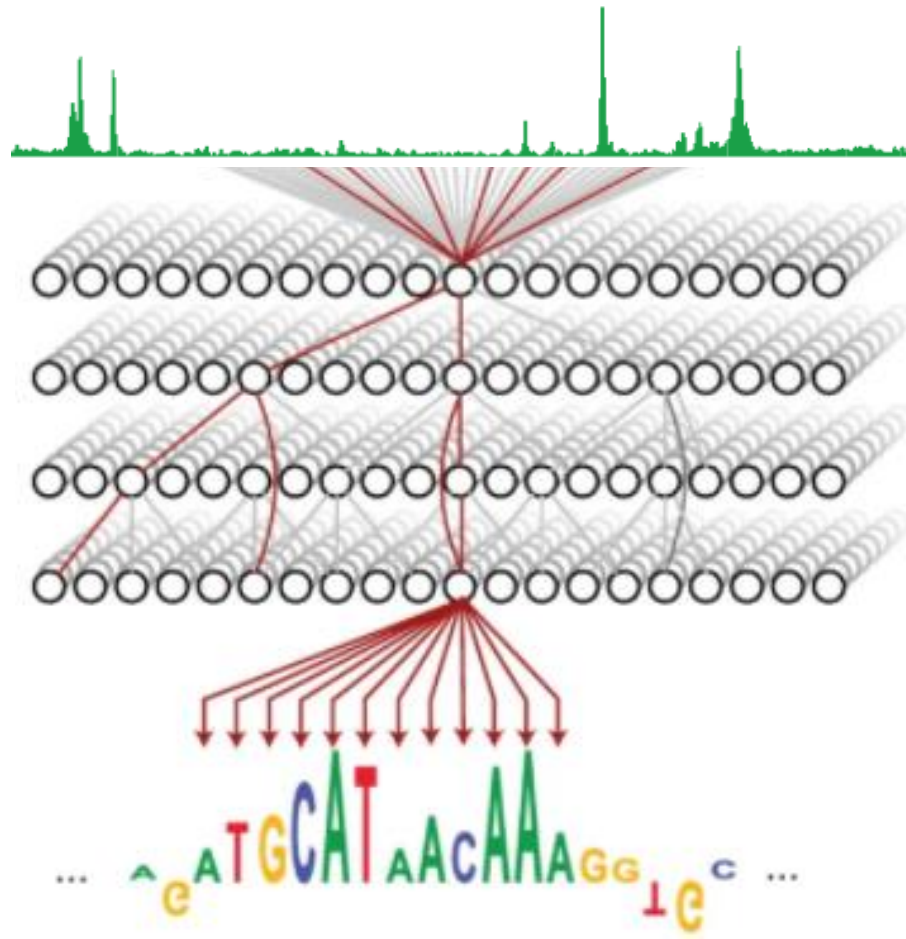
Anna Shcherbina



BpNet

(maps sequence to base-resolution profiles)
One model for every expt.

Avsec et al. 2021, Nature Genetics
Shrikumar et al. 2017, ICML
Tseng et al. 2020, NeurIPS
Nair et al, 2022, Bioinformatics
Schreiber et al. 2022, Biorxiv



DeepLIFT, FastISM, Yuzu

(infers contribution of every base in each control sequence thru lens of model)



Avanti Shrikumar



Alex Tseng

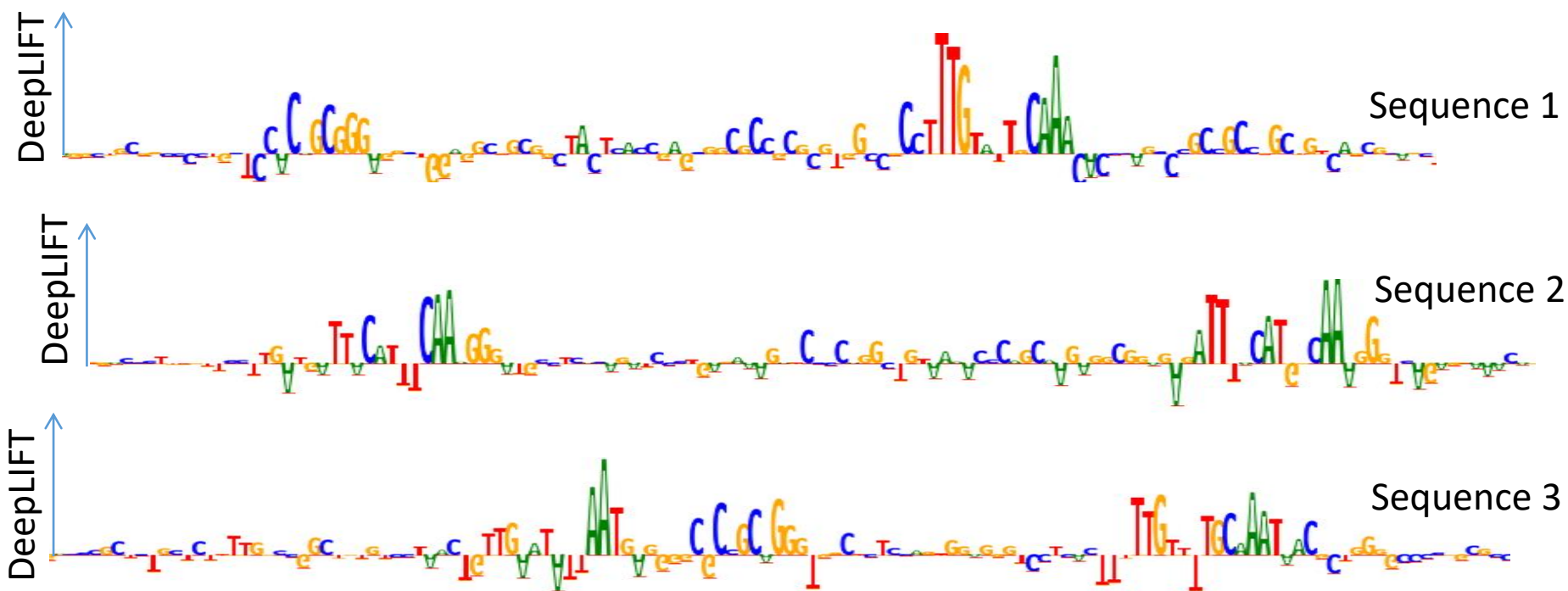


Surag Nair



Jacob Schreiber

TF-MoDISCO: Consolidate predictive subsequences into non-redundant motif representations

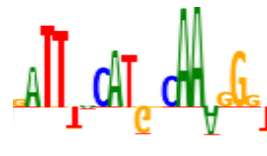
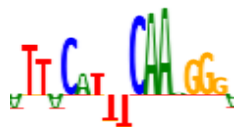
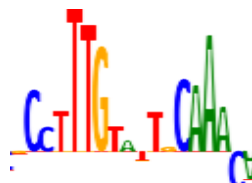


Avanti Shrikumar



Alex Tseng

TF-MoDISCO: Consolidate predictive subsequences into non-redundant motif representations

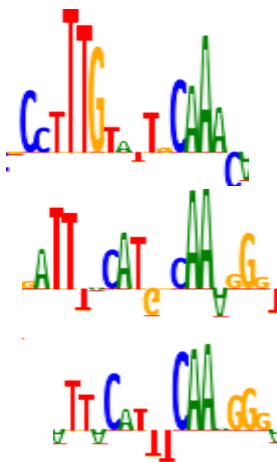
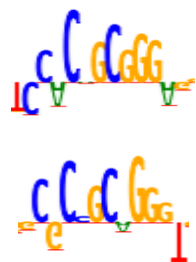


Avanti Shrikumar



Alex Tseng

TF-MoDISCO: Consolidate predictive subsequences into non-redundant motif representations

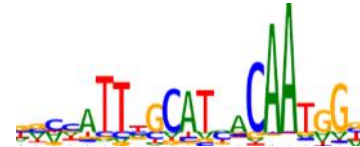
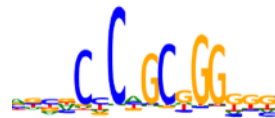


Avanti Shrikumar



Alex Tseng

TF-MoDISCO: Consolidate predictive subsequences into non-redundant motif representations

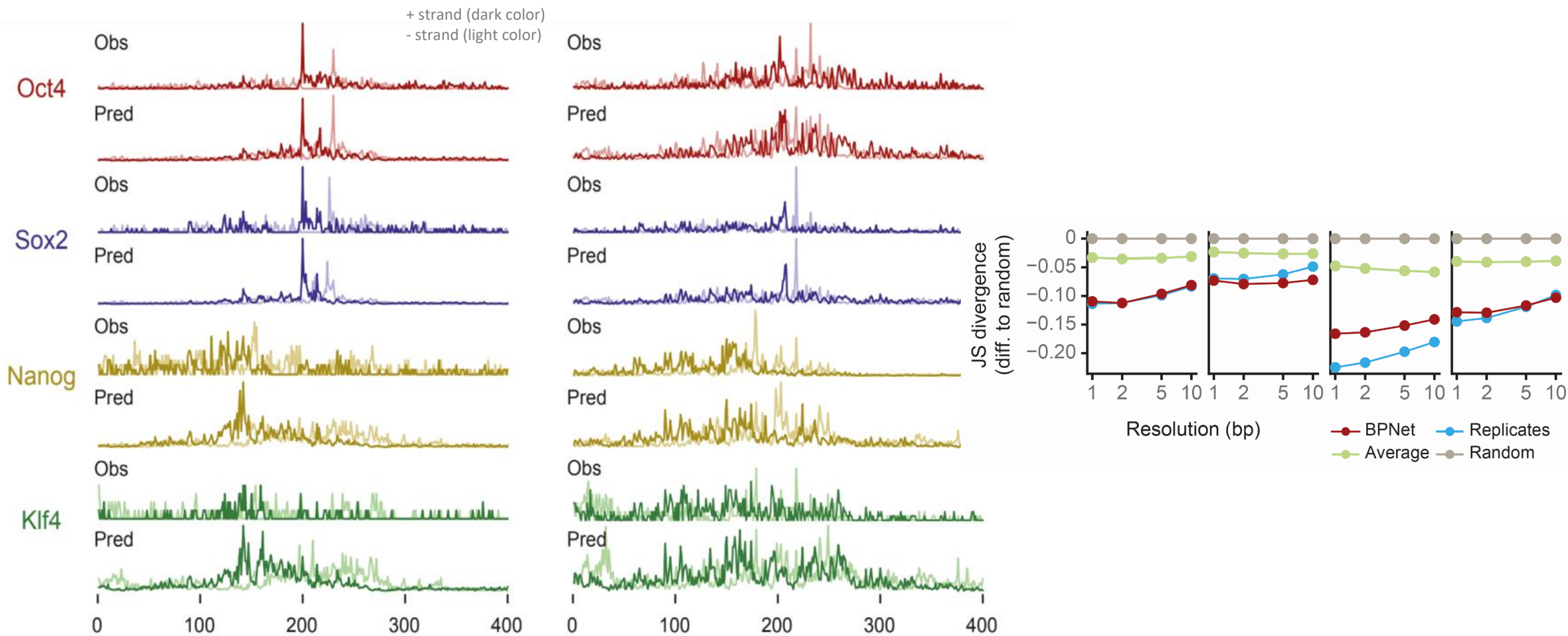


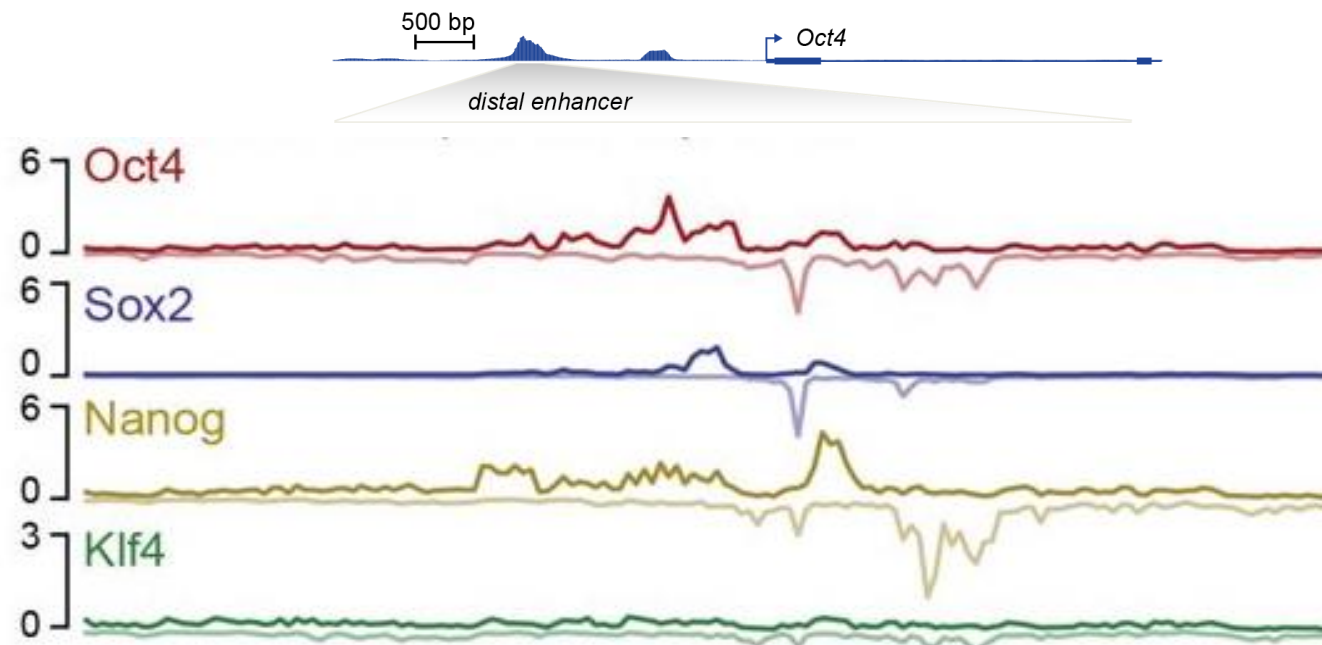
Avanti Shrikumar



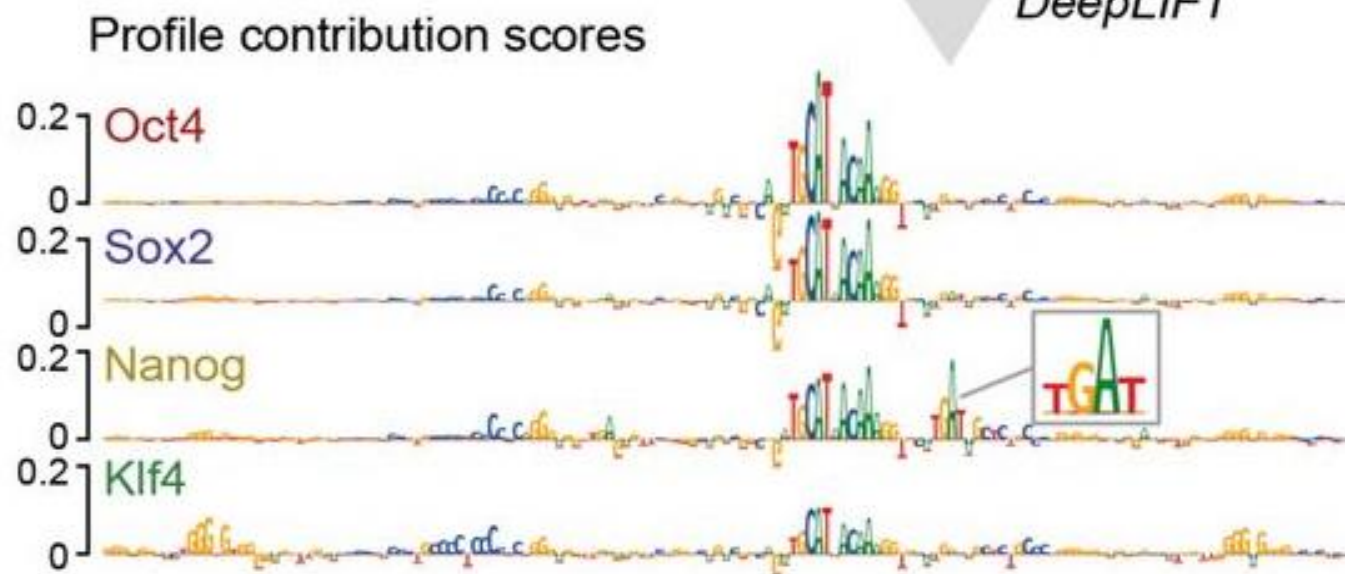
Alex Tseng

BpNet maps DNA sequence to base-resolution molecular profiles with accuracy on par with replicate expt. concordance





DeepLIFT

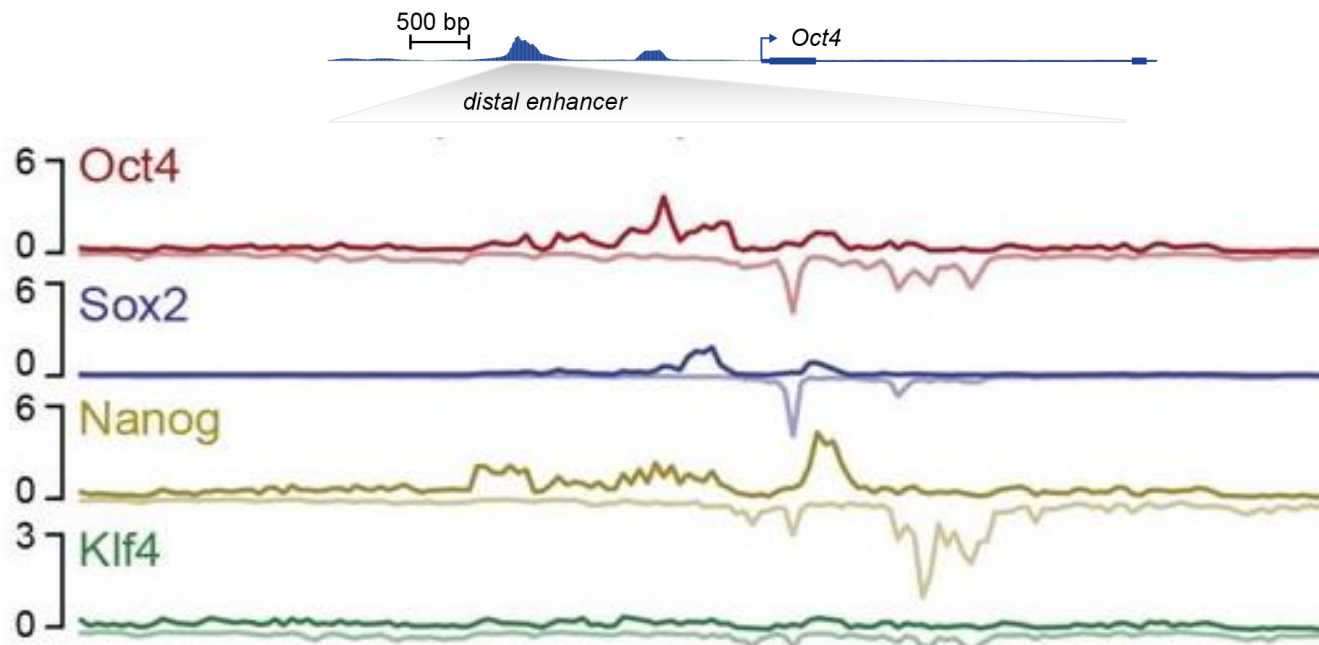


Avanti Shrikumar



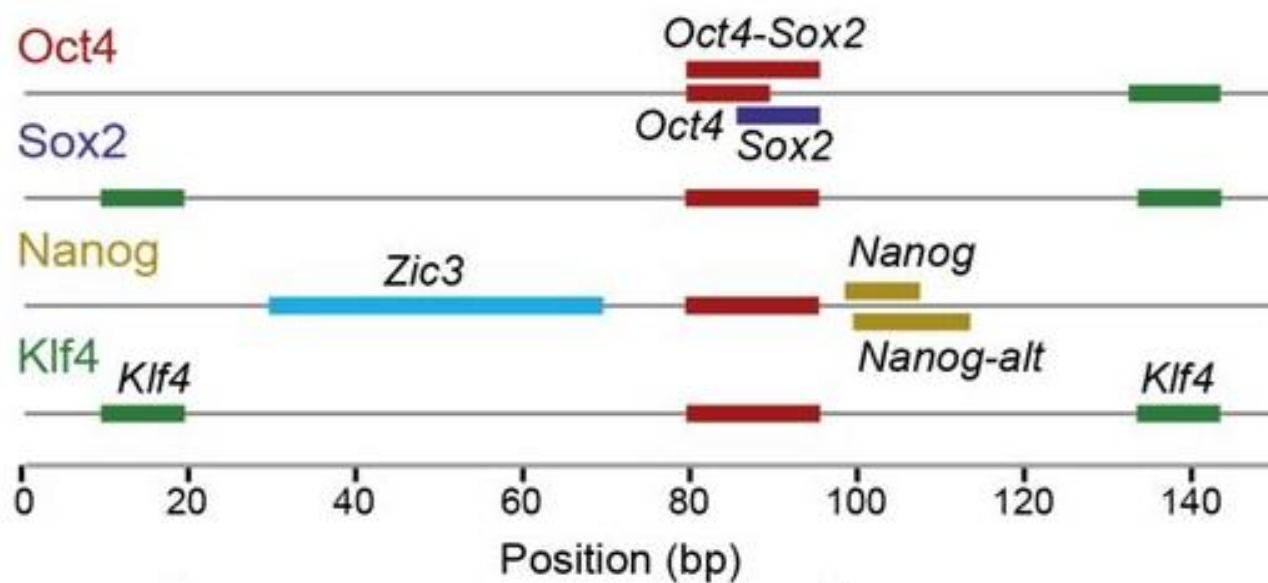
Alex Tseng

Shrikumar et al. 2017 ICML
 Shrikumar et al. 2019 ISMB
 Tseng et al. 2020 NeurIPS
 Greenside et al. 2018, ECCB



Profile contribution scores

DeepLIFT



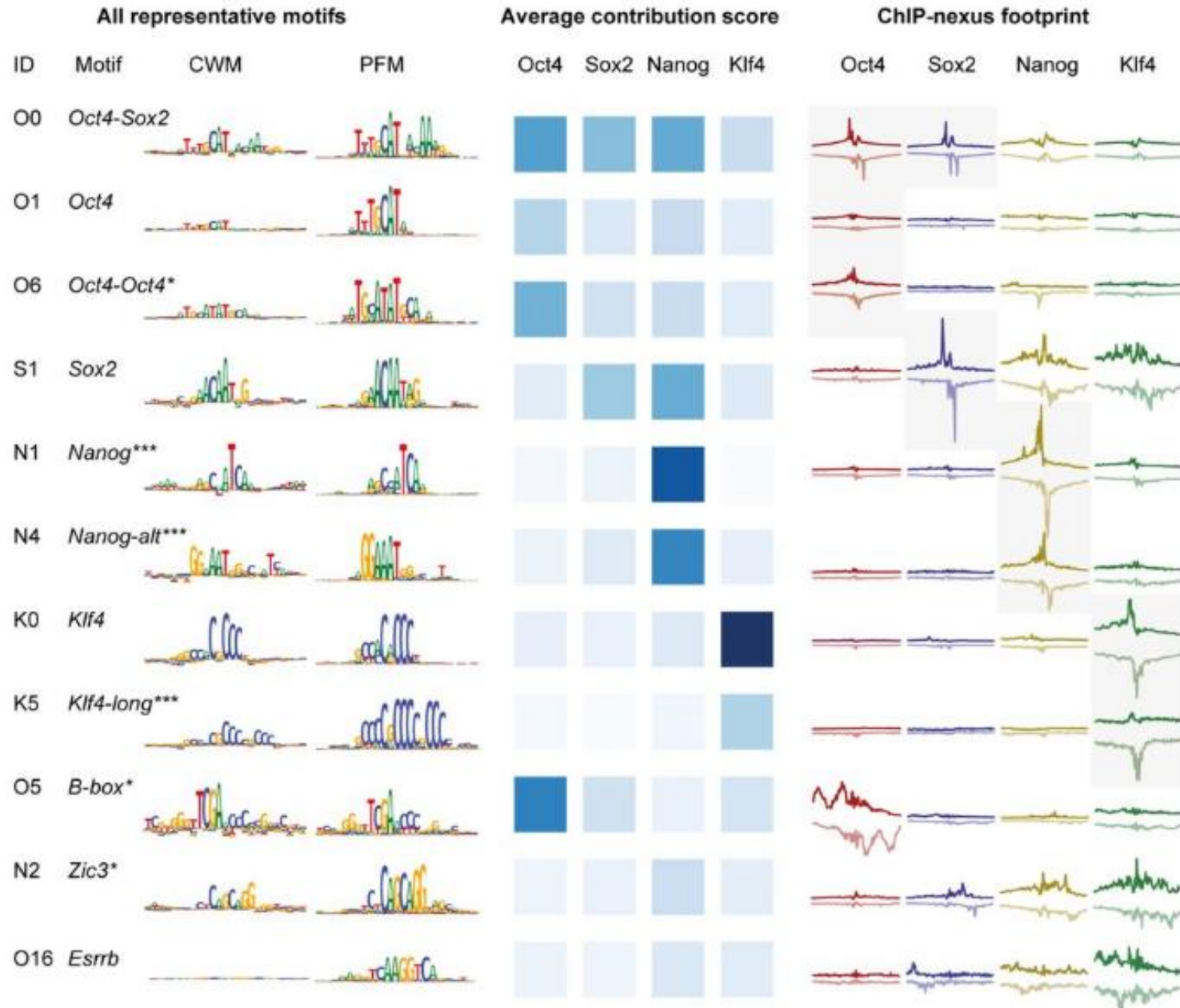
Avanti Shrikumar



Alex Tseng

Shrikumar et al. 2017 ICML
 Shrikumar et al. 2019 ISMB
 Tseng et al. 2020 NeurIPS
 Greenside et al. 2018, ECCB

Complex repertoire of motifs due to cooperative binding



50 motifs for 4 TFs!

Syntax discovery using *in-silico* perturbations

Use BpNet model as in-silico oracle to perform perturbation experiments



- 1) On synthetic sequences

Syntax discovery using *in-silico* perturbations

Use BpNet model as in-silico oracle to perform perturbation experiments



- 1) On synthetic sequences

Syntax discovery using *in-silico* perturbations

Use BpNet model as in-silico oracle to perform perturbation experiments



1) On synthetic sequences



2) By mutating genomic sequences

Syntax discovery using *in-silico* perturbations

Use BpNet model as in-silico oracle to perform perturbation experiments



1) On synthetic sequences



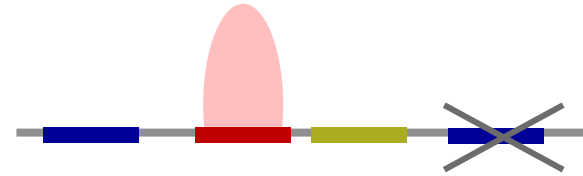
2) By mutating genomic sequences

Syntax discovery using *in-silico* perturbations

Use BpNet model as in-silico oracle to perform perturbation experiments



1) On synthetic sequences



2) By mutating genomic sequences

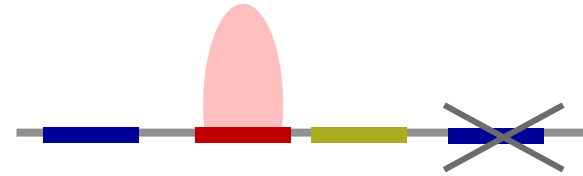
Syntax discovery using *in-silico* perturbations

Use BpNet model as in-silico oracle to perform perturbation experiments



1) On synthetic sequences

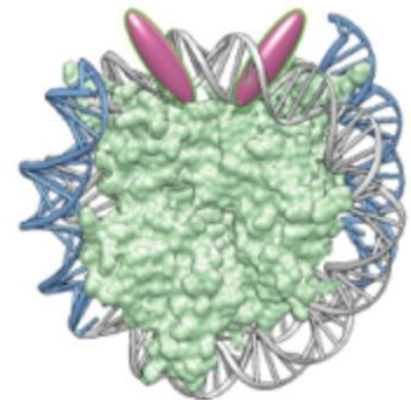
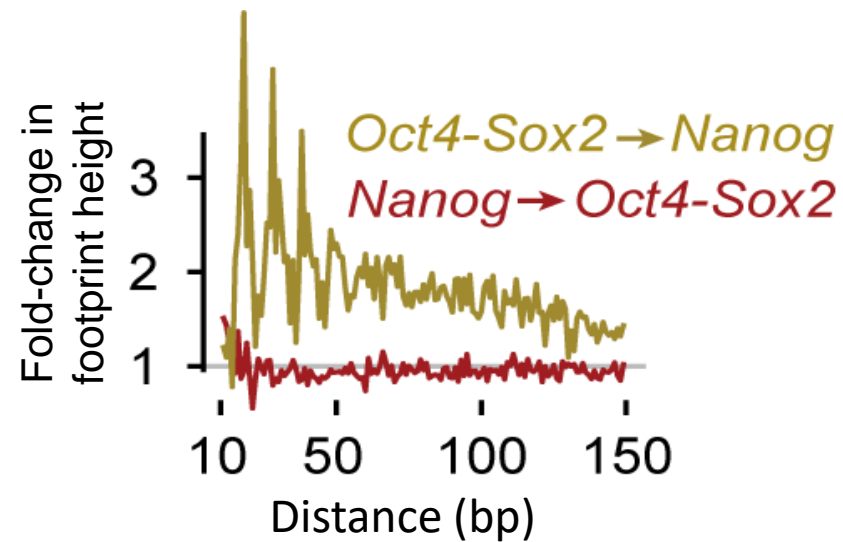
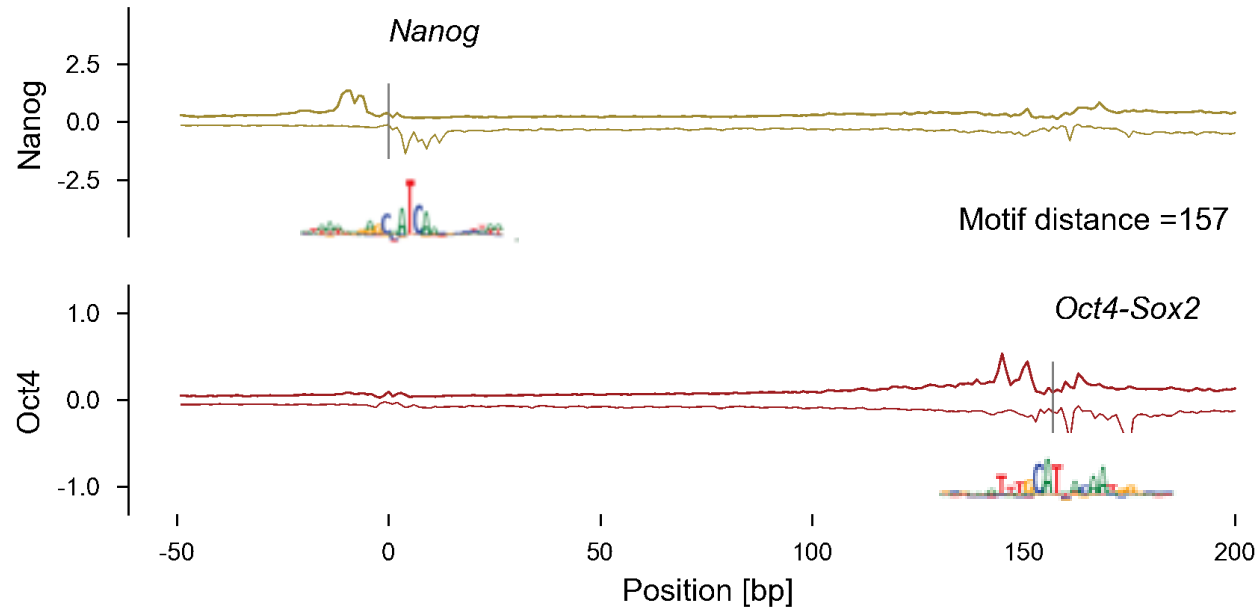
In silico biochemistry



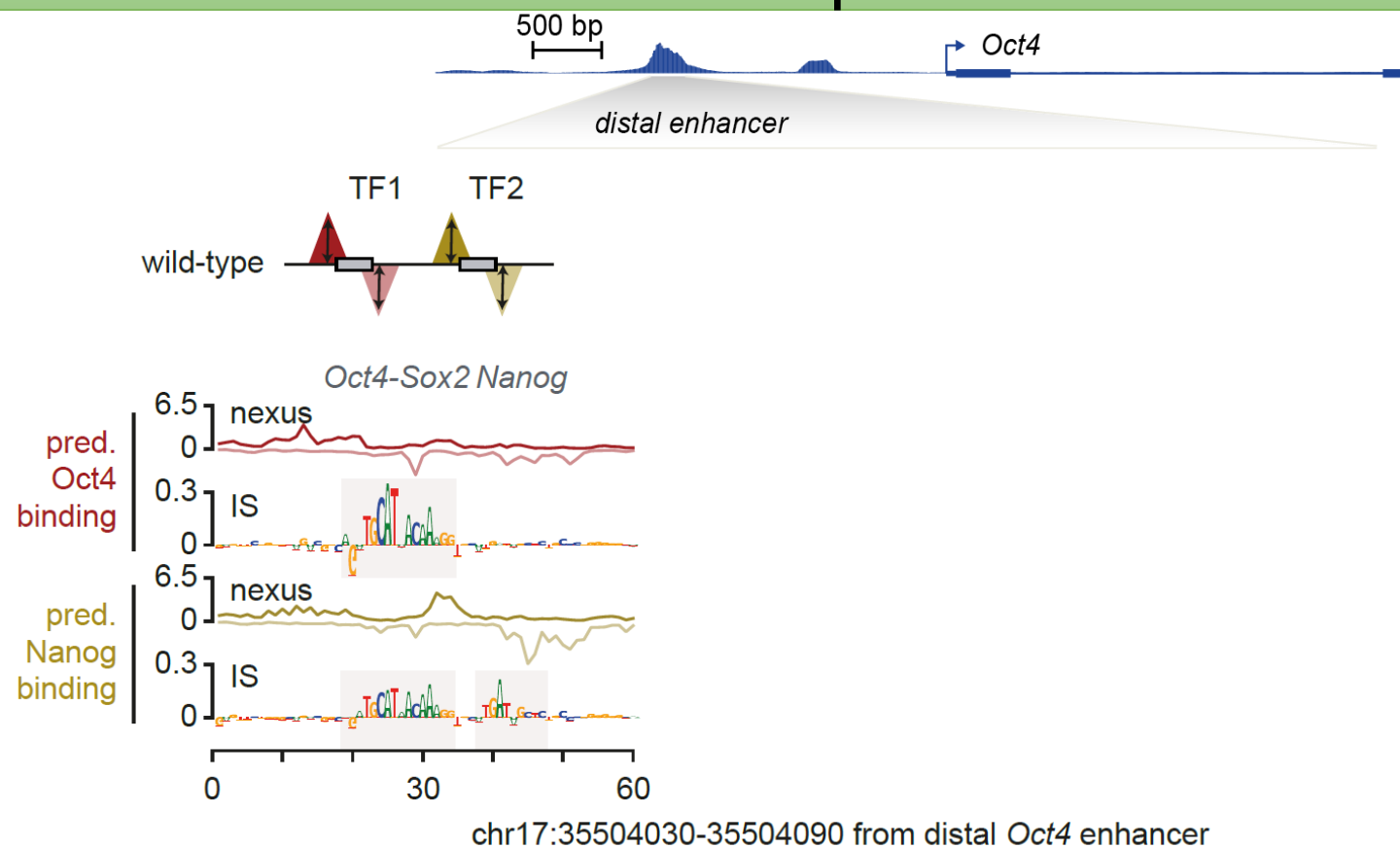
2) By mutating genomic sequences

In silico genetics

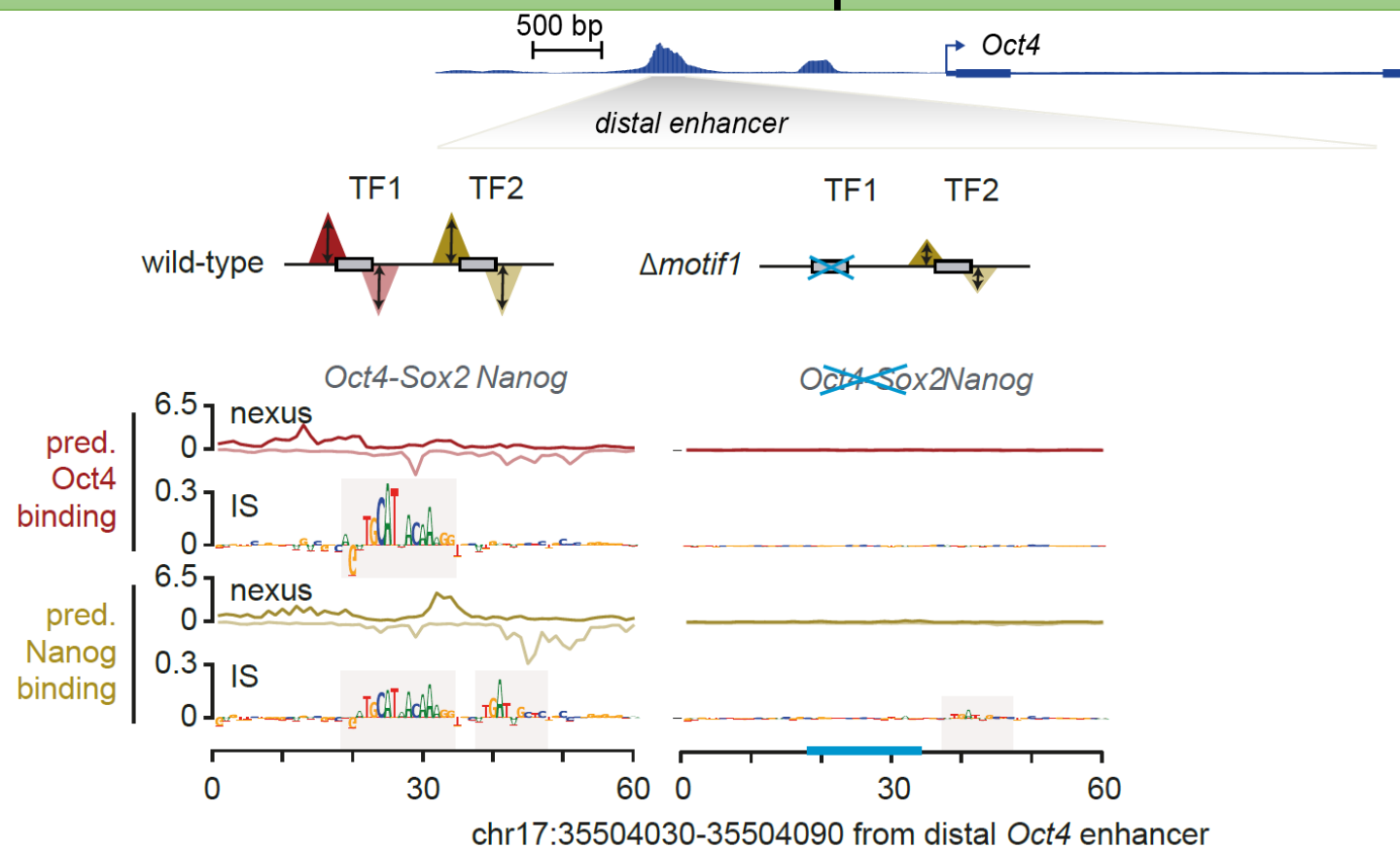
In-silico reporters: Designing synthetic sequences to query models to reveal syntax



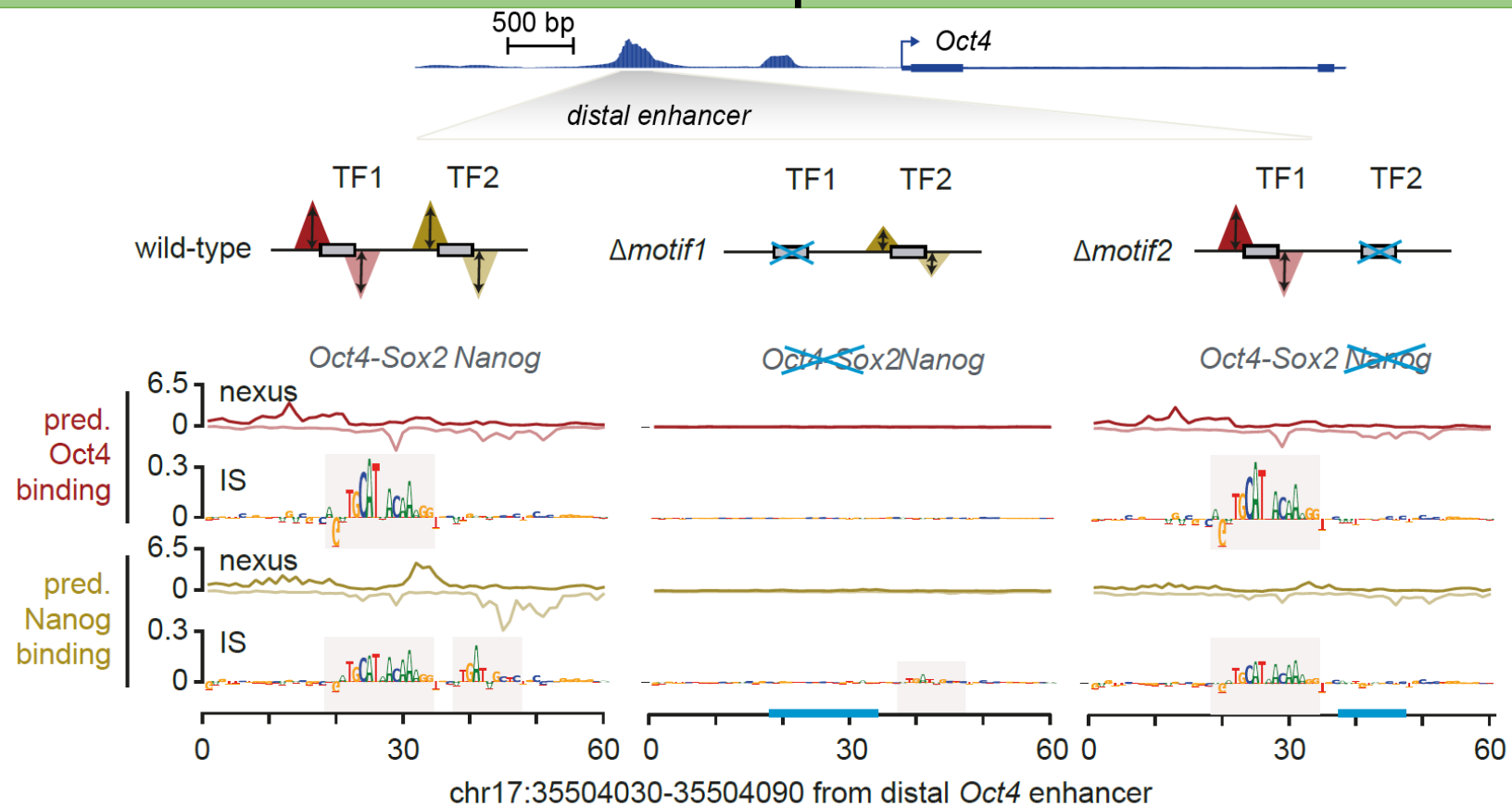
in-silico genome editing: Deciphering syntax by perturbing genomic sequences



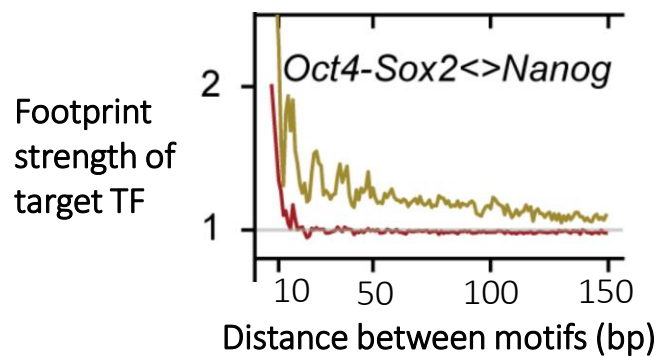
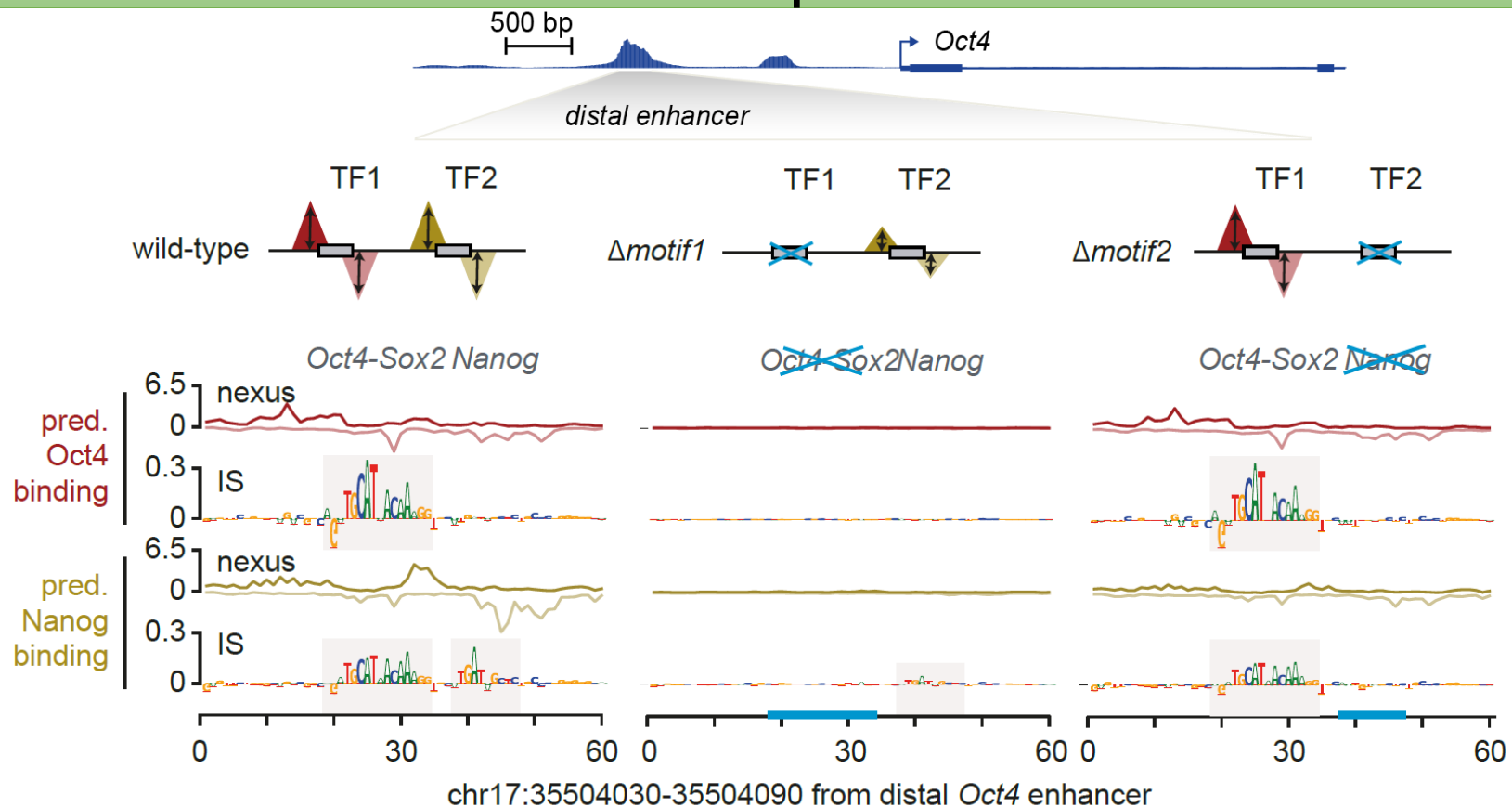
in-silico genome editing: Deciphering syntax by perturbing genomic sequences



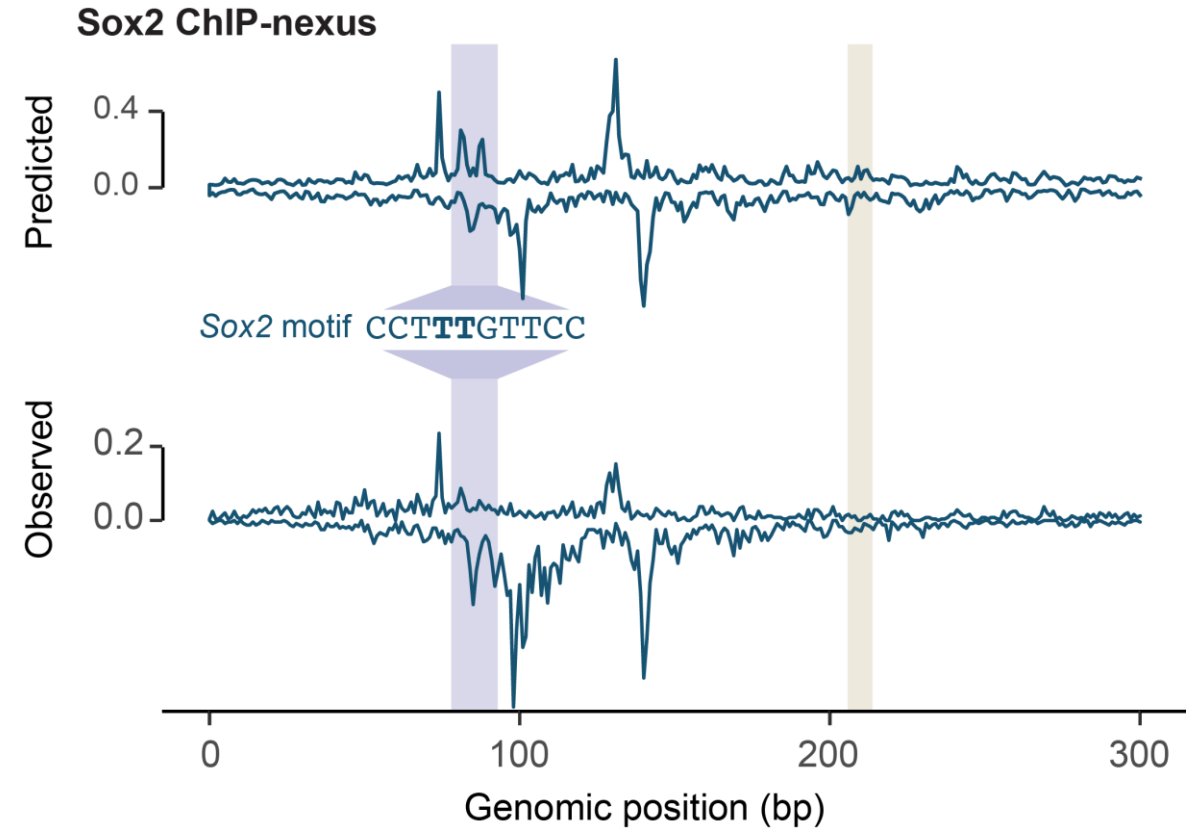
in-silico genome editing: Deciphering syntax by perturbing genomic sequences



in-silico genome editing: Deciphering syntax by perturbing genomic sequences

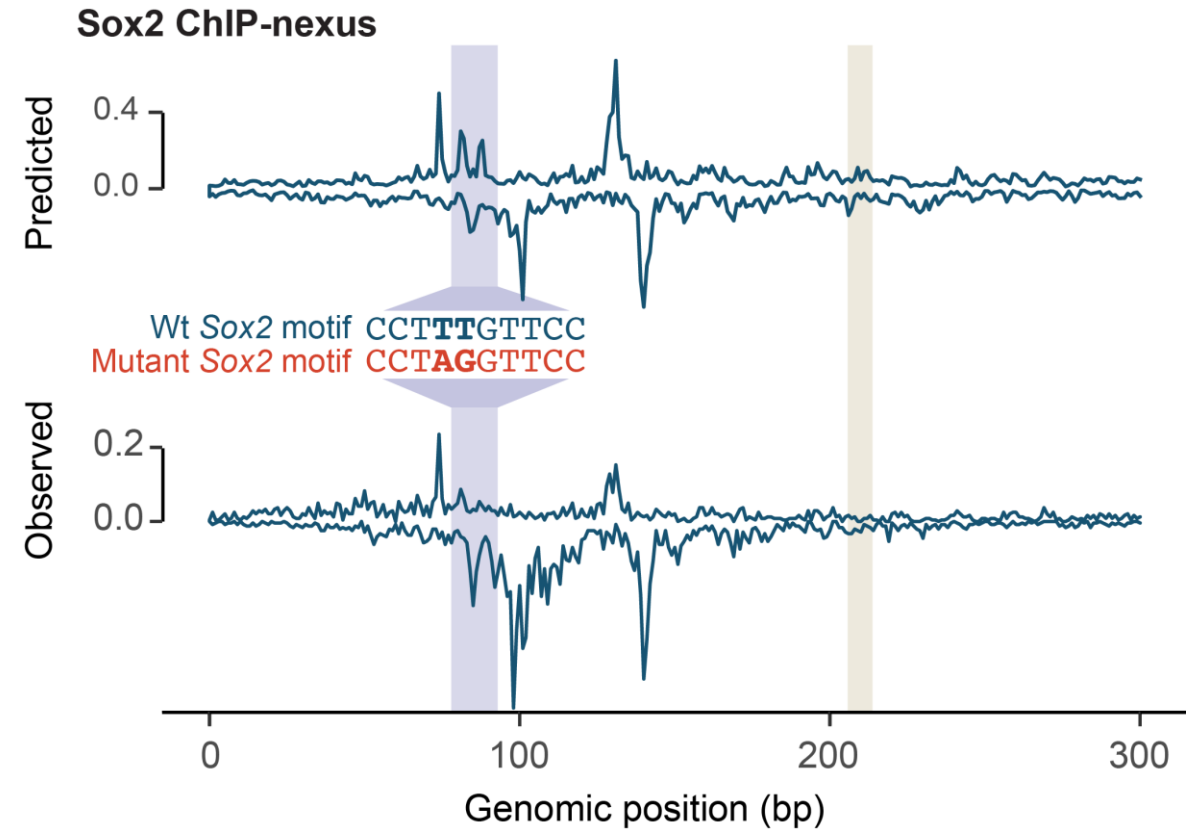


Designing CRISPR experiments to validate motif syntax

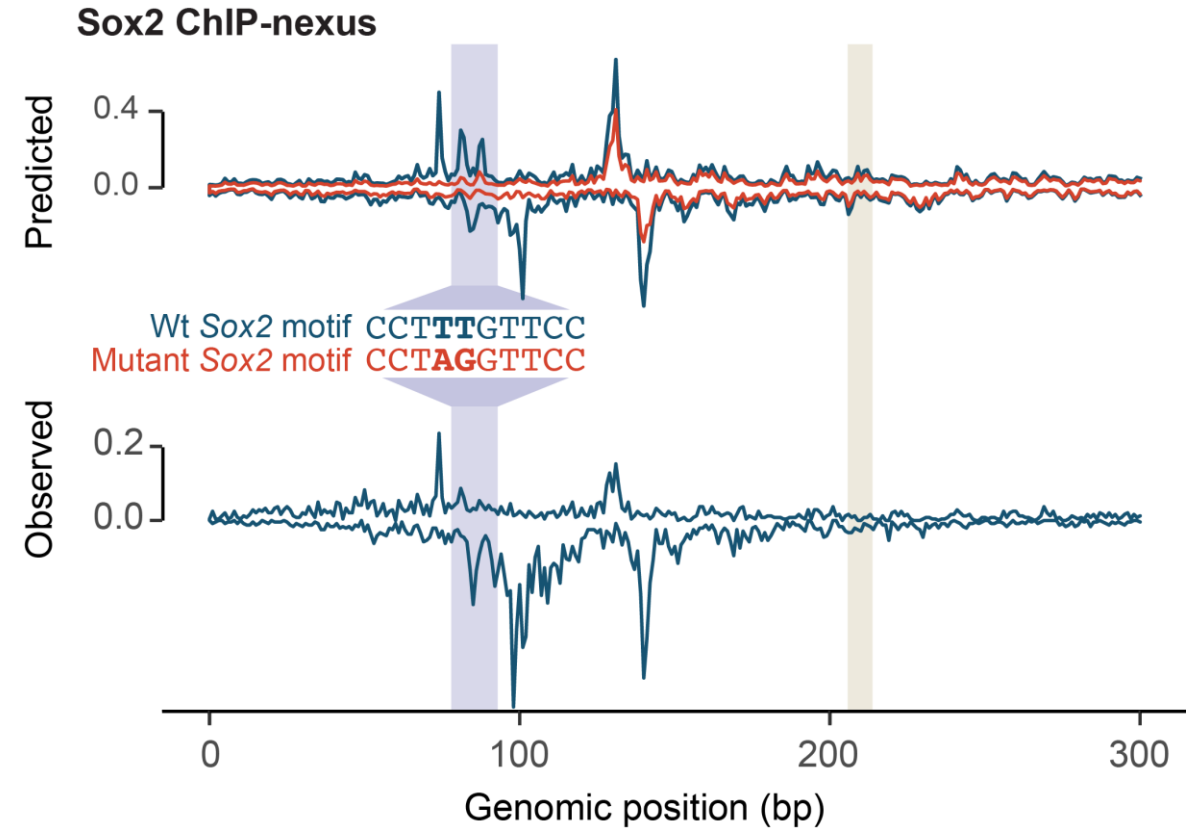


Julia Zeitlinger, Sabrina Krueger, Melanie Weilert

Designing CRISPR experiments to validate motif syntax

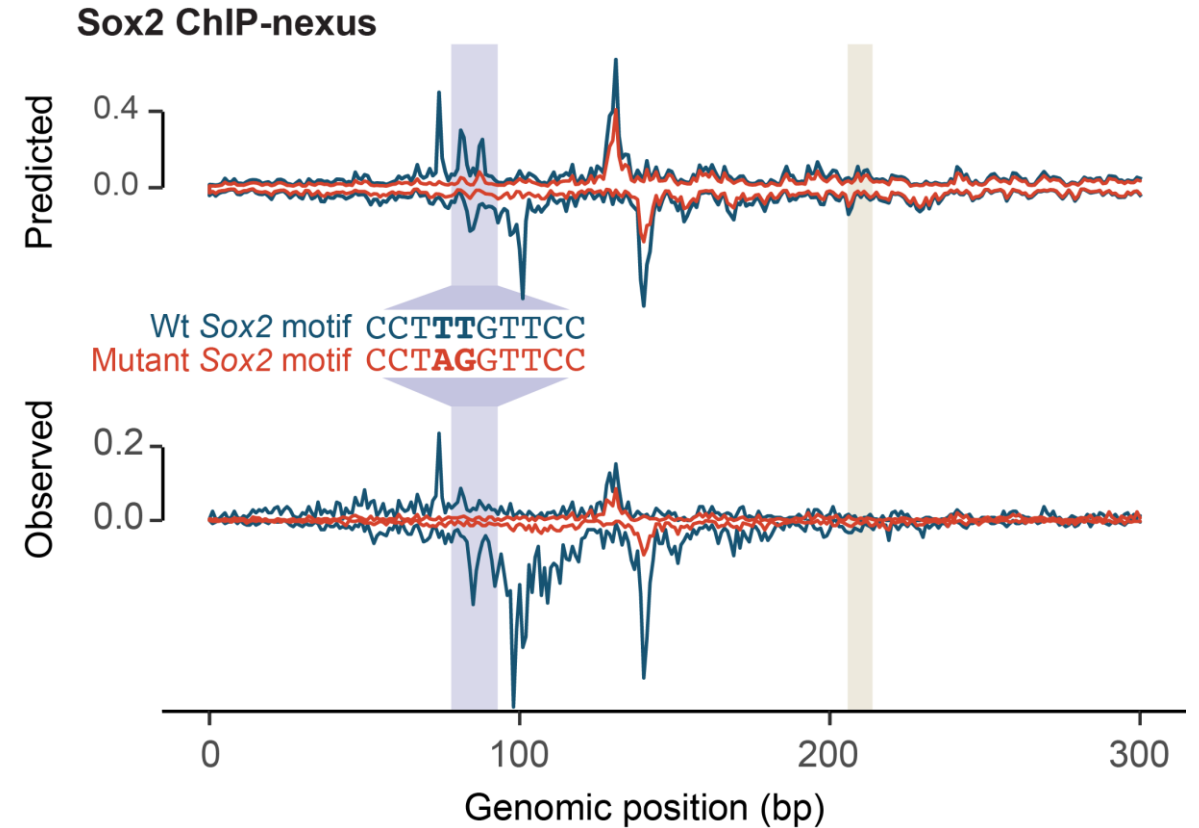


Designing CRISPR experiments to validate motif syntax



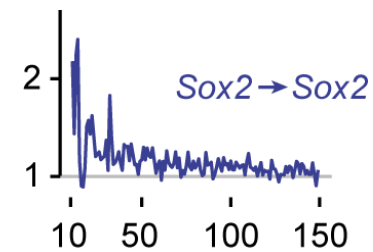
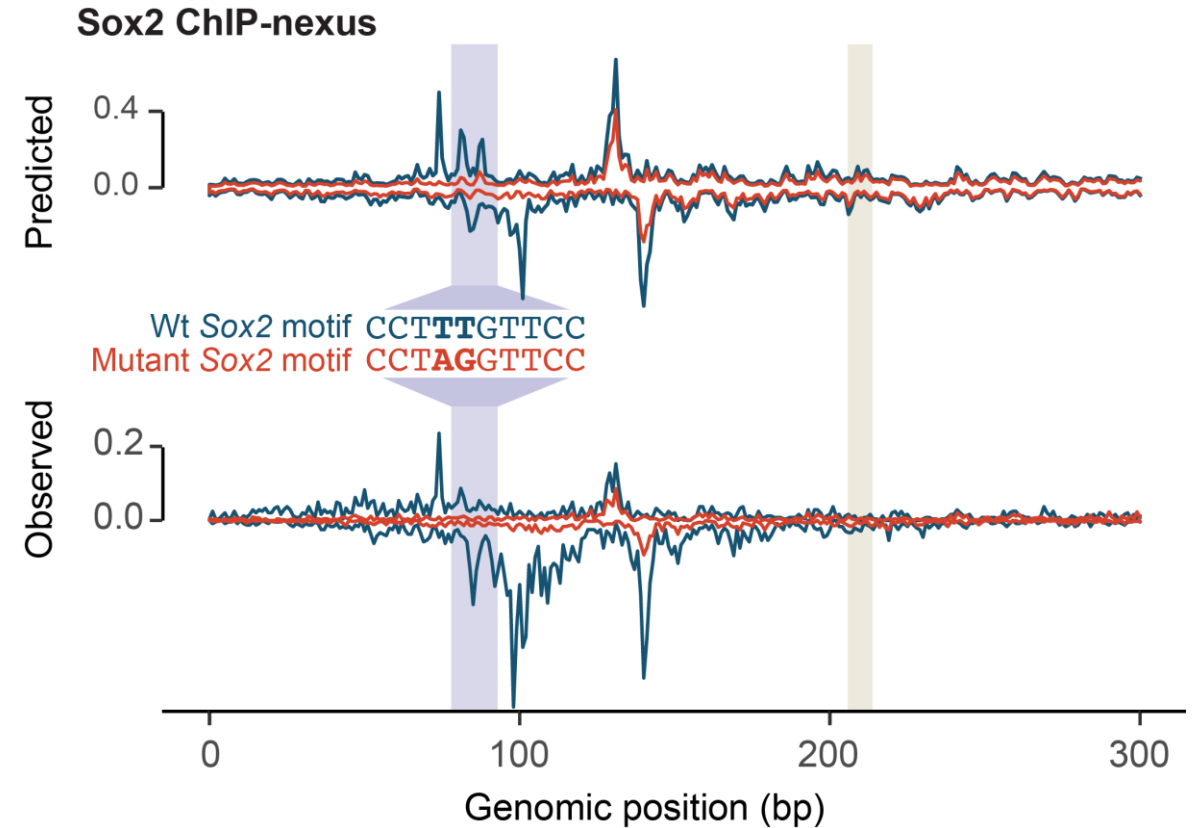
Julia Zeitlinger, Sabrina Krueger, Melanie Weilert

Designing CRISPR experiments to validate motif syntax

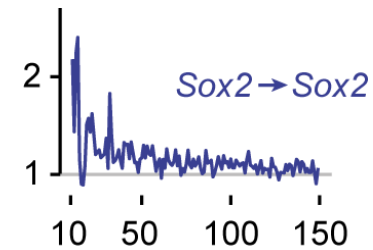
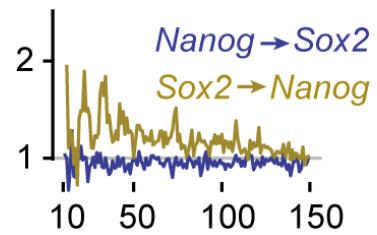
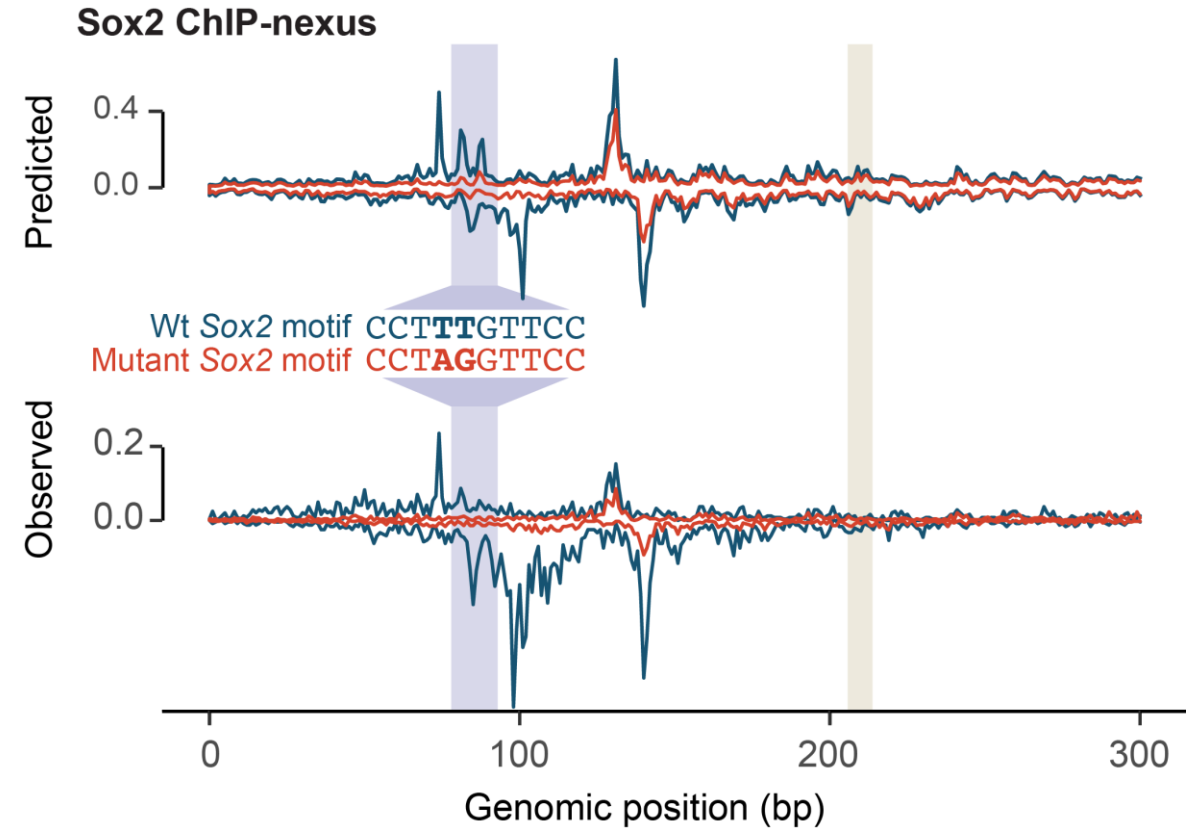


Julia Zeitlinger, Sabrina Krueger, Melanie Weilert

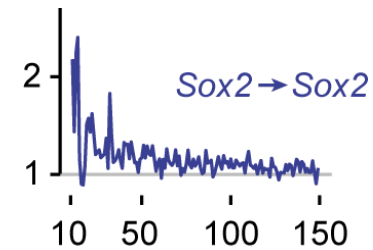
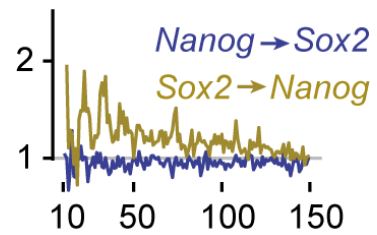
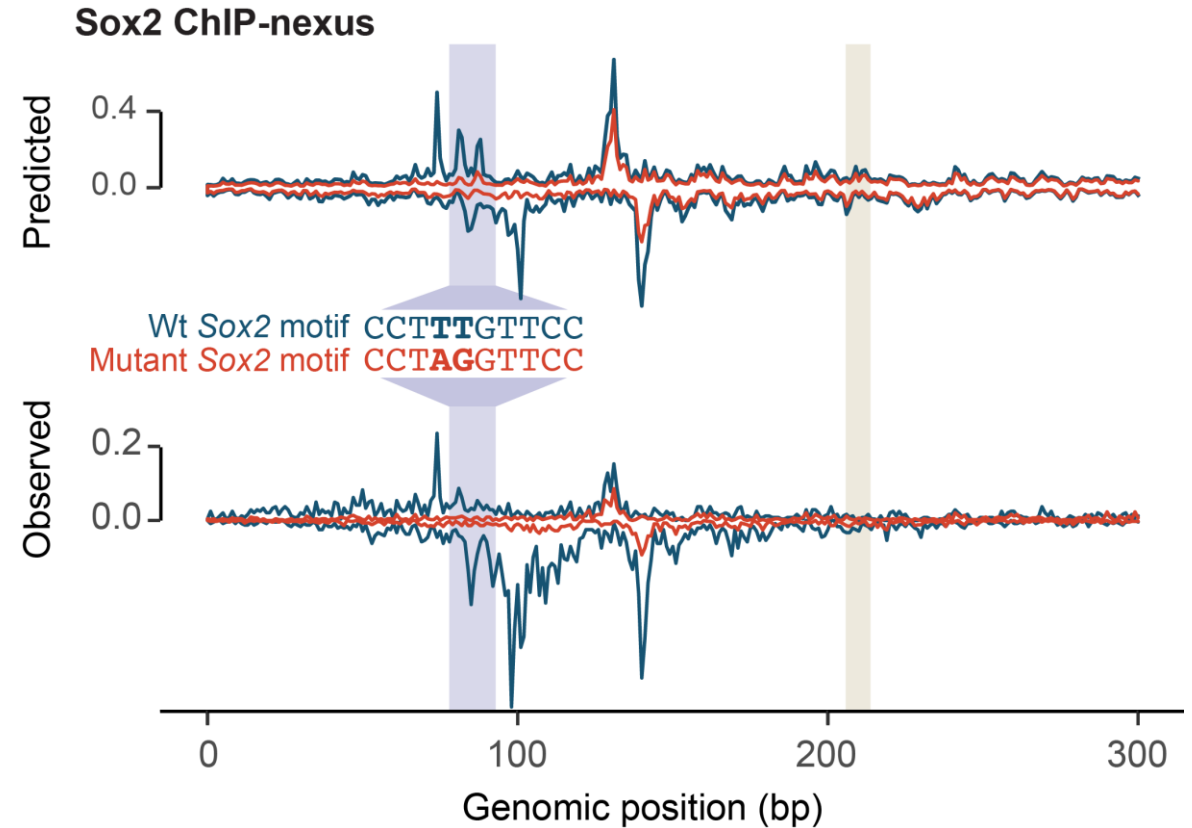
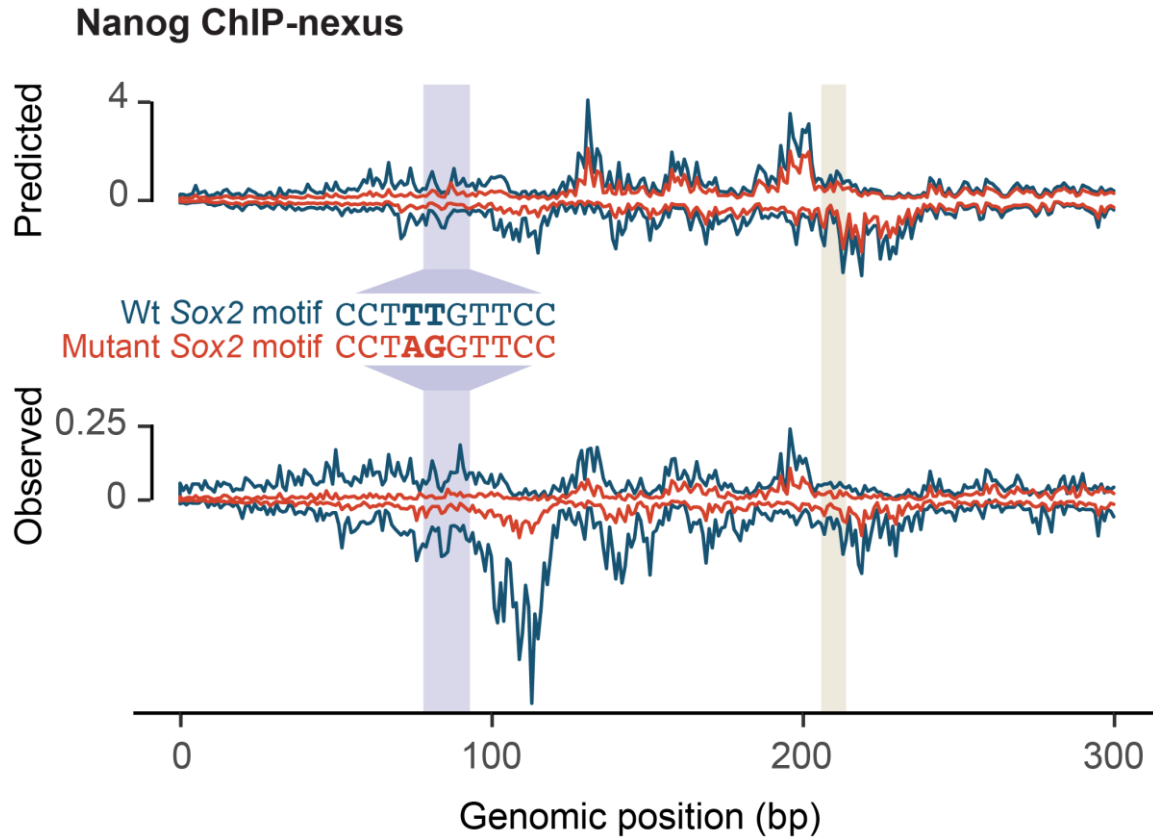
Designing CRISPR experiments to validate motif syntax



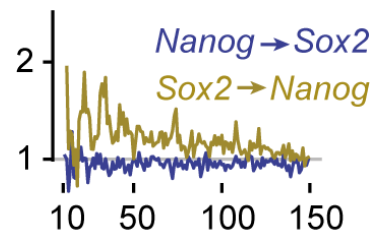
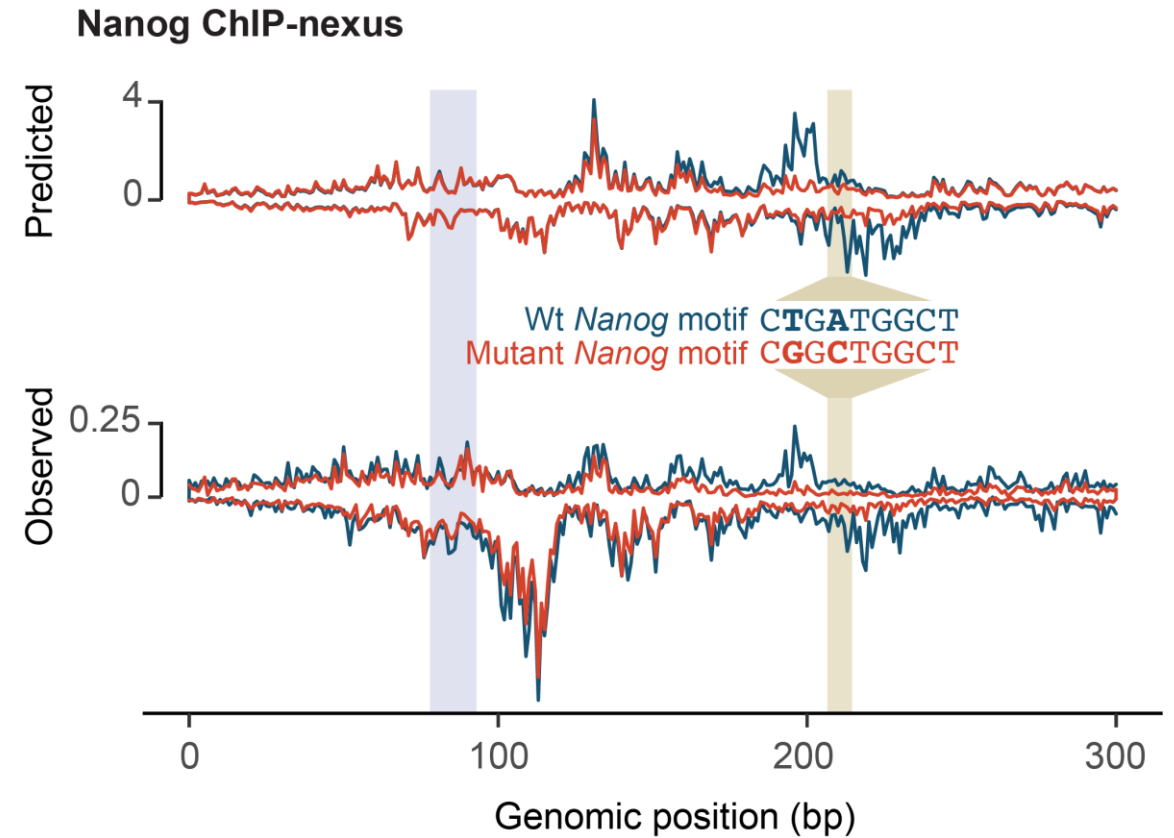
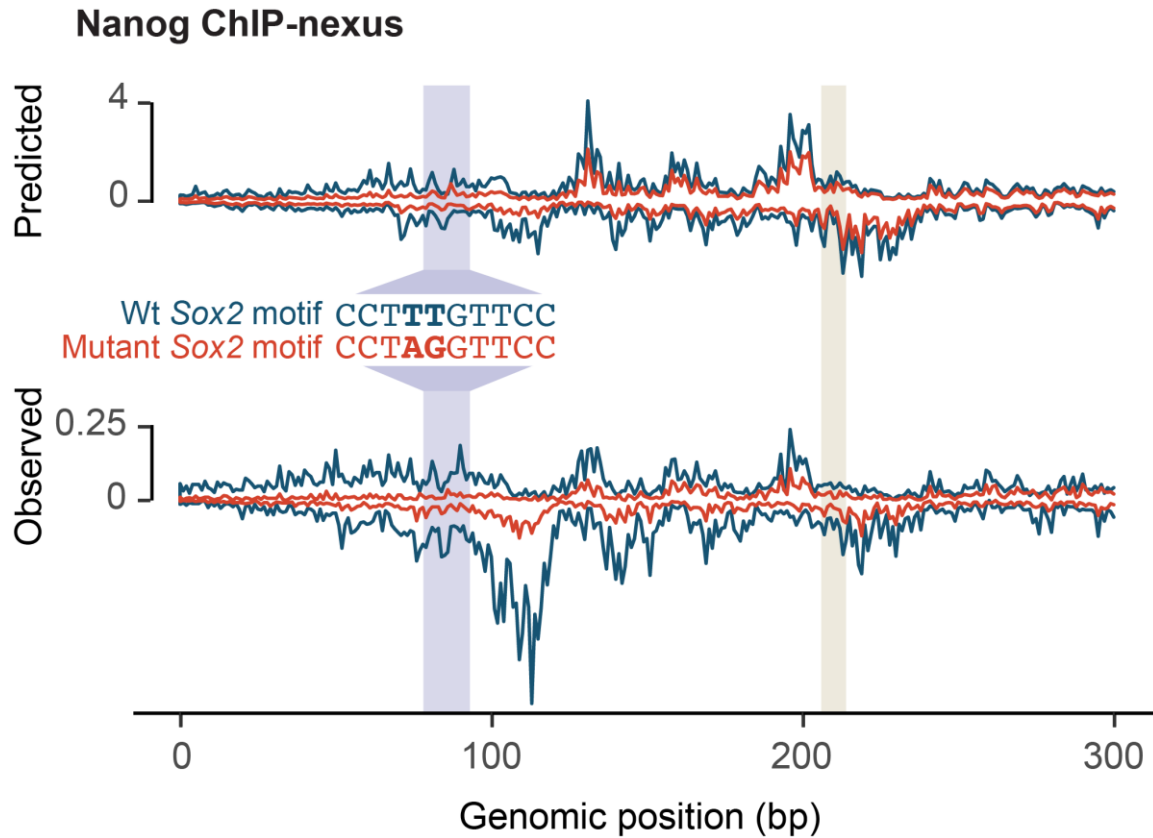
Designing CRISPR experiments to validate motif syntax



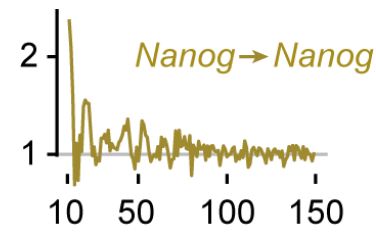
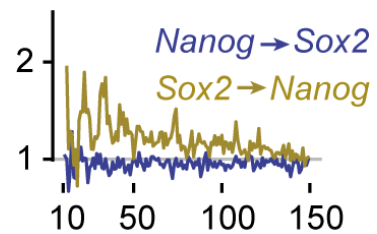
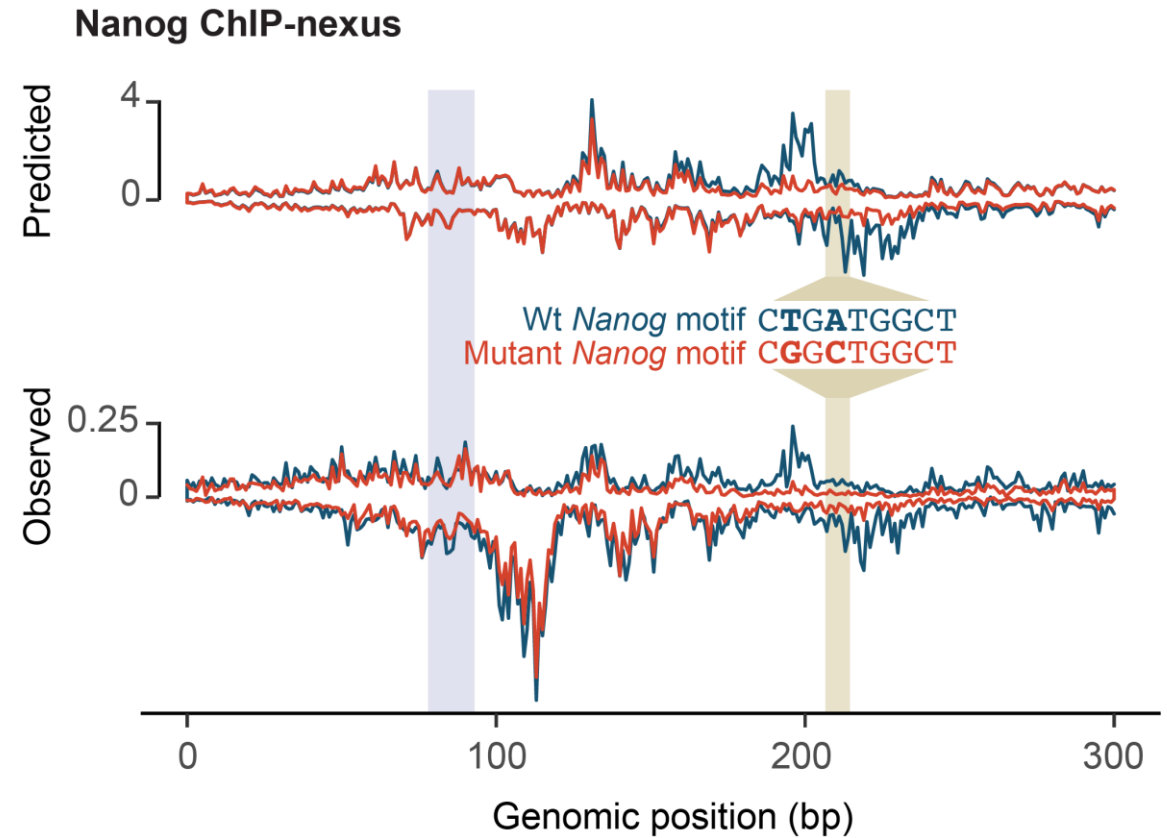
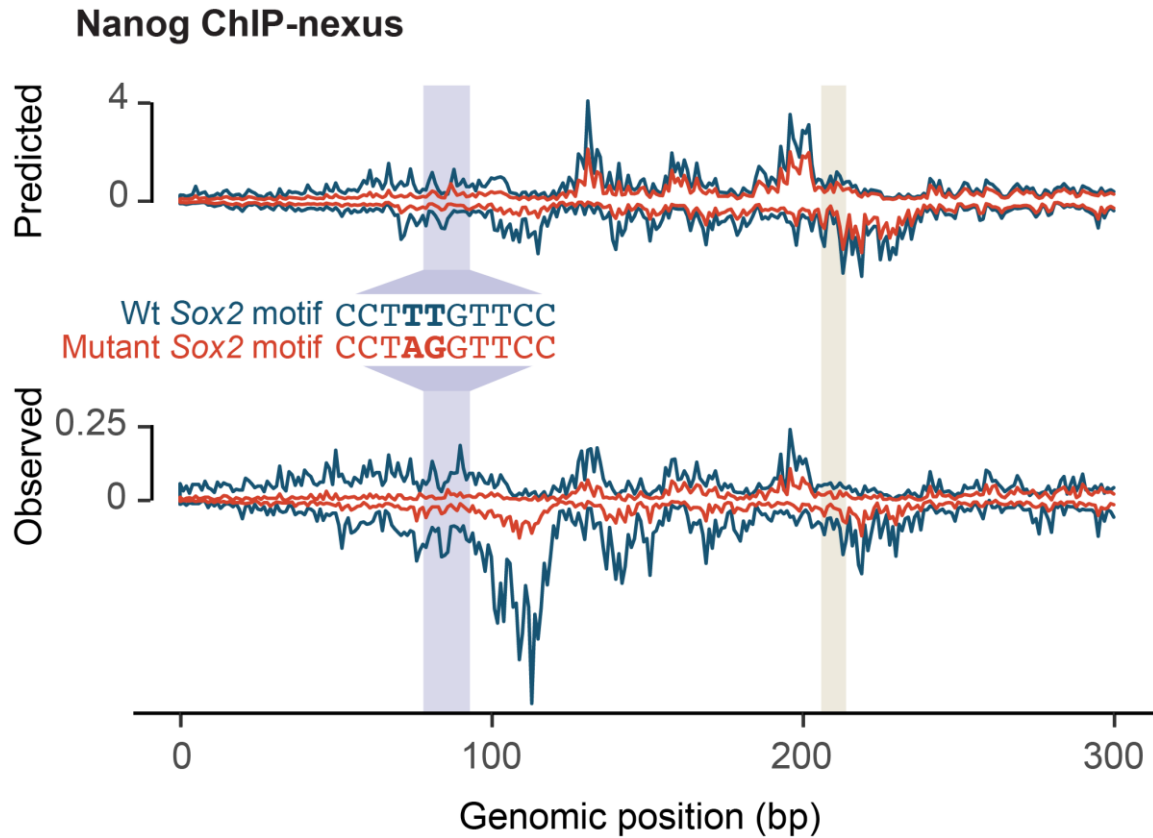
Designing CRISPR experiments to validate motif syntax



Designing CRISPR experiments to validate motif syntax

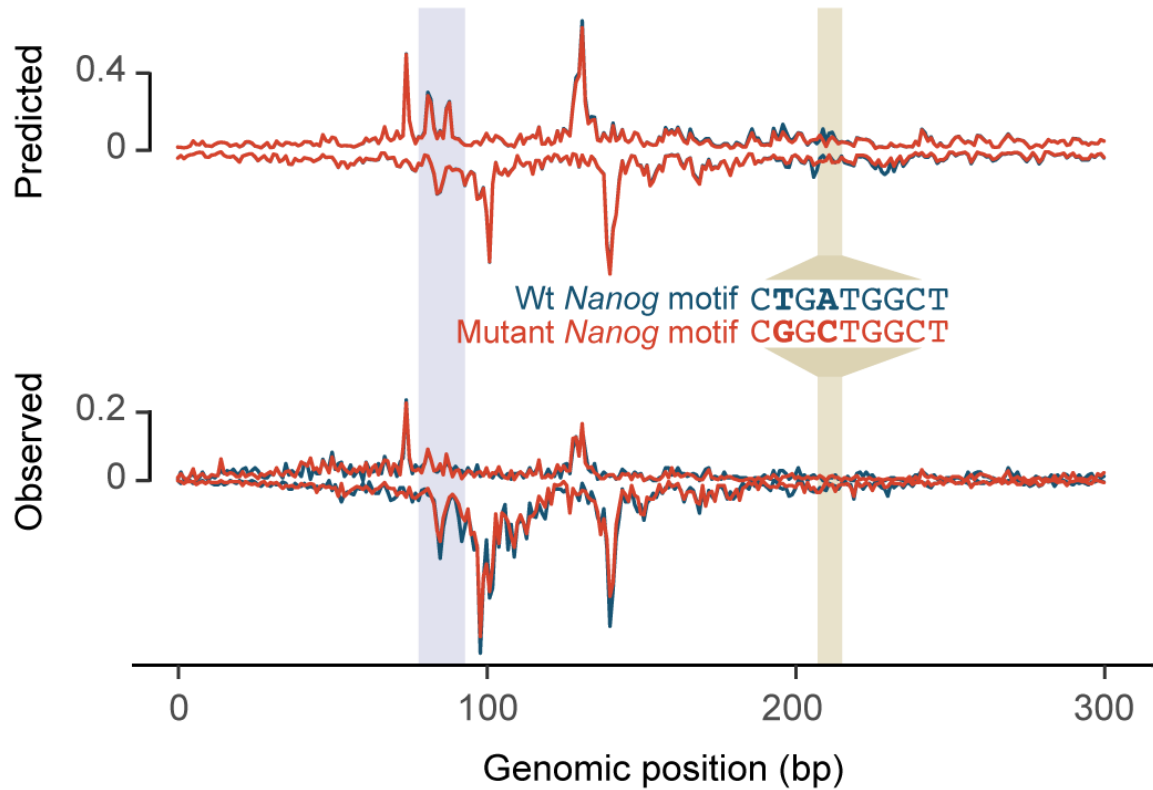


Designing CRISPR experiments to validate motif syntax

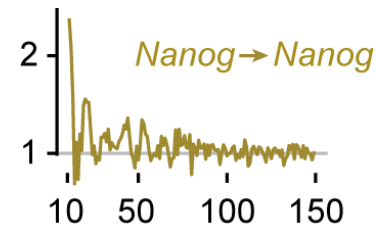
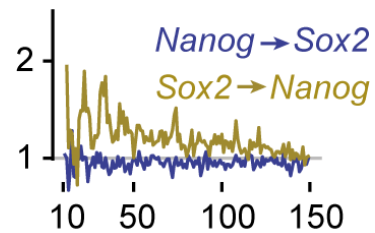
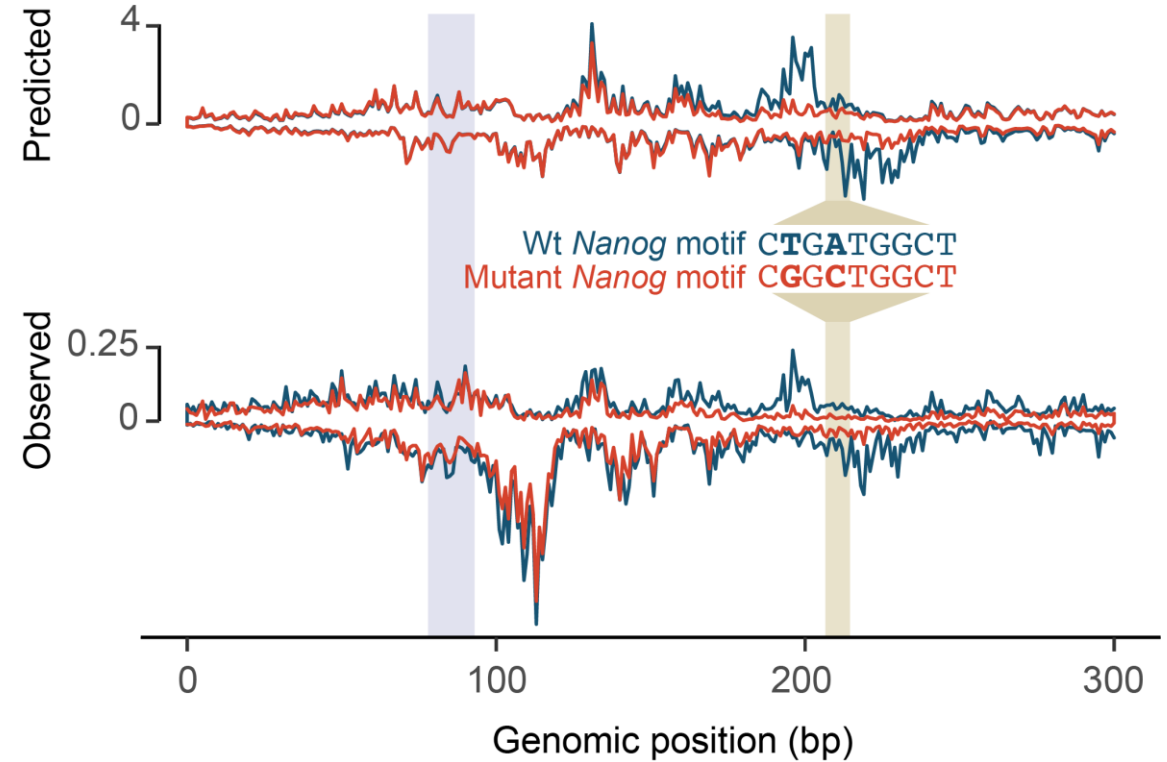


Designing CRISPR experiments to validate motif syntax

Sox2 ChIP-nexus



Nanog ChIP-nexus



Deciphering genetic variants



Soumya Kundu



Lakshman Sundaram



Ryan Corces



Howard Chang



Tom Montine



Mo Ameen



Tom
Quertermous



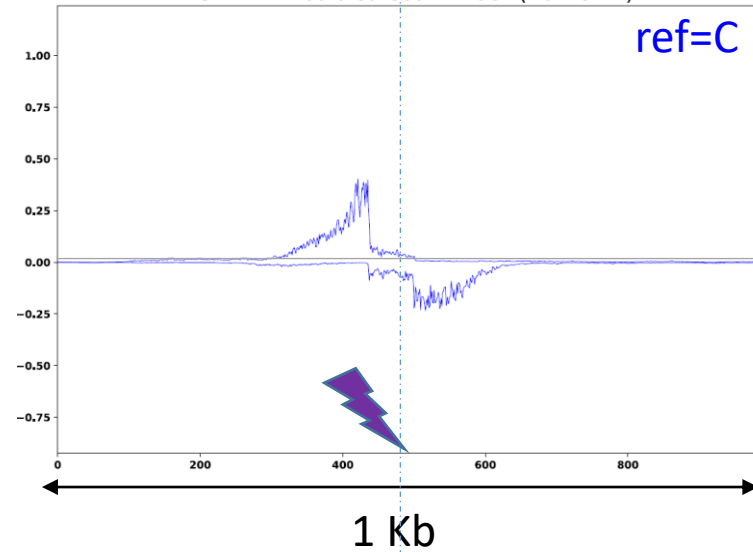
Will Greenleaf



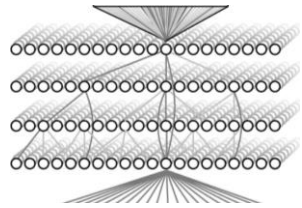
Sergiu Pasca

In-silico mutagenesis: Predict effect of genetic variant on molecular activity

Predicted molecular profile of protein-DNA binding



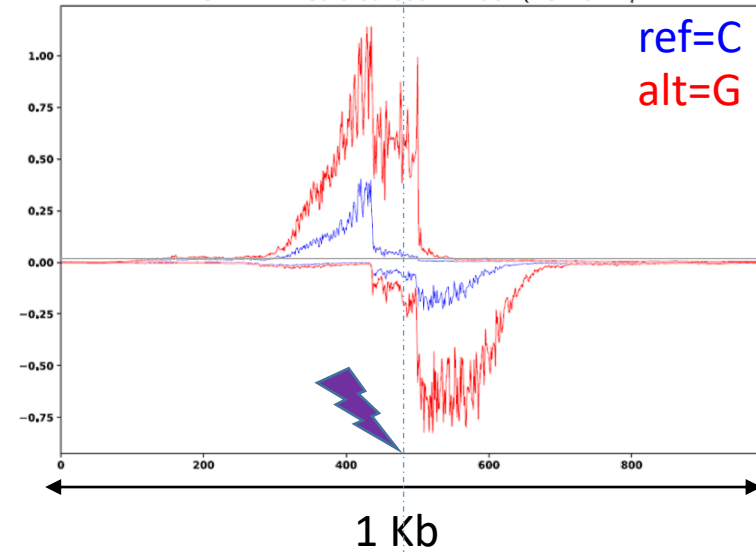
PredictedSignal



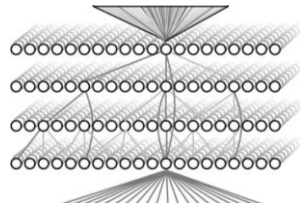
.....ACTGAT **C**GCAATCG.....

In-silico mutagenesis: Predict effect of genetic variant on molecular activity

Predicted molecular profile of protein-DNA binding



Δ PredictedSignal

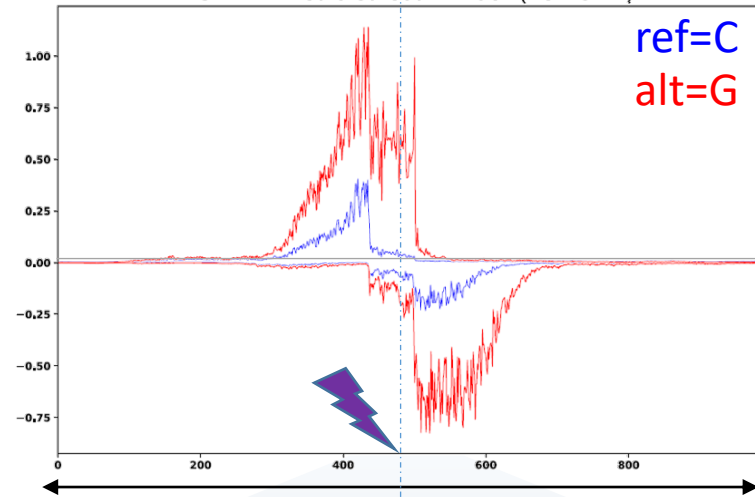


.....ACTGAT **C** GCAATCG.....

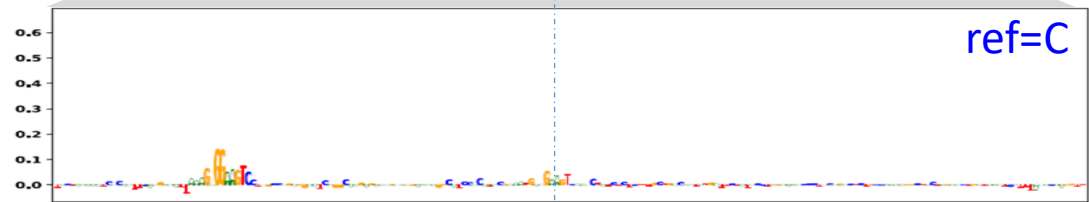
.....ACTGAT **G** GCAATCG.....

Interpret disrupted predictive sequence syntax

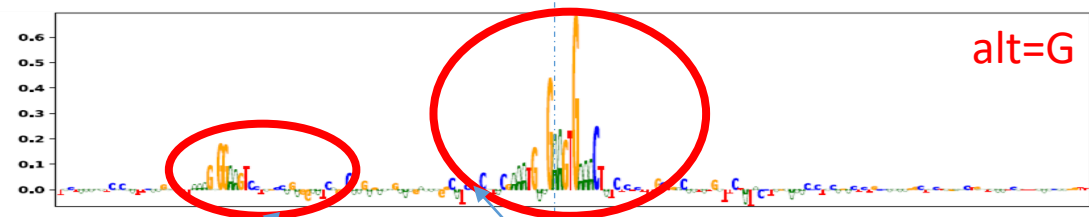
Predicted molecular profile of protein-DNA binding



1 Kb



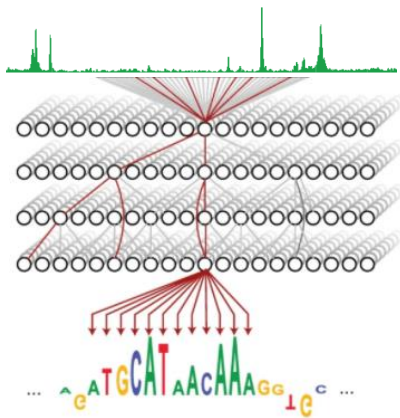
ref=C



alt=G

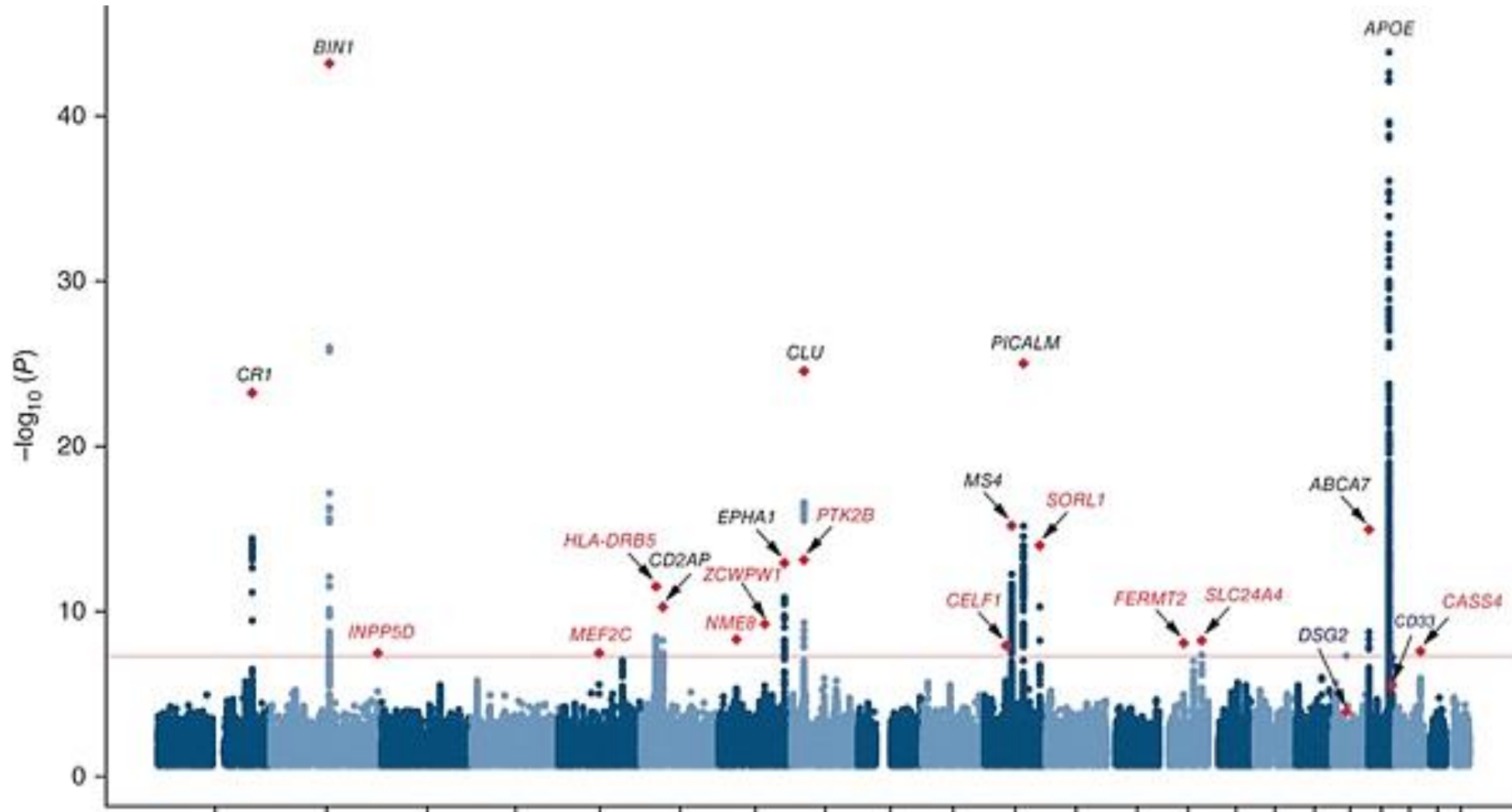
200 bp

Sequence binding motifs of SPI1 DNA binding protein



Genetic variants associated with Alzheimer's disease

Statistical significance of association



(Lambert et al., Nat. Genet., 2013)

Genomic position →

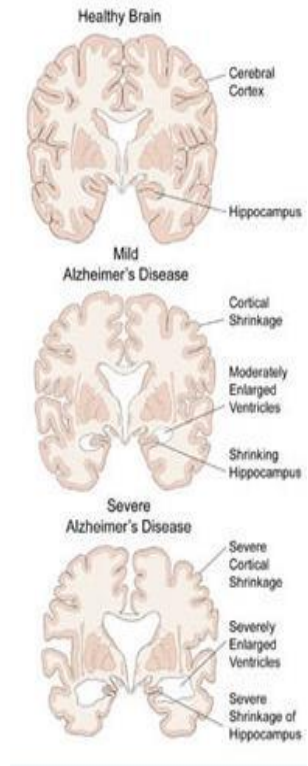
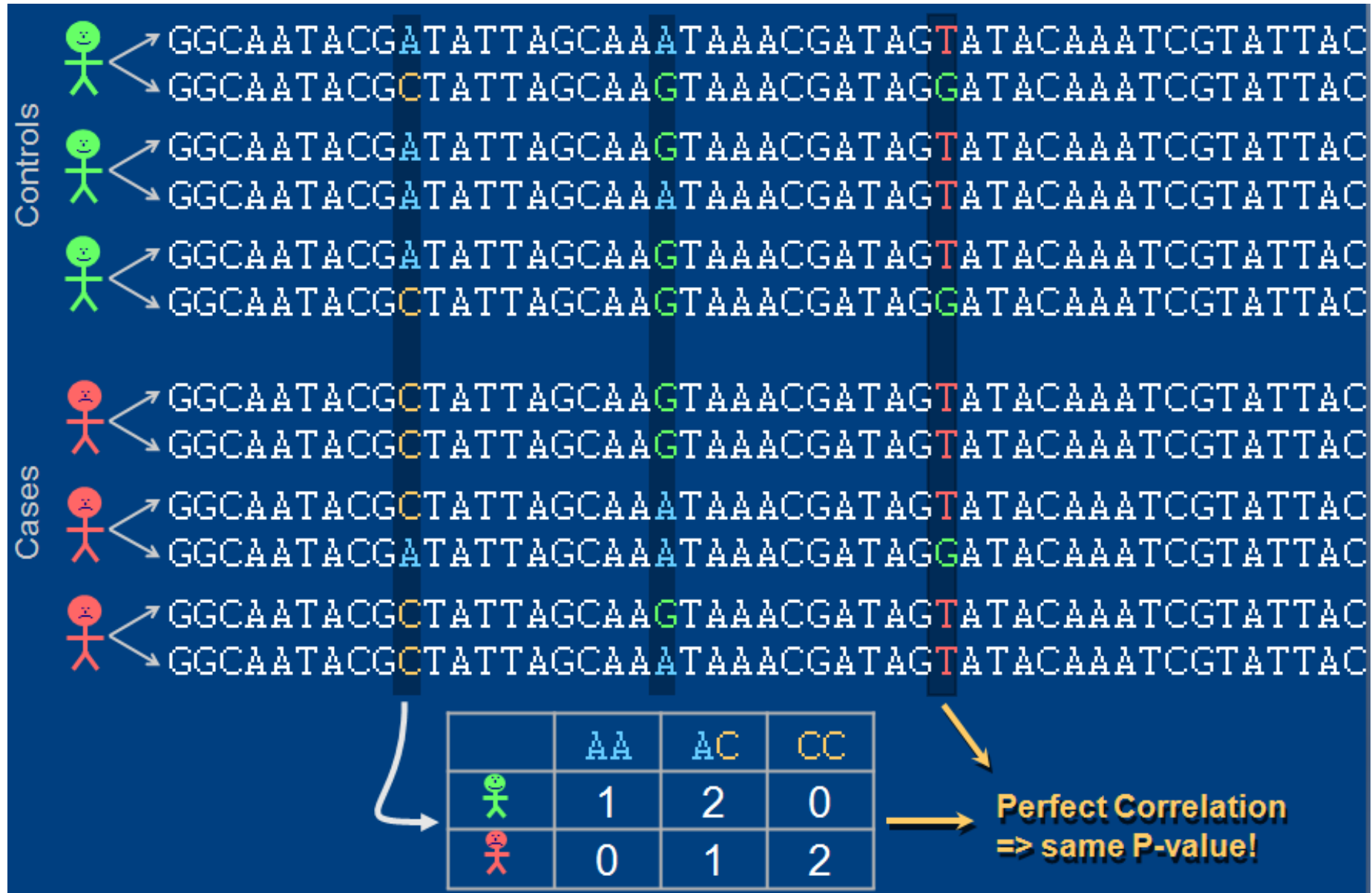
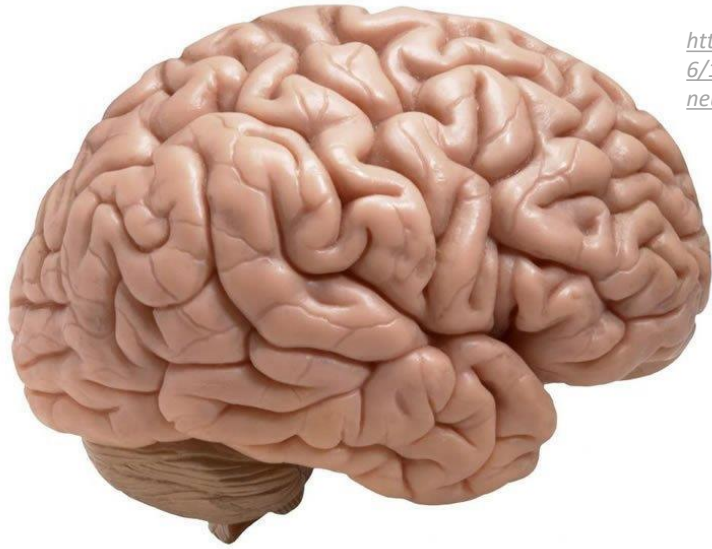


Illustration by Bob Morreale, American Health Assistance Foundation

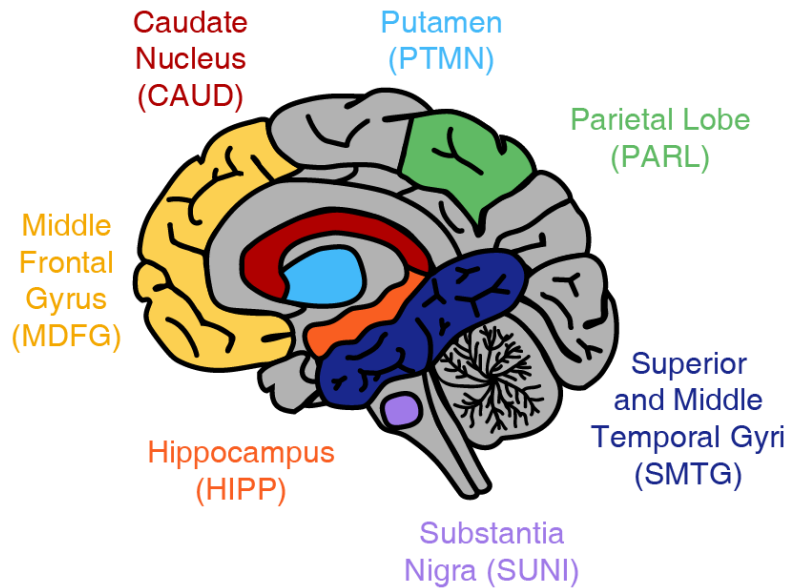
Problem: Which variant in the disease-associated locus is causal & what does it do?



The brain is a complex tissue with many different cell types

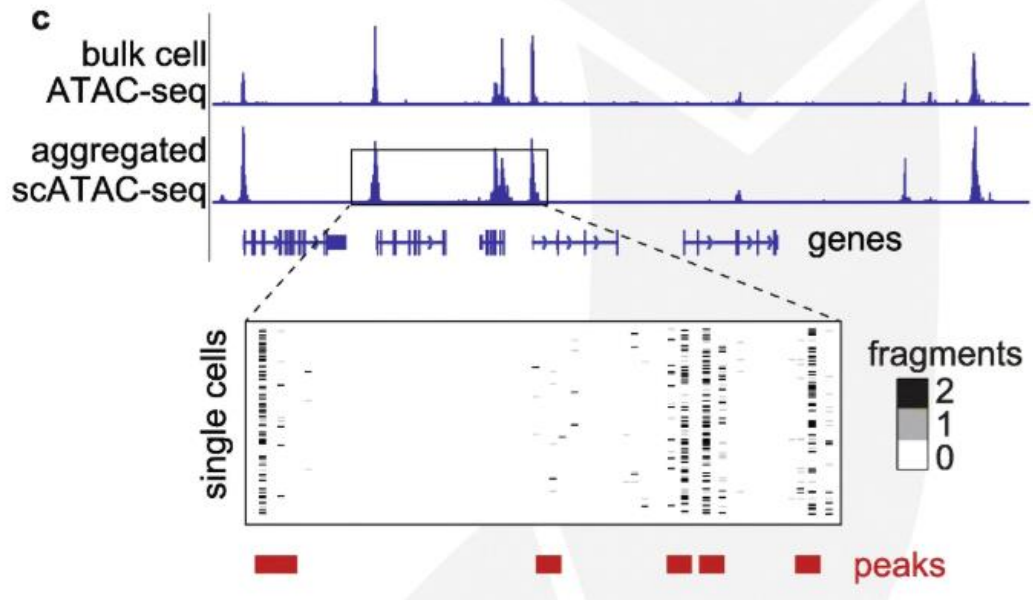


<https://i0.wp.com/neurosciencenews.com/files/2016/11/neurodegeneration-brain-neurosciencenews.jpg>



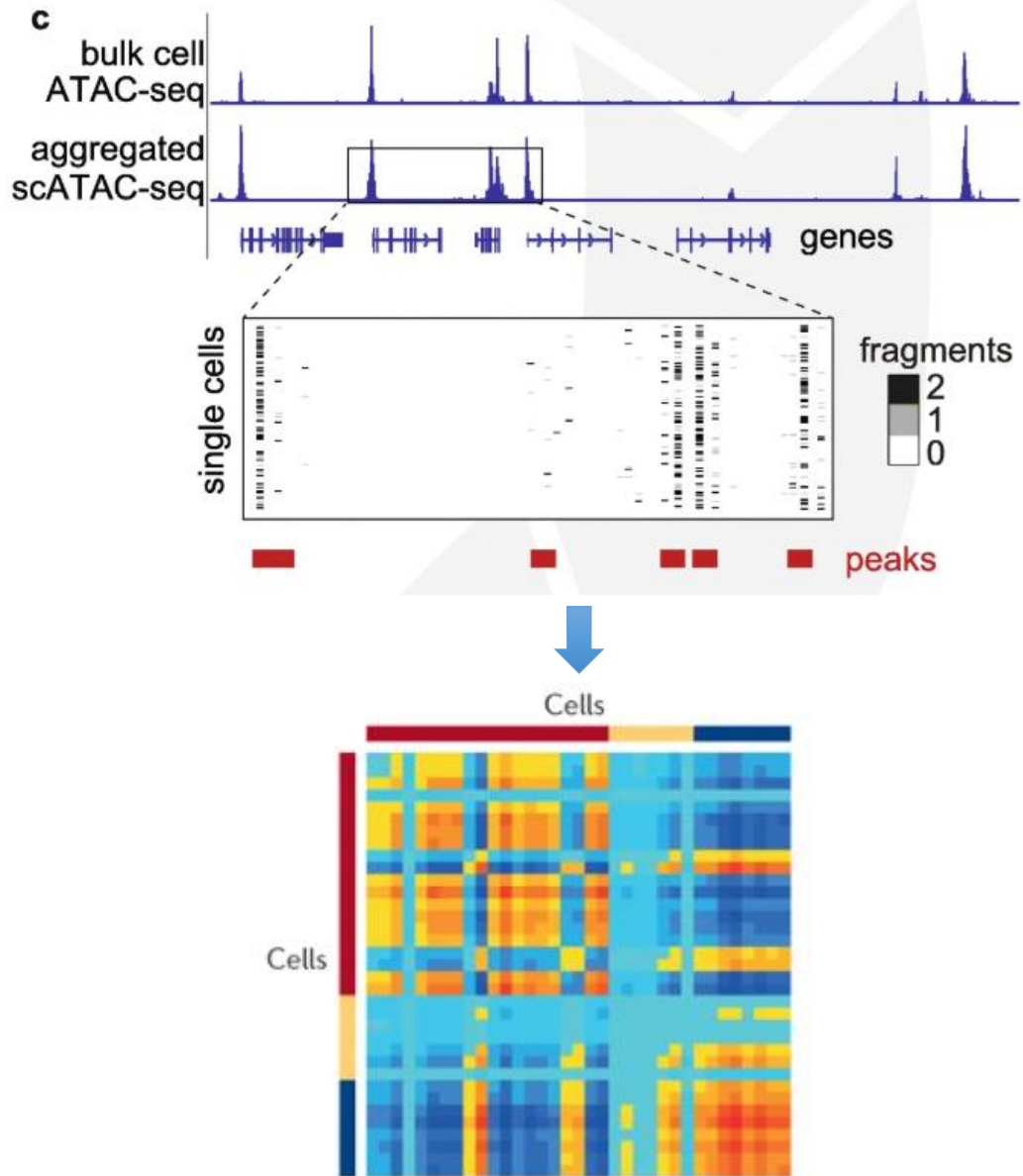
Healthy	Alzheimer's Disease
<p>Neurons</p>	<ul style="list-style-type: none"> - Aβ plaque buildup - Neurofibrillary tau tangles - Synapse loss - Excitotoxicity - Axonal/dendritic injury - Degeneration/cell death - Release of pro-inflammatory molecules - Oxidative stress
<p>Astrocytes</p>	<ul style="list-style-type: none"> - Proliferation - Activated morphology - Tight-junction dysfunction - Inflammatory cytokine release - Altered trafficking and degradation pathways
<p>Microglia</p>	<ul style="list-style-type: none"> - Proliferation - Activated morphology - Inflammatory cytokine release - Dysregulated immune response - Altered trafficking and degradation pathways
<p>Oligodendrocytes</p>	<ul style="list-style-type: none"> - Demyelination - Degeneration/cell death - Oxidative stress
<p>Pericytes</p>	<ul style="list-style-type: none"> - Tight-junction disruption - Blood-brain barrier breakdown - Degeneration/cell death
<p>Vascular endothelial cells</p>	<ul style="list-style-type: none"> - Tight-junction disruption - Blood-brain barrier breakdown - Degeneration/cell death

Computationally deciphering cell types in the brain



Ryan Corces

Computationally deciphering cell types in the brain

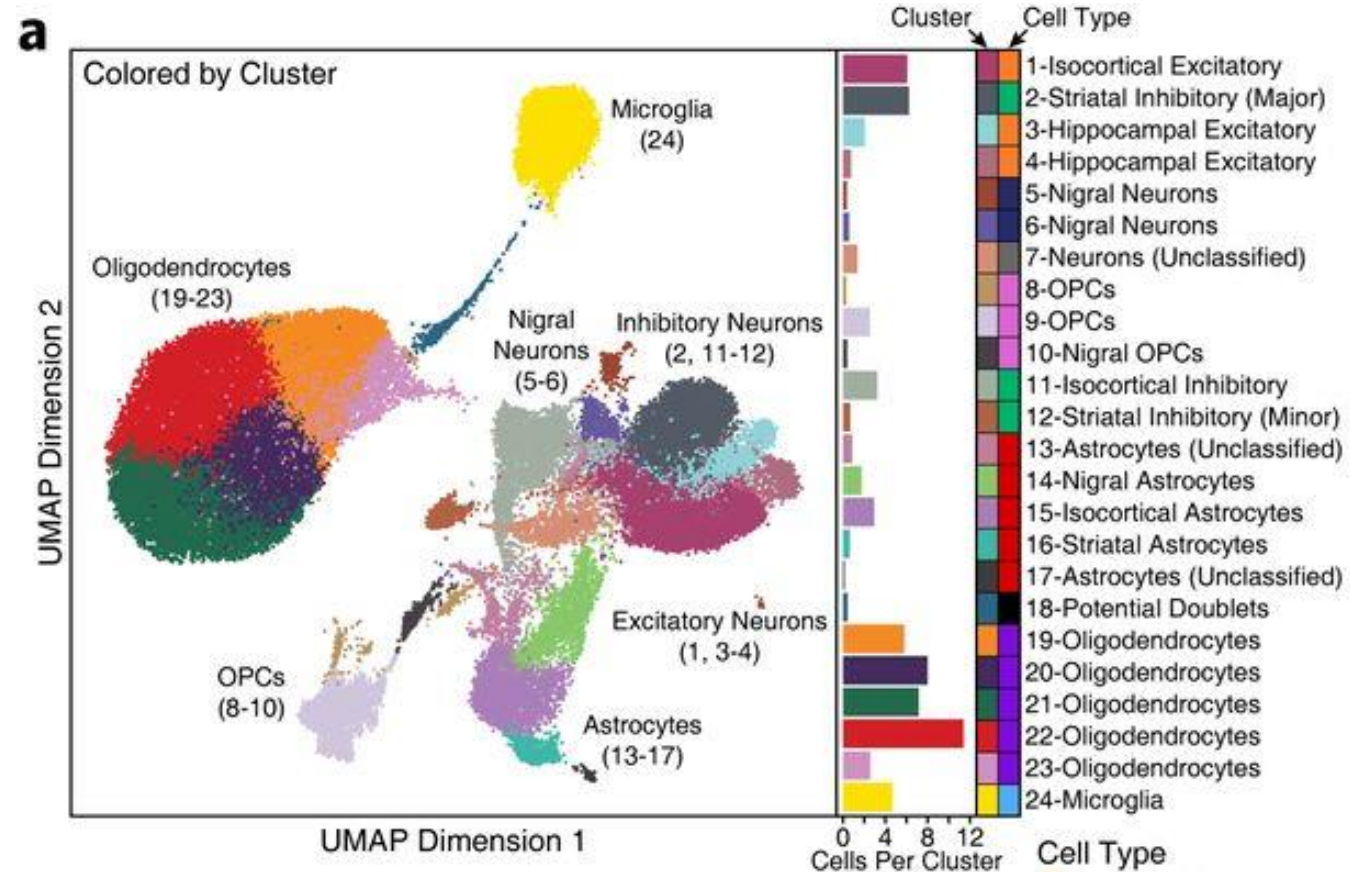
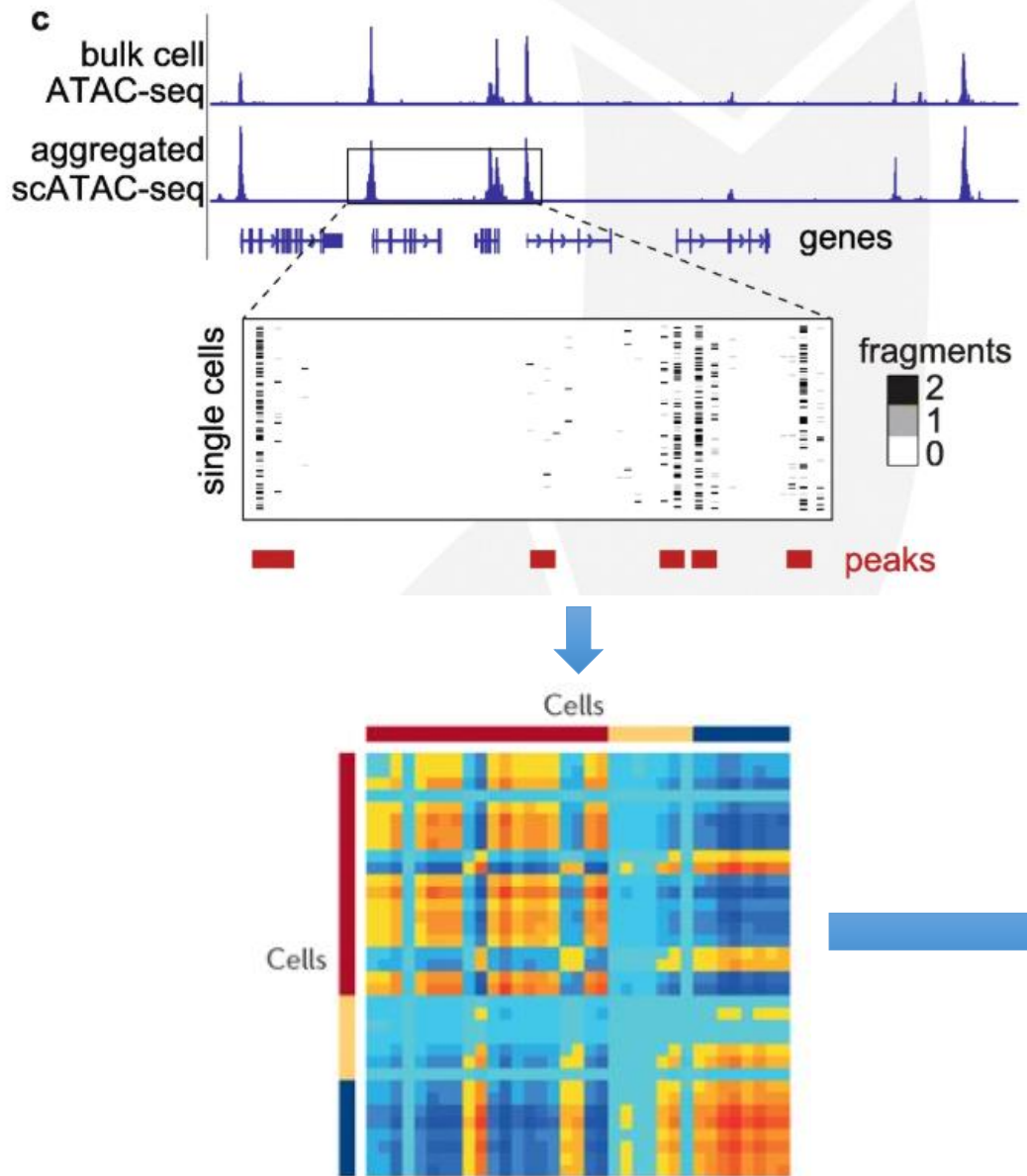


Ryan Corces

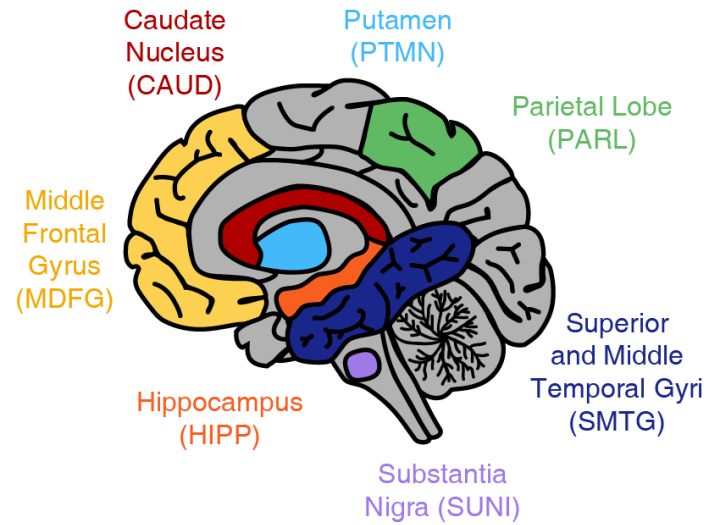
Computationally deciphering cell types in the brain



Ryan Corces

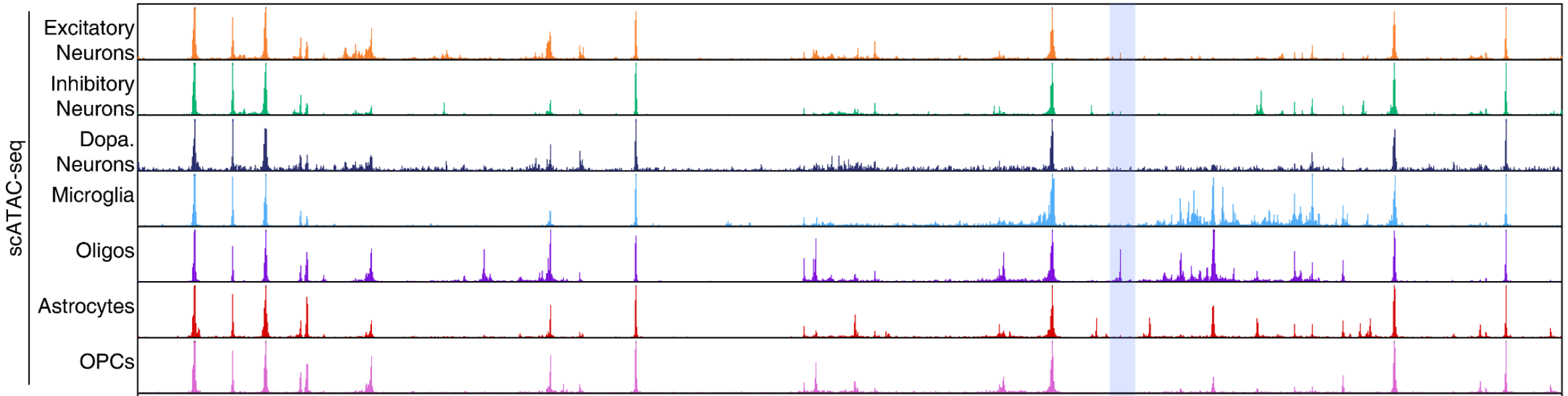


Molecular profiling of cell types in the brain

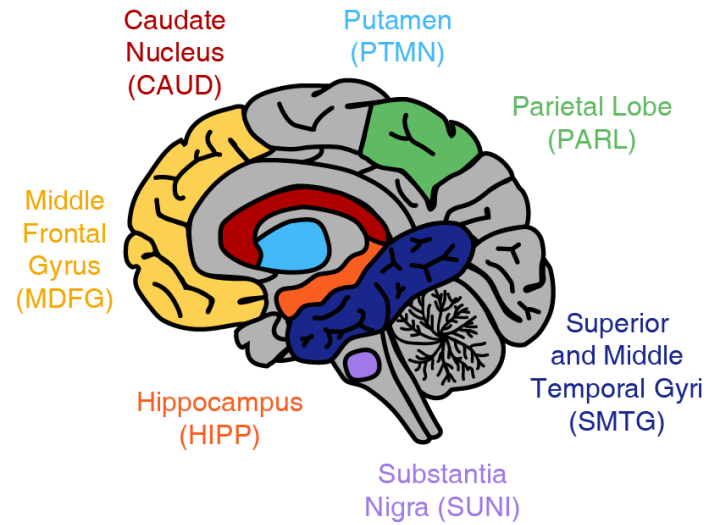


Corces et al. 2020, Nature Genetics

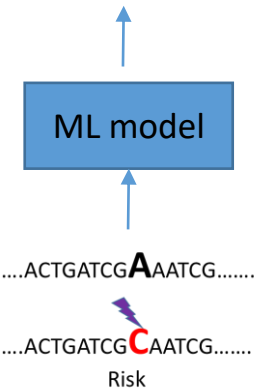
chr11:85599000-86331000 - Alzheimer's Disease rs1237999 - PICALM Locus



Molecular profiling of cell types in the brain

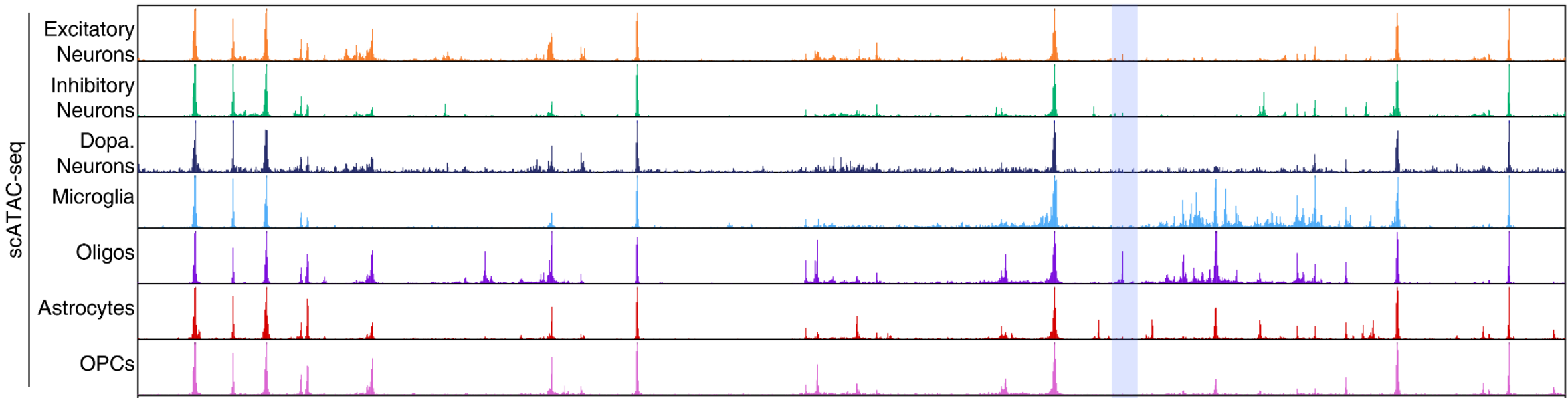


Δ Predicted signal



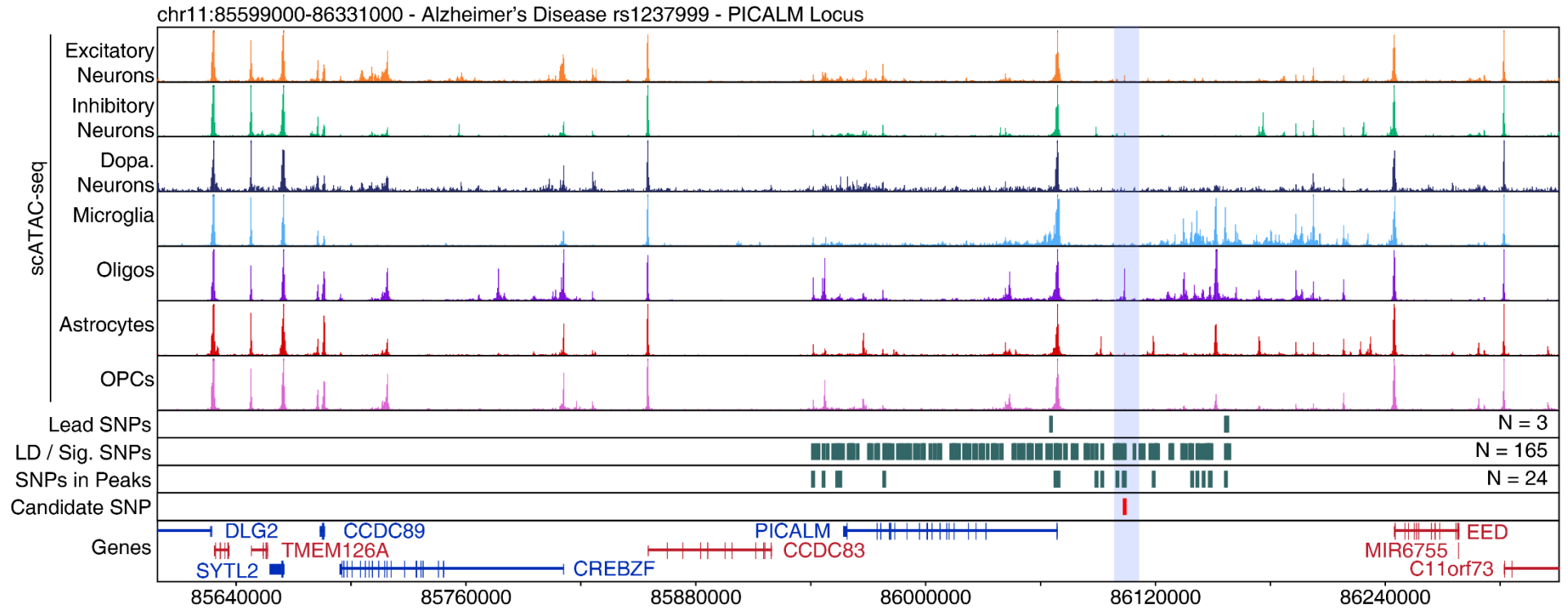
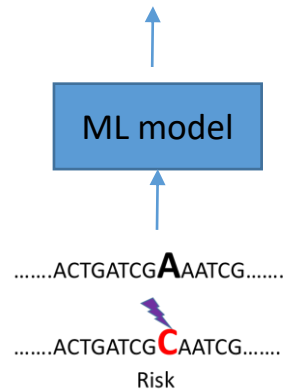
Corces et al. 2020, Nature Genetics

chr11:85599000-86331000 - Alzheimer's Disease rs1237999 - PICALM Locus



Predicting and interpreting causal AD variants

Δ Predicted signal



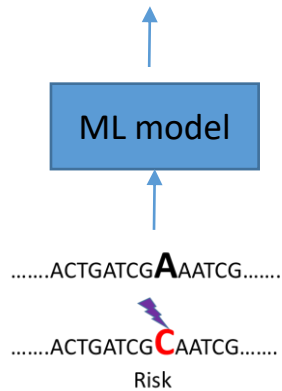
Anna Shcherbina



Soumya Kundu

Predicting and interpreting causal AD variants

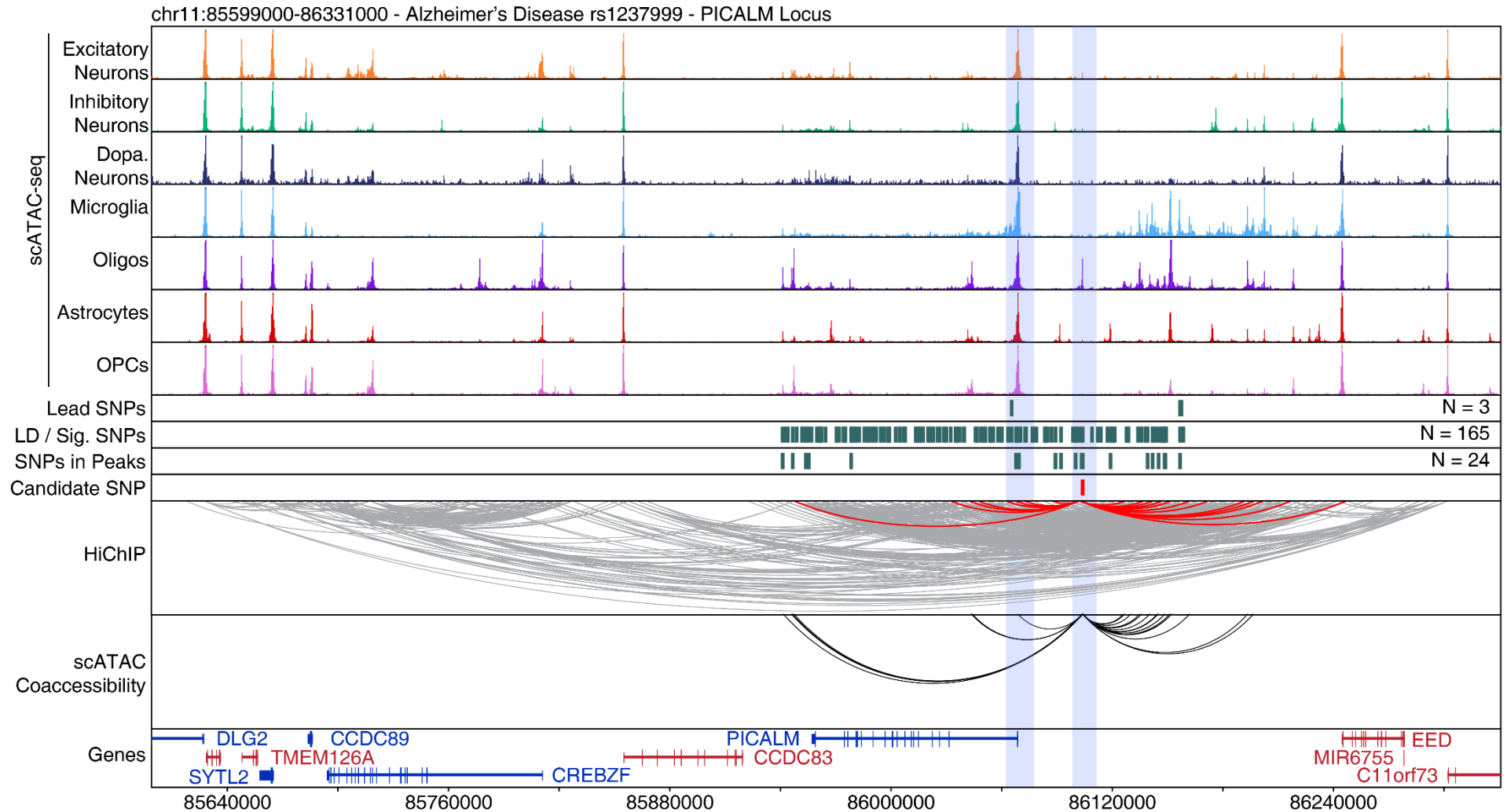
Δ Predicted signal



Anna Shcherbina

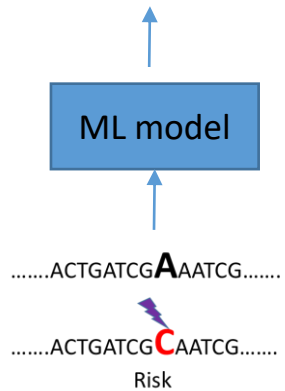


Soumya Kundu



Predicting and interpreting causal AD variants

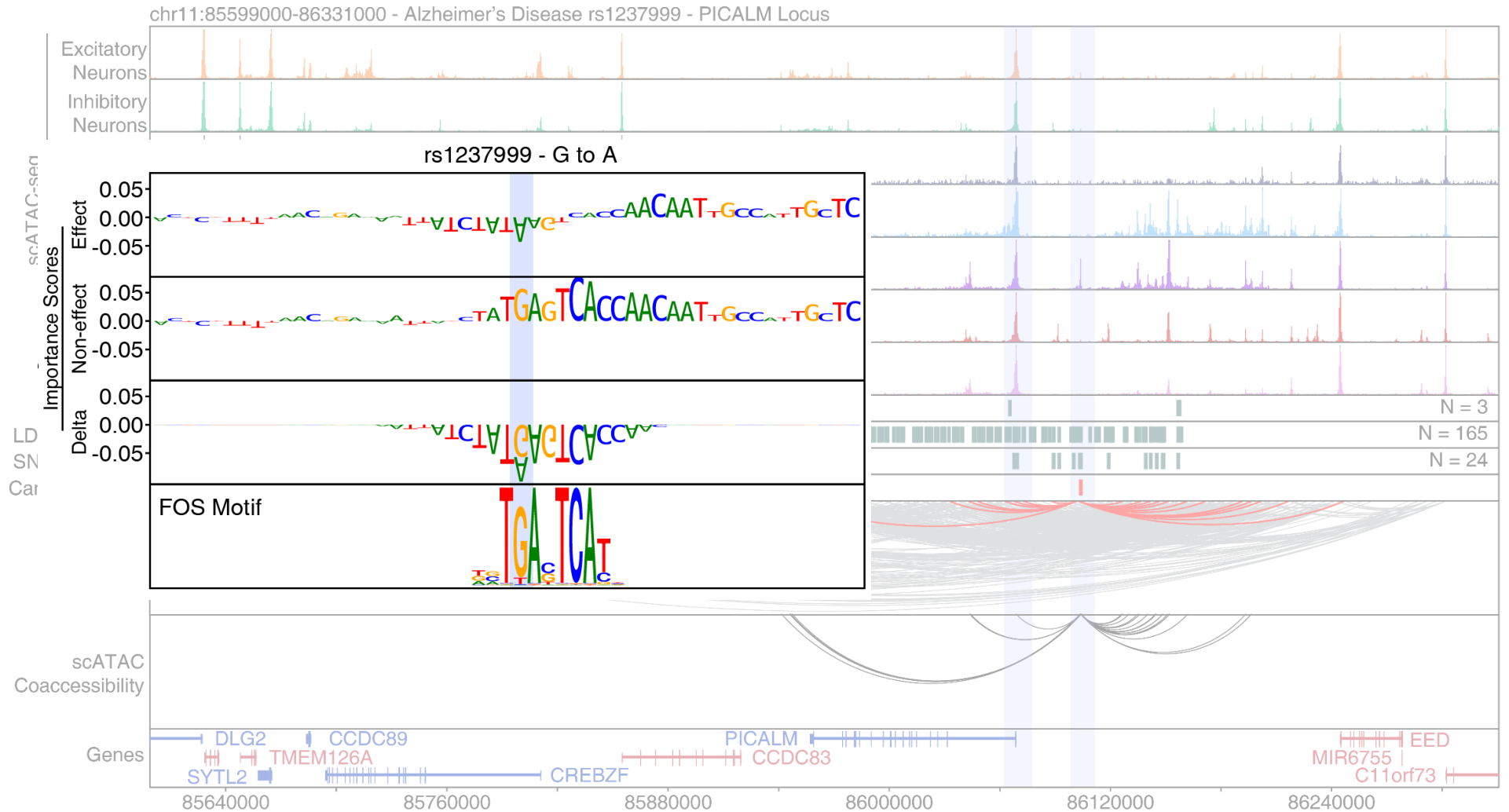
Δ Predicted signal



Anna Shcherbina

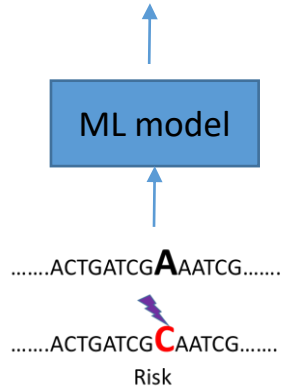


Soumya Kundu



Predicting and interpreting causal AD variants

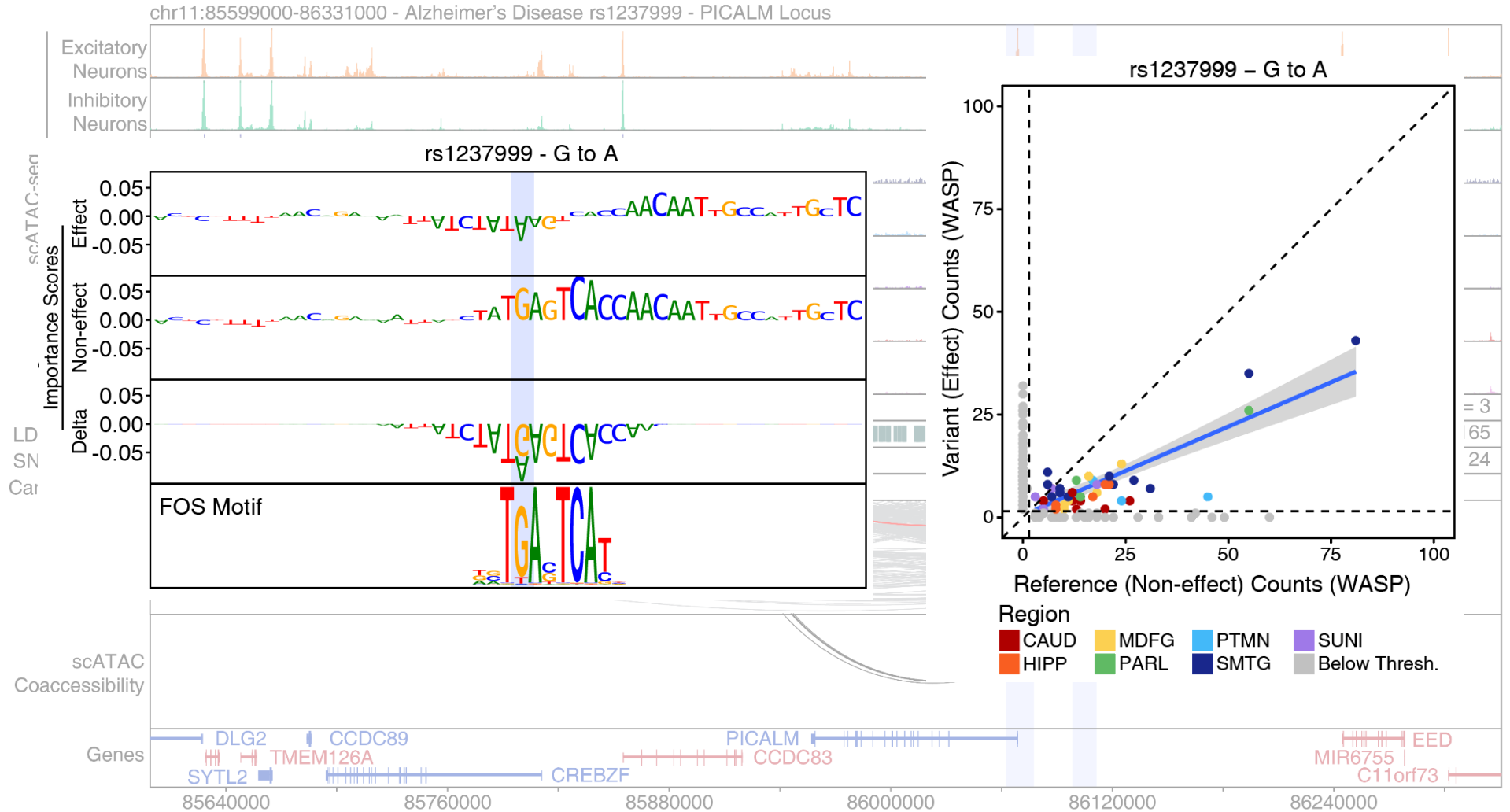
Δ Predicted signal



Anna Shcherbina

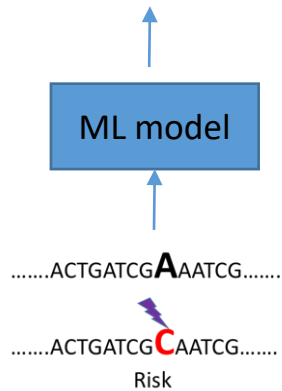


Soumya Kundu



Predicting and interpreting causal AD variants

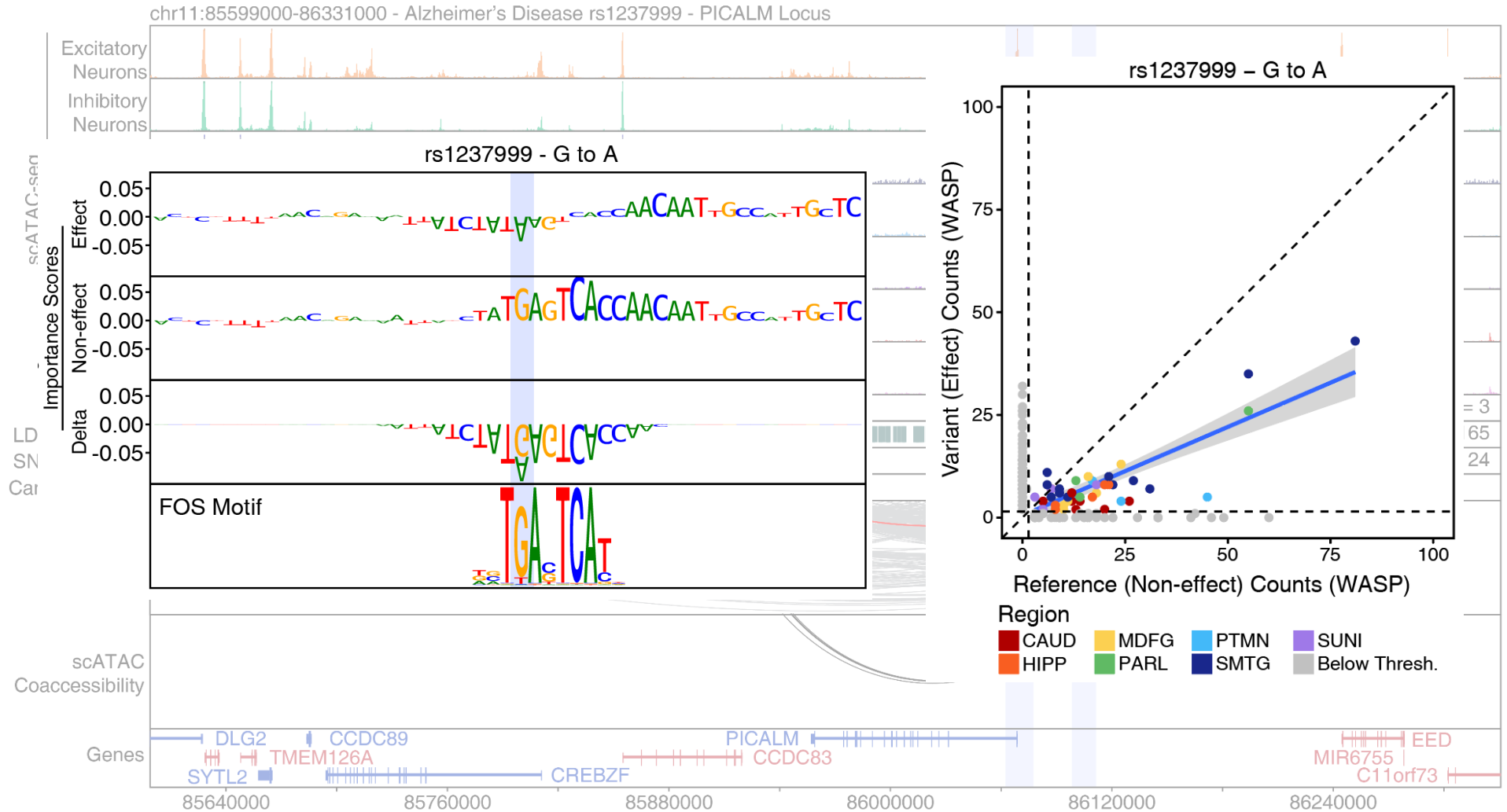
Δ Predicted signal



Anna Shcherbina

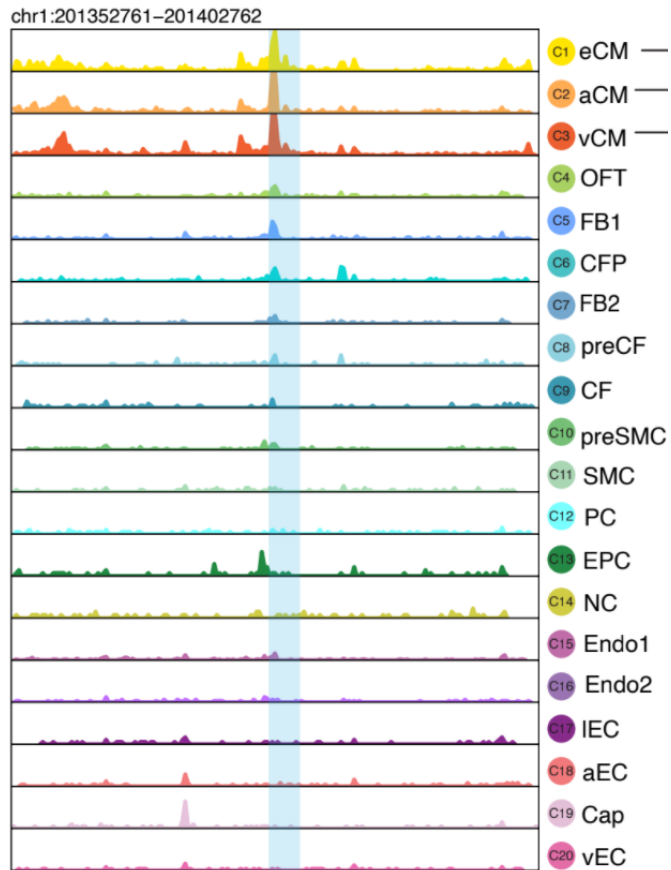
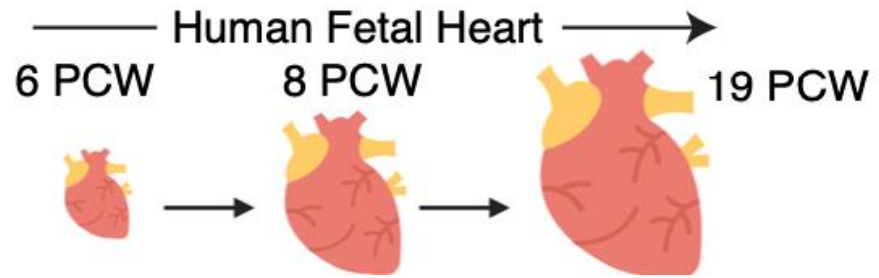


Soumya Kundu



Genetic variant rs1237999 disrupts a sequence motif of the FOS protein in a control element of the PICALM gene in oligodendrocyte cells in the brain

Predicting *de-novo* non-coding mutations in congenital heart disease

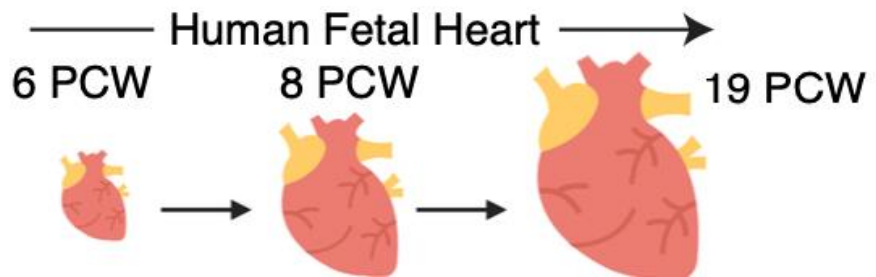


Lakshman
Sundaram

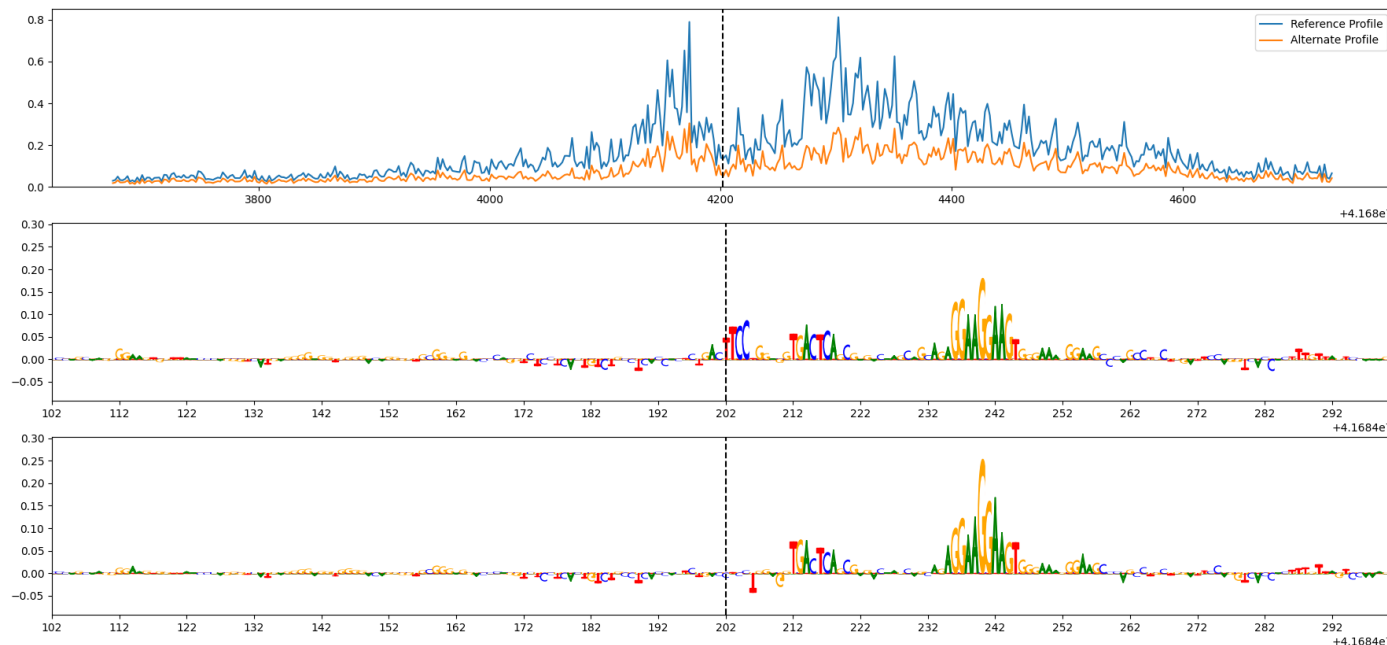
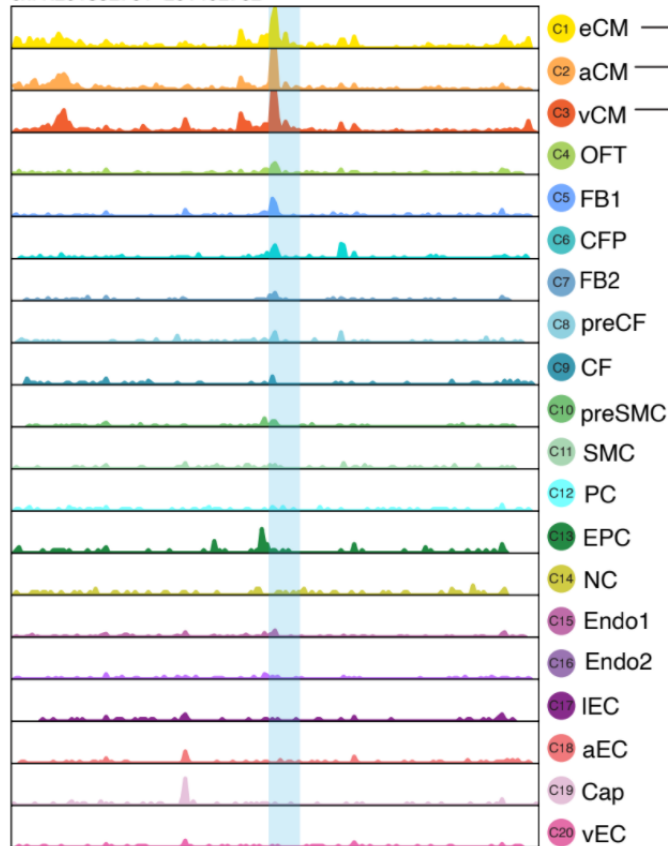


Mo Ameen

Predicting *de-novo* non-coding mutations in congenital heart disease



chr1:201352761-201402762



Prediction: Mutation disrupts ETV motif in control element active in arterial endothelial cells

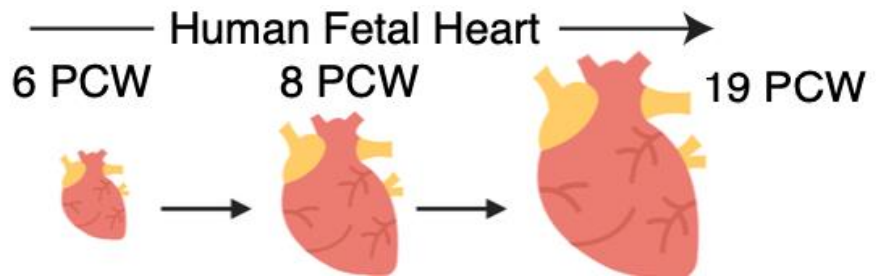


Lakshman Sundaram

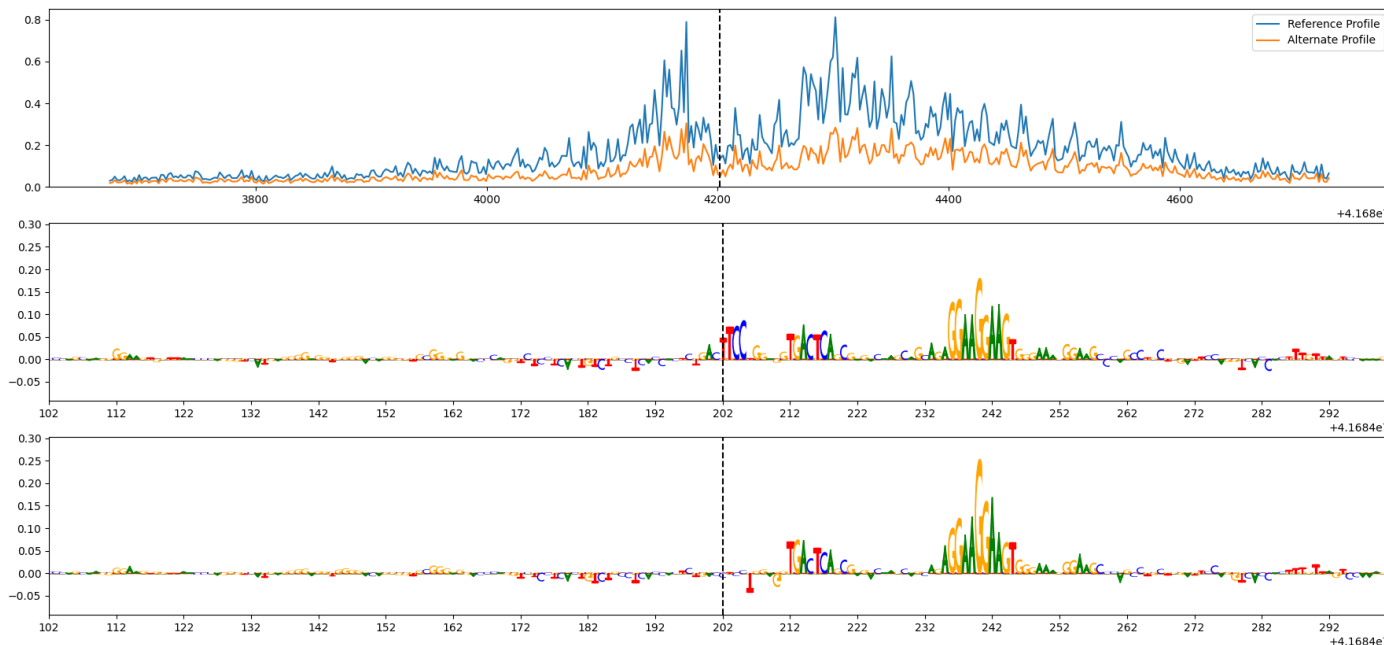
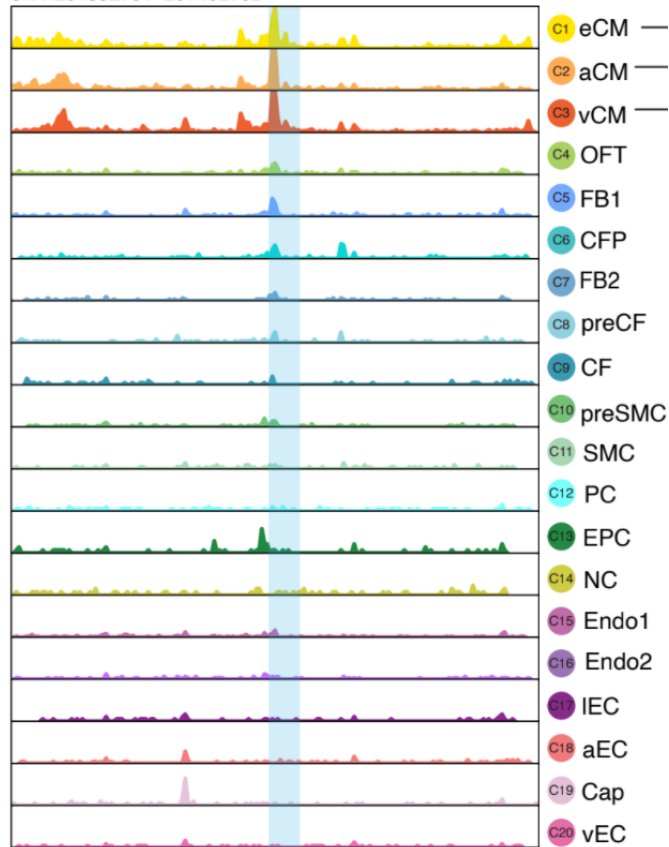


Mo Ameen

Predicting *de-novo* non-coding mutations in congenital heart disease



chr1:201352761-201402762

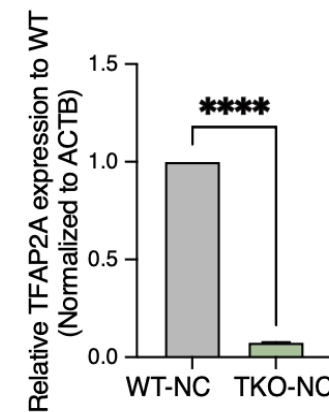
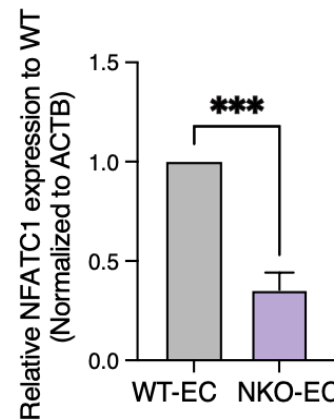
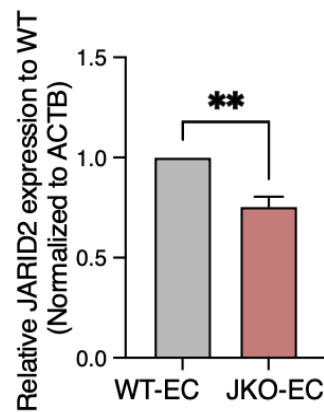


Lakshman Sundaram



Mo Ameen

Prediction: Mutation disrupts ETV motif in control element active in arterial endothelial cells



CRISPR/Cas9 experiments validate downstream target genes

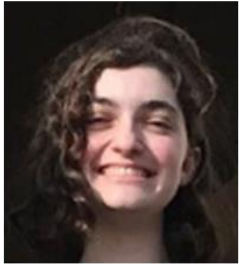
Summary

- Large-scale molecular profiling datasets => decipher genome function
- Neural networks can map DNA sequence to molecular profiles with unprecedented accuracy
- Models can be interpreted to decipher functional DNA letters, words and syntax
- Models can be used to decipher disease-associated mutations
- Predictions are validated by genome editing experiments
- Predictions can provide clues for therapeutic interventions

Kundaje lab



Daniel Kim (BMI)



Kelly Cochran (CS)



Soumya Kundu (CS)



Surag Nair (CS)



Maxim Zaslavsky (CS)



Vivek Ramalingam (Postdoc)



Caleb Lareau (Postdoc)



Akshay Balsubramani (Postdoc)



Georgi Marinov (Postdoc)



Alex Tseng (CS)



Amr Alexandari (CS)



Abhimanyu Banerjee (Physics)



Lakshman Sundaram (CS)



Anusri Pampari (CS)



Kristy Mualim (Bioinformatician)



Jacob Schreiber (Postdoc)



Mahfuza Sharmin (Postdoc)



Eran Kotler (Postdoc)



Zahoor Zafrulla (ML engineer)

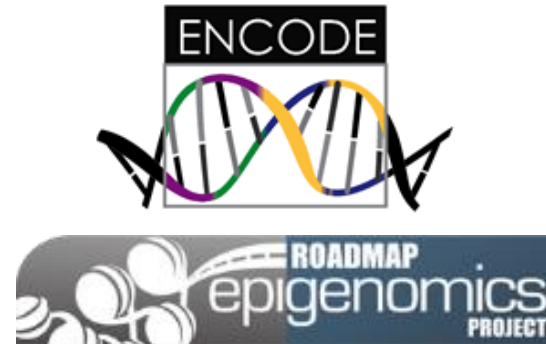
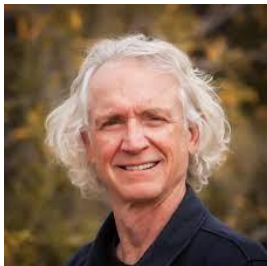
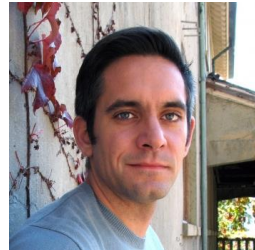
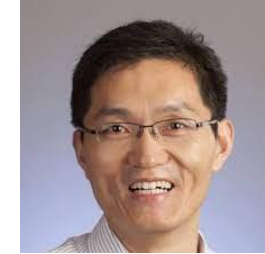


Chris Probert (Genetics)



Jin Wook Lee (Software engineer)

Collaborator labs



Funding



1R01HG009674
1U01HG009431
1U24HG009446
R01ES02500902

1DP2OD022870

