Genetic variation and regulation of the 3D genome

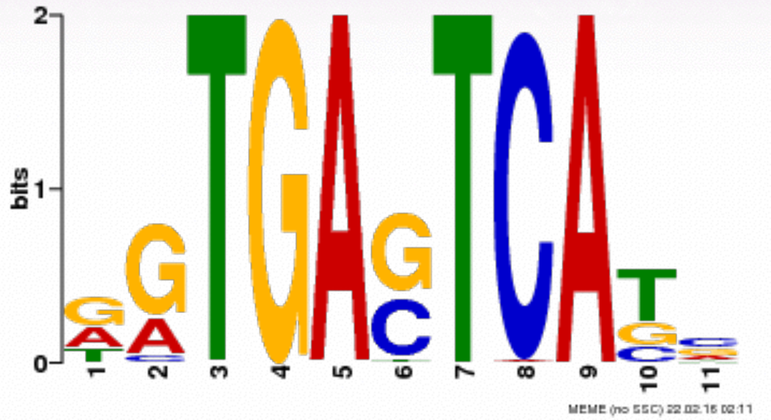
Katie Pollard
Gladstone Institute of Data Science & Biotechnology
University of California San Francisco
Chan-Zuckerberg Biohub

SAGES Conference

University of Pennsylvania • June 7, 2019

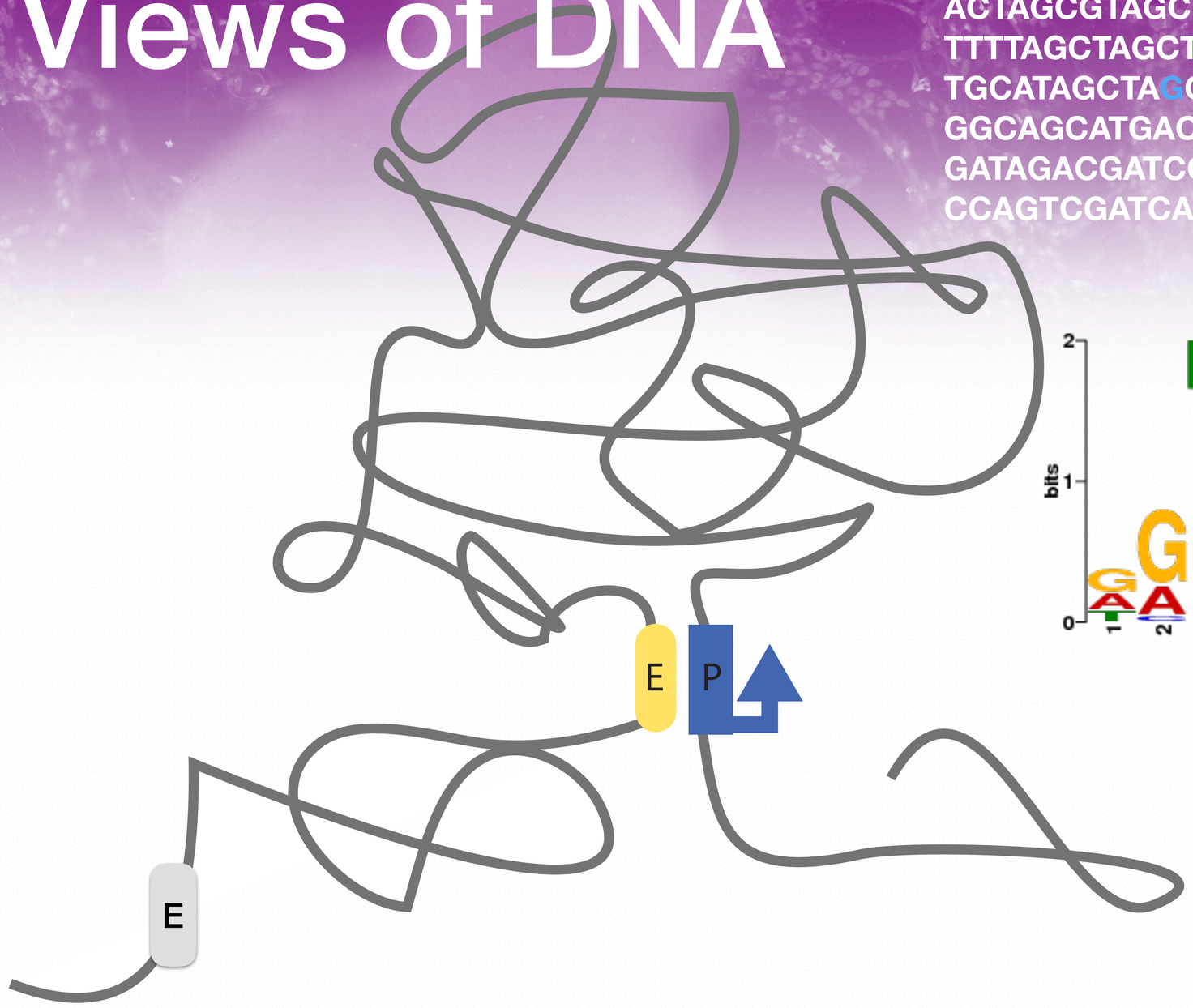
Views of DNA

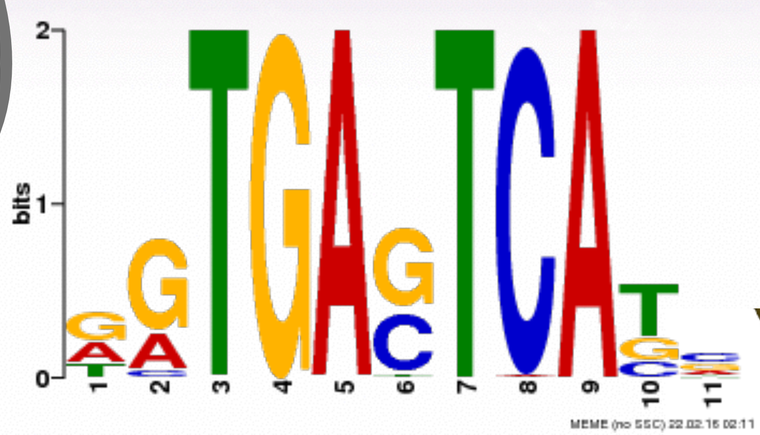
ACTAGCGTAGCTAGCGATATCTAGGGCGATCGATGCTACGTATCGAGC
TTTTAGCTAGCTAGCTAGCATCGATGCATCGATCGTACGATCGATCGTA
TGCATAGCTAGCTAGCATGCATGCATCGATCGAATCGATATTAGCTAGC
GGCAGCATGACTAGTCAGATATCGTACGATGTCGAAAACTGATCAGTC
GATAGACGATCGATCGATCGATCGAGGCGCATCGATCGATGCTAGCAT
CCAGTCGATCAGTCGATCGATCGATCGATCGATCGATCGACTAGATCG

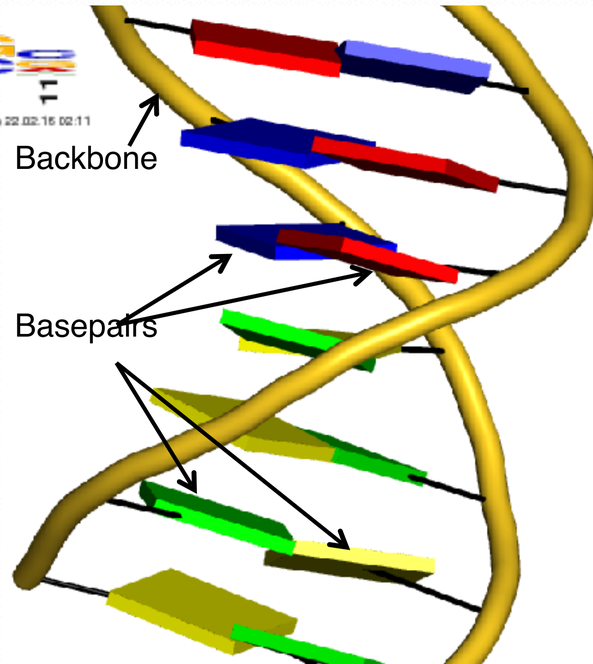


Views of DNA

ACTAGCGTAGCTAGCGATATCTAGGGCGATCGATGCTACGTATCGAGC
TTTTAGCTAGCTAGCTAGCATCGATGCATCGATCGTACGATCGATCGTA
TGCATAGCTAGCTAGCATGCATGCATCGATCGAATCGATATTAGCTAGC
GGCAGCATGACTAGTCAGATATCGTACGATGTCGAAAACTGATCAGTC
GATAGACGATCGATCGATCGATCGAGGCGCATCGATCGATGCTAGCAT
CCAGTCGATCAGTCGATCGATCGATCGATCGATCGATCGACTAGATCG







Chromatin Structure Meets Population Genetics

What is the relationship between genetic variation and chromatin structure?

1. Open chromatin in the human developing telencephalon sheds light on non-coding mutations in autism.

   Eirene Markenscoff-Papadimitriou, Pawel Przytycki & Sean Whalen

2. Chromatin boundaries are under strong negative selection.

   Geoff Fudenberg

3. Chromatin interactions and linkage disequilibrium are uncorrelated along the human genome.

   Sean Whalen

Chromatin Structure Meets Population Genetics

What is the relationship between genetic variation and chromatin structure?

1. Open chromatin in the human developing telencephalon sheds light on non-coding mutations in autism.

   Eirene Markenscoff-Papadimitriou, Pawel Przytycki & Sean Whalen

2. Chromatin boundaries are under strong negative selection.

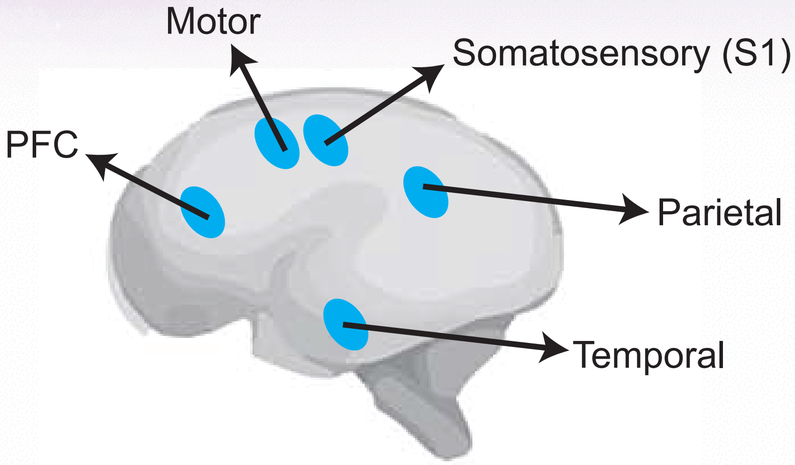
   Geoff Fudenberg

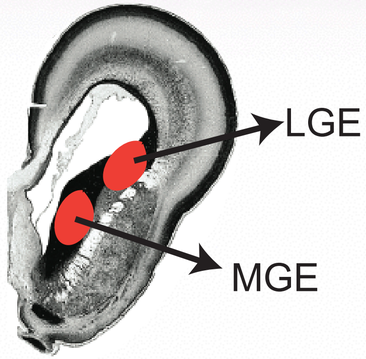
3. Chromatin interactions and linkage disequilibrium are uncorrelated along the human genome.

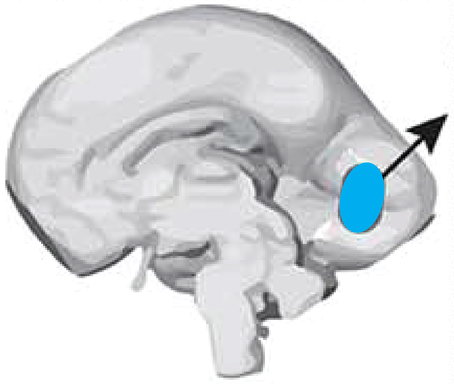
   Sean Whalen

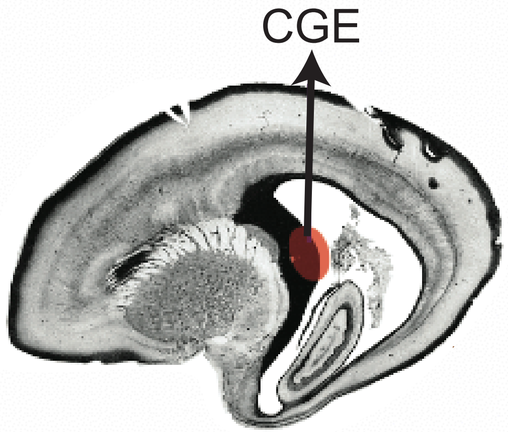
Telencephalon Open Chromatin Atlas

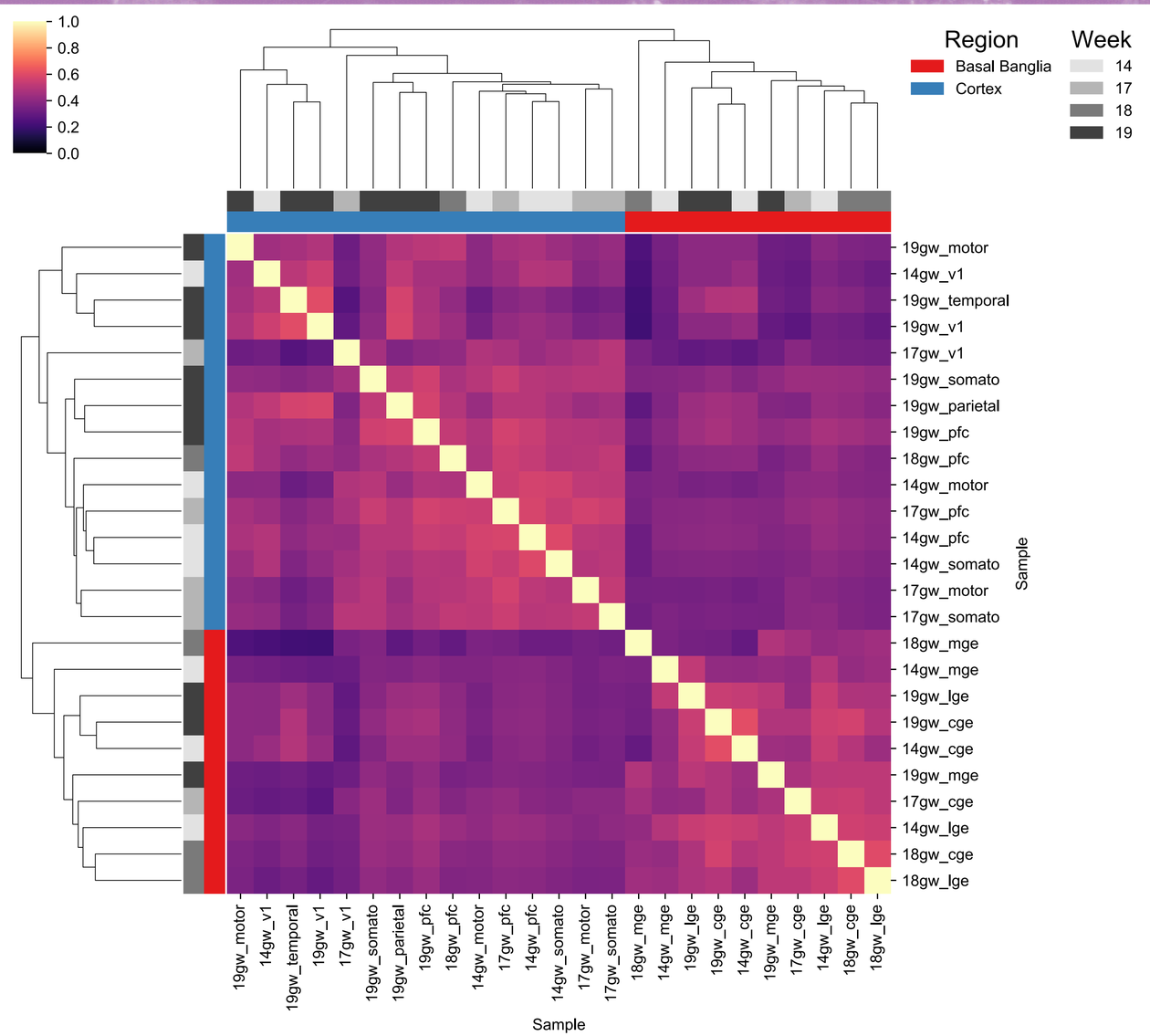
ATACseq on microdissected human brain regions 14-19 GW





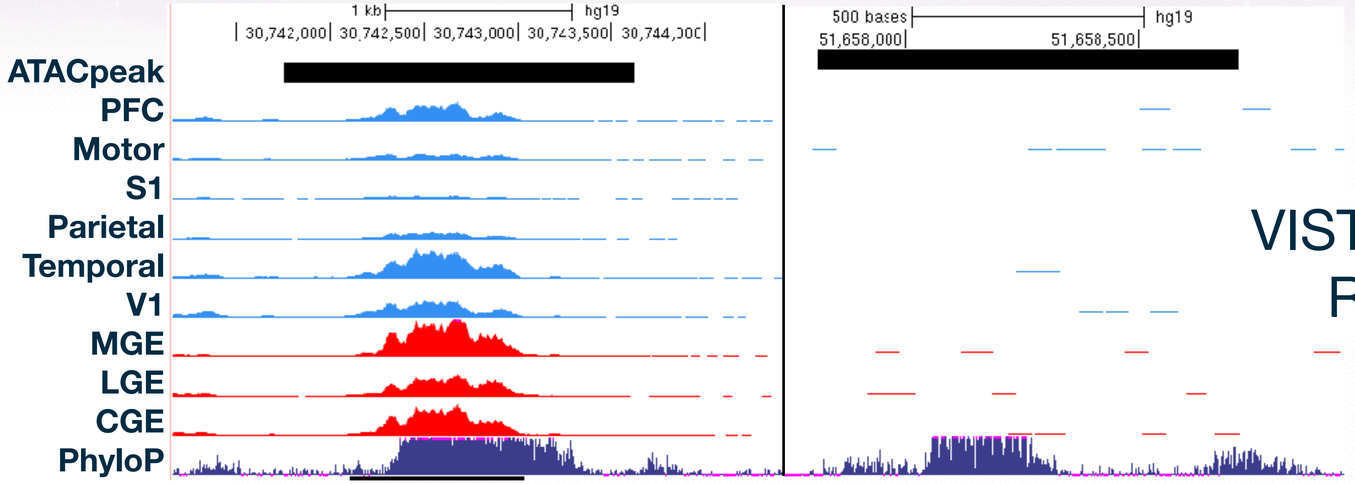






EnhancerFinder predicts active regions

Ensemble learning model trained on VISTA enhancers

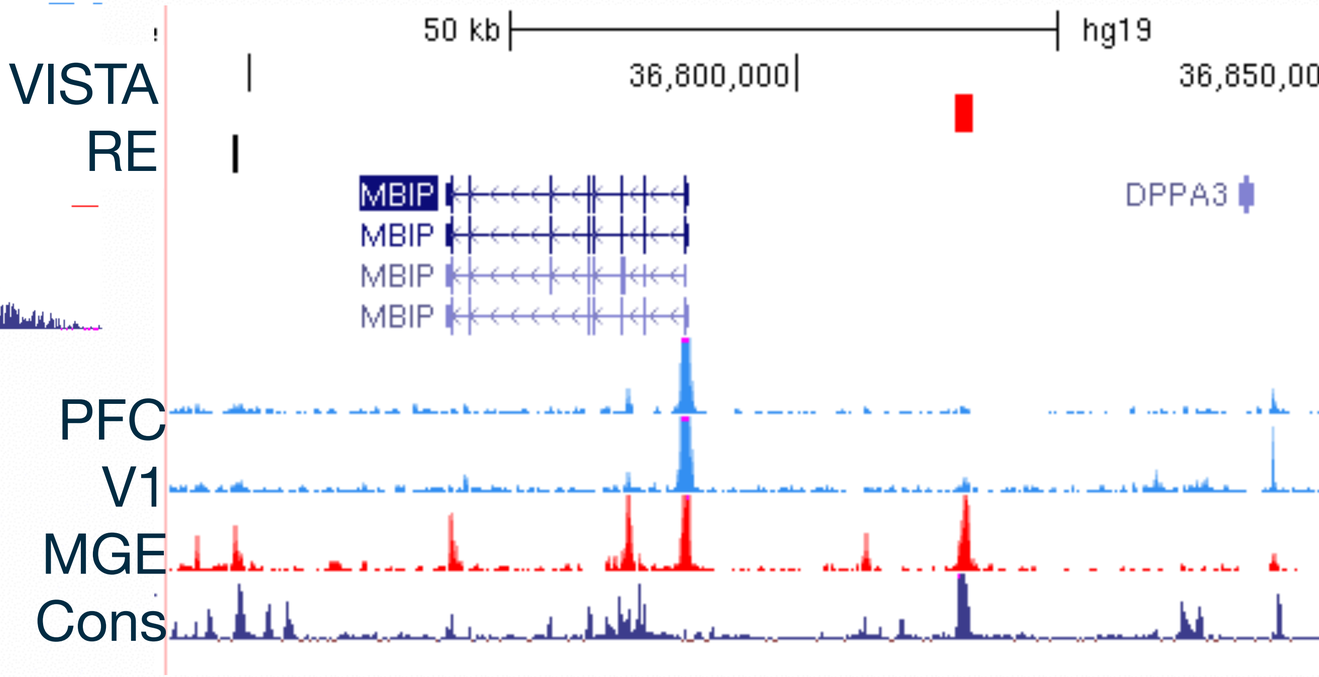


hs433

VS.

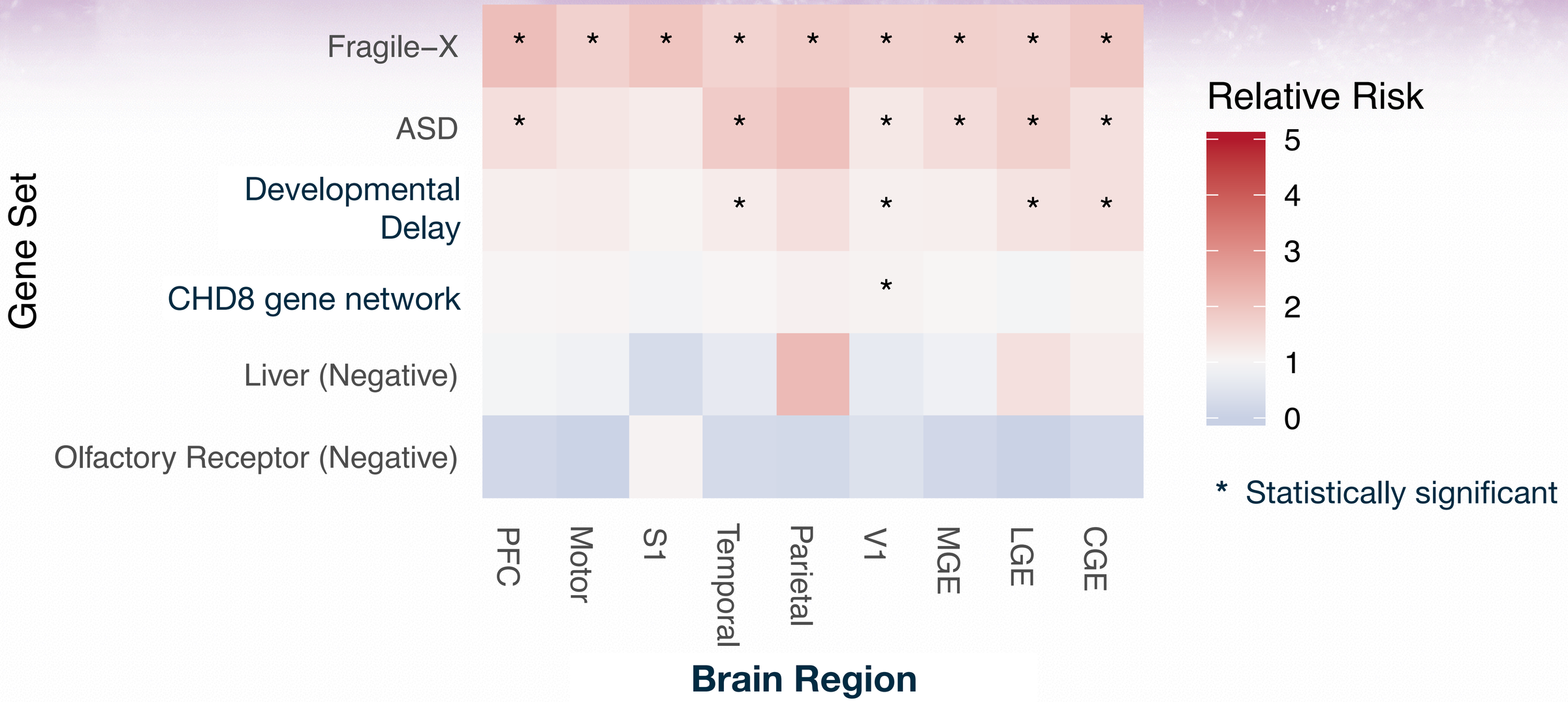
hs72

Predict which open chromatin regions are active (REs) and annotate as regional, temporal, layer specific using ATACseq

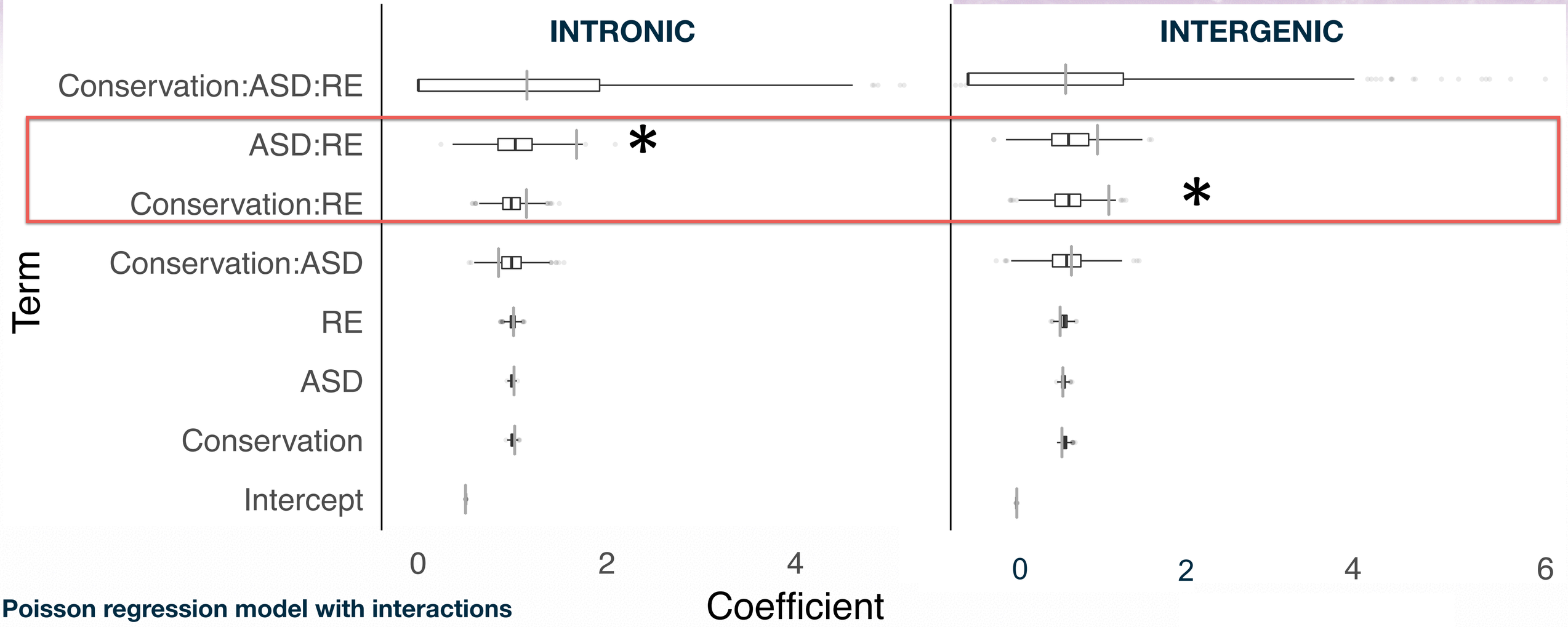


10-20% per region predicted as REs

REs link disease risk to brain regions

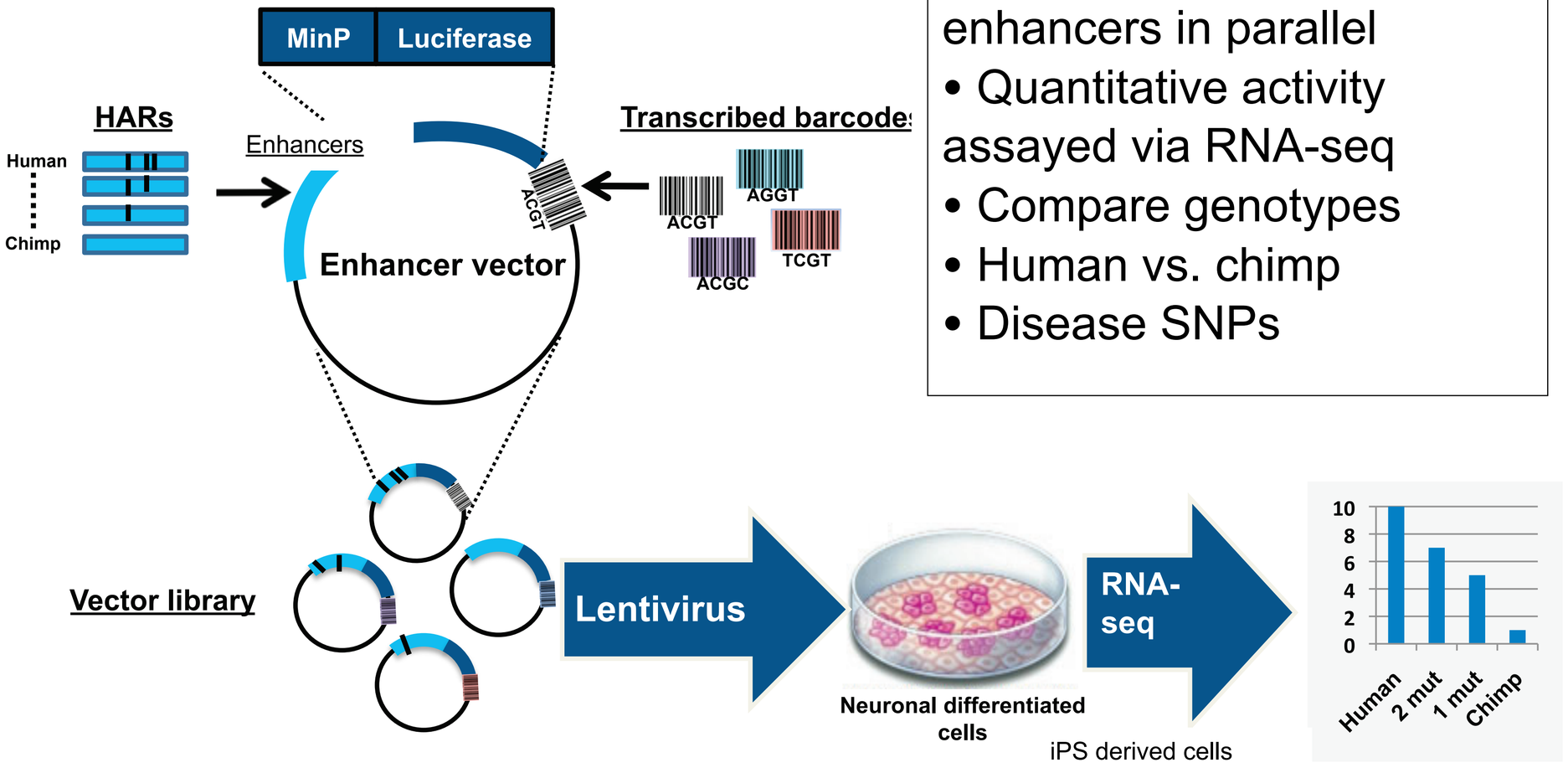


REs link disease risk to specific subsets of noncoding elements

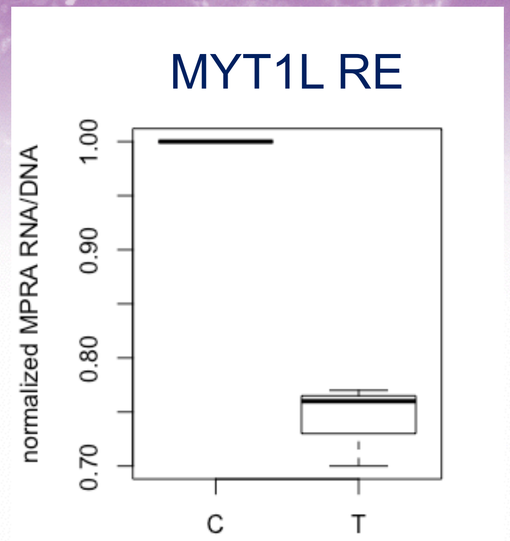


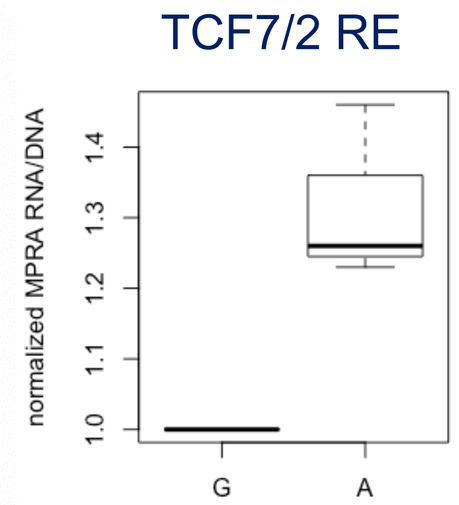
Poisson regression model with interactions (forward model selection)

Massively Parallel Reporter Assays for validation & mutation testing



- Test >12,000 170bp enhancers in parallel
- Quantitative activity assayed via RNA-seq
- Compare genotypes
- Human vs. chimp
- Disease SNPs





with Hane Ryu, Fumitaka Inoue, Nadav Ahituv, Jay Shendure

Conclusions

- Open chromatin is dynamic between brain regions and layers.
- Machine learning identifies a subset of open chromatin regions most likely to be enhancers. These are enriched for association with neurodevelopmental genes and psychiatric disease genes.
- Diseases can be mapped to dynamic enhancers and the brain regions in which they are active.
- Autism risk alleles are enriched in intronic enhancers of ASD genes and conserved sites in intergenic enhancers, and not in all open chromatin around ASD genes or all predicted enhancers.
- MPRAs quantify differential activity of enhancer alleles.

Chromatin Structure Meets Population Genetics

What is the relationship between genetic variation and chromatin structure?

1. Open chromatin in the human developing telencephalon sheds light on non-coding mutations in autism.

   Eirene Markenscoff-Papadimitriou, Pawel Przytycki & Sean Whalen

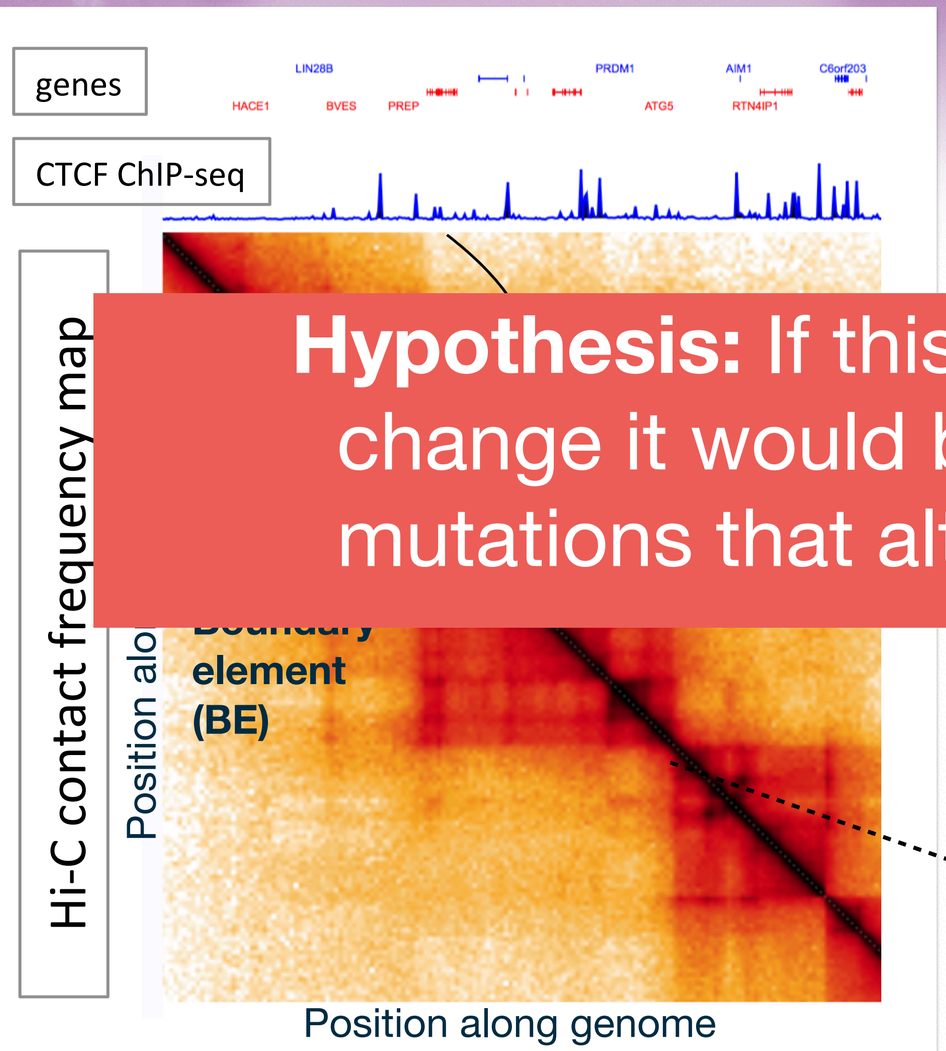
2. Chromatin boundaries are under strong negative selection.

   Geoff Fudenberg

3. Chromatin interactions and linkage disequilibrium are uncorrelated along the human genome.

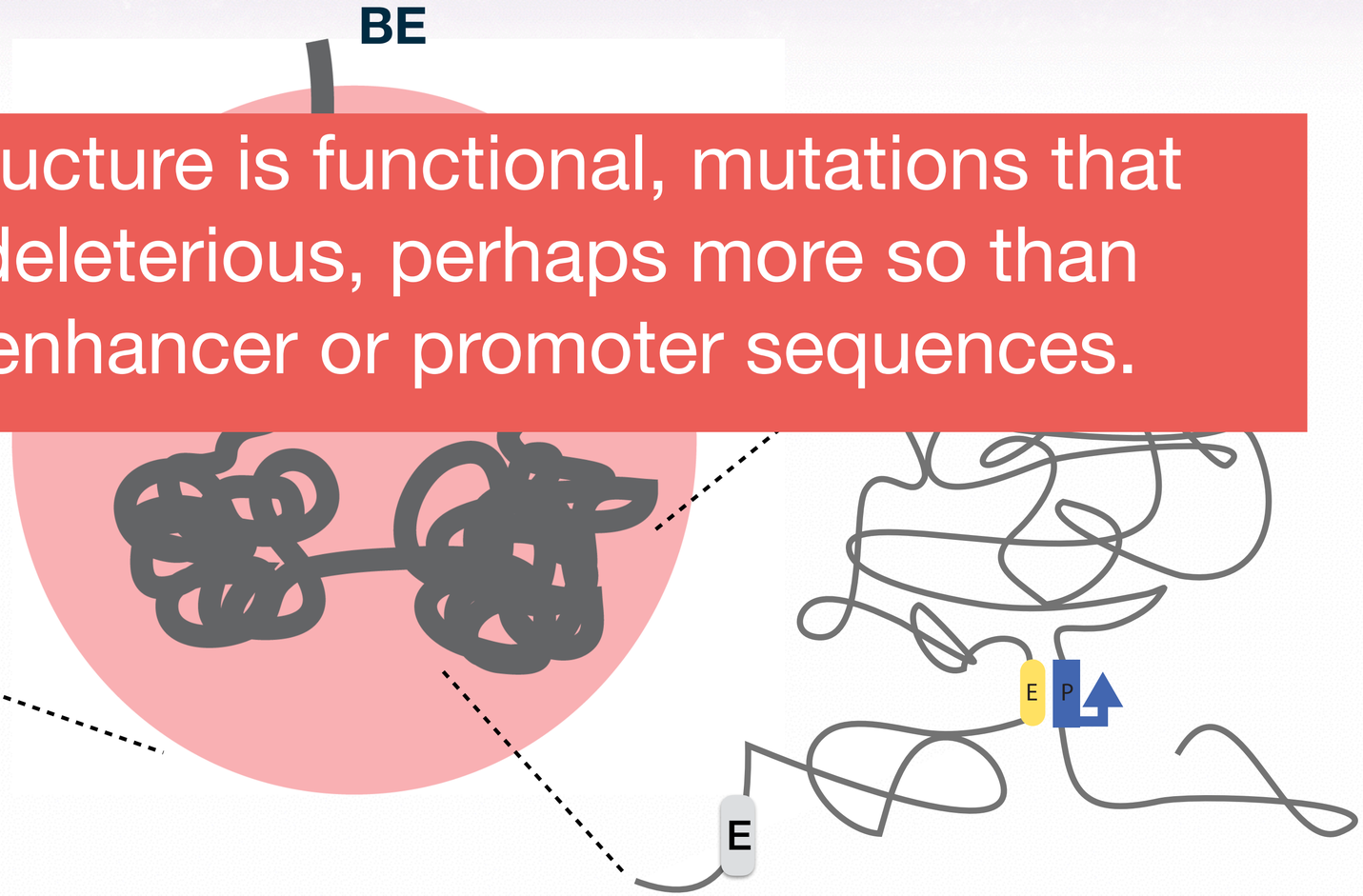
   Sean Whalen

How important are boundaries (BEs)?

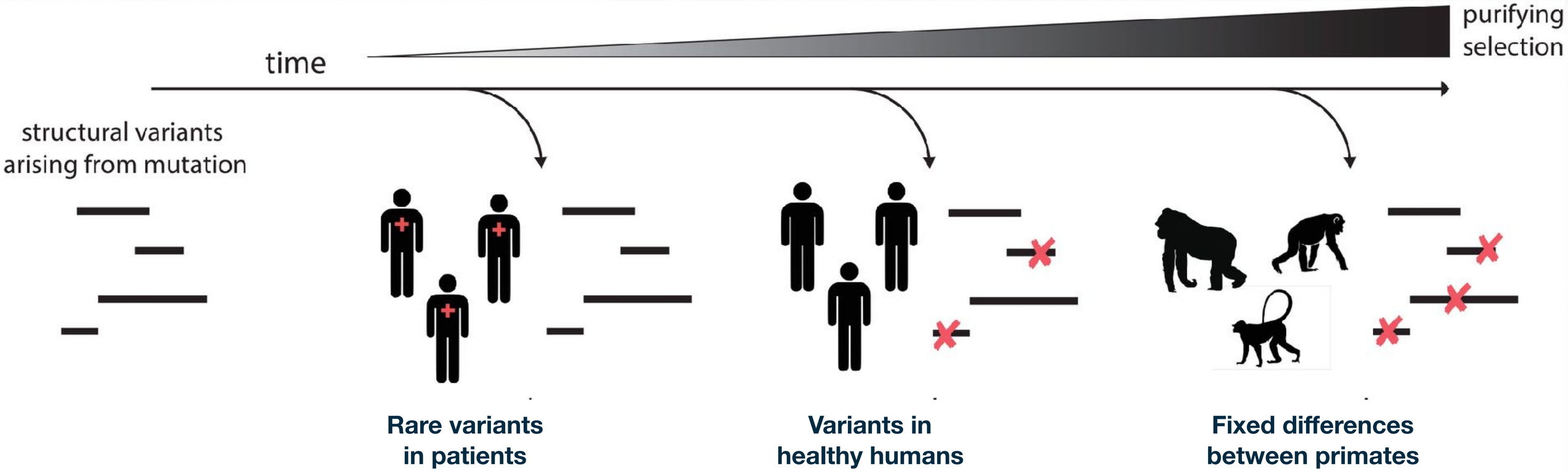


Map downloaded from higlass.io, also see Kerpedjiev, bioRixv, 2017

Hypothesis: If this structure is functional, mutations that change it would be deleterious, perhaps more so than mutations that alter enhancer or promoter sequences.

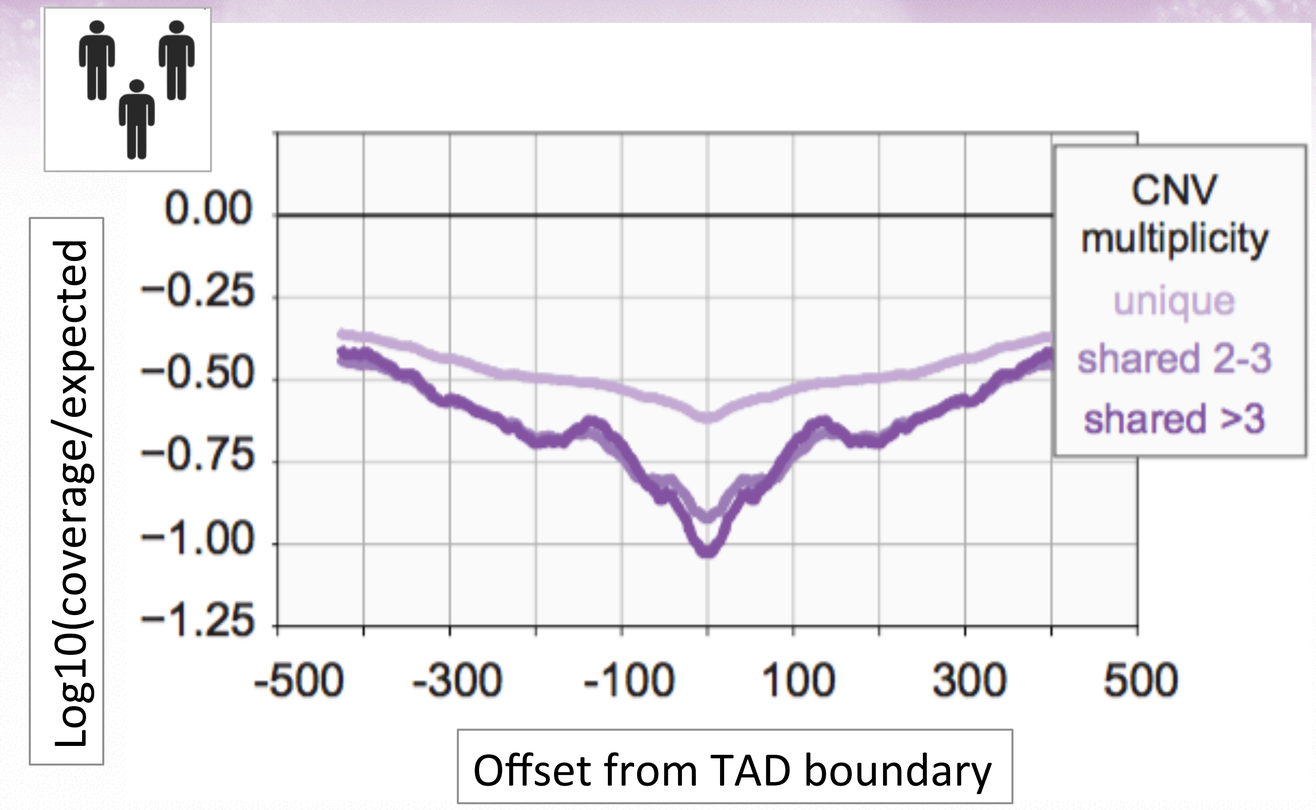


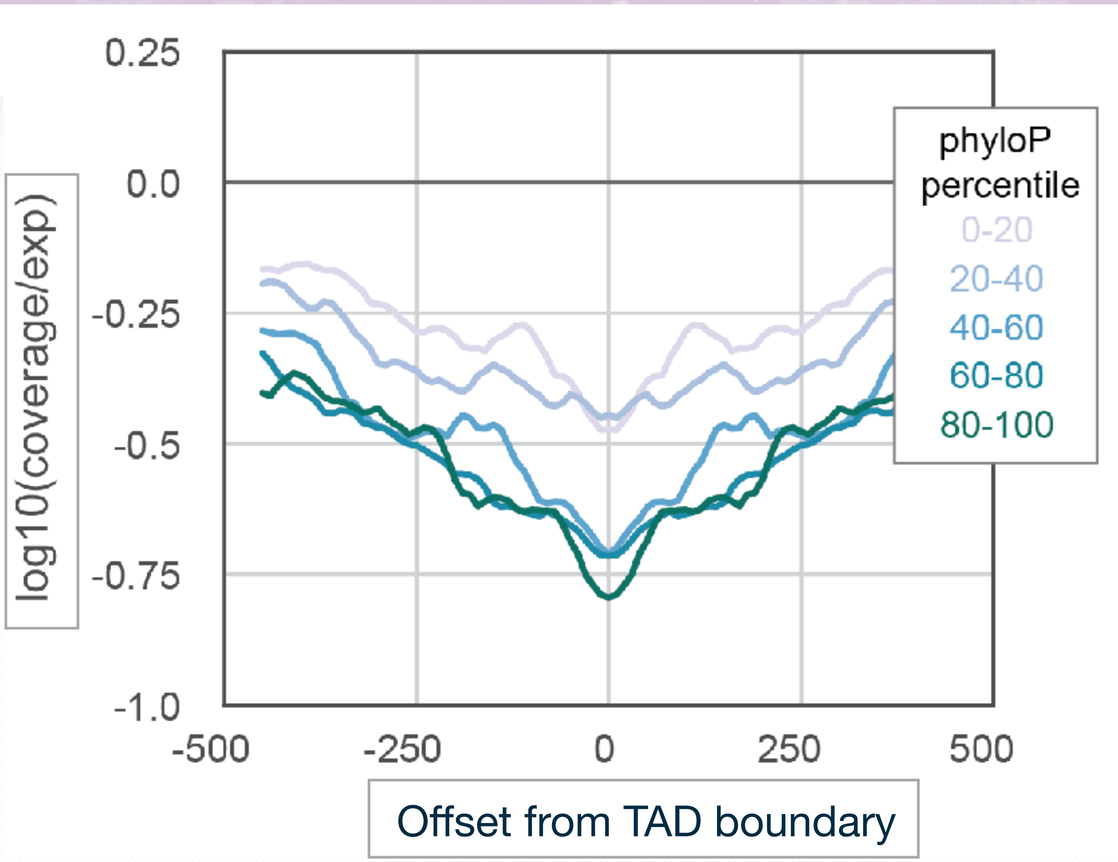
Approach: deleterious deletions will be depleted over time



Structural Variant Data: Apes: Sudmant 2013; Controls, Cases: Coe, 2014

Deletions are depleted at BEs

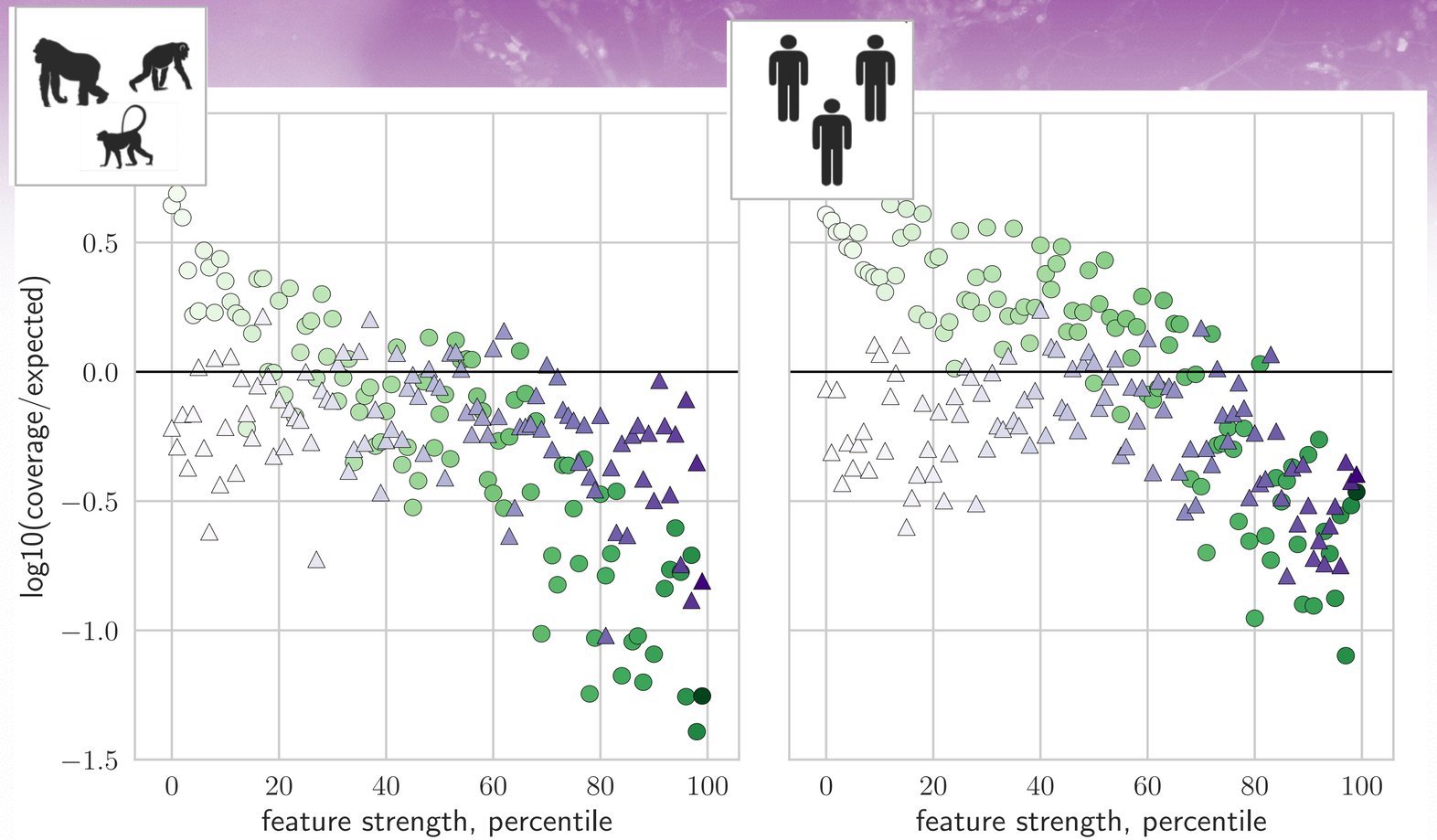




Hi-C Data: Rao et al 2015
Expression Data: GTeX

Genomic Element Data: ENCODE, Epigenomics Roadmap
Structural Variant Data: Apes: Sudmant 2013; Controls, Cases: Coe, 2014

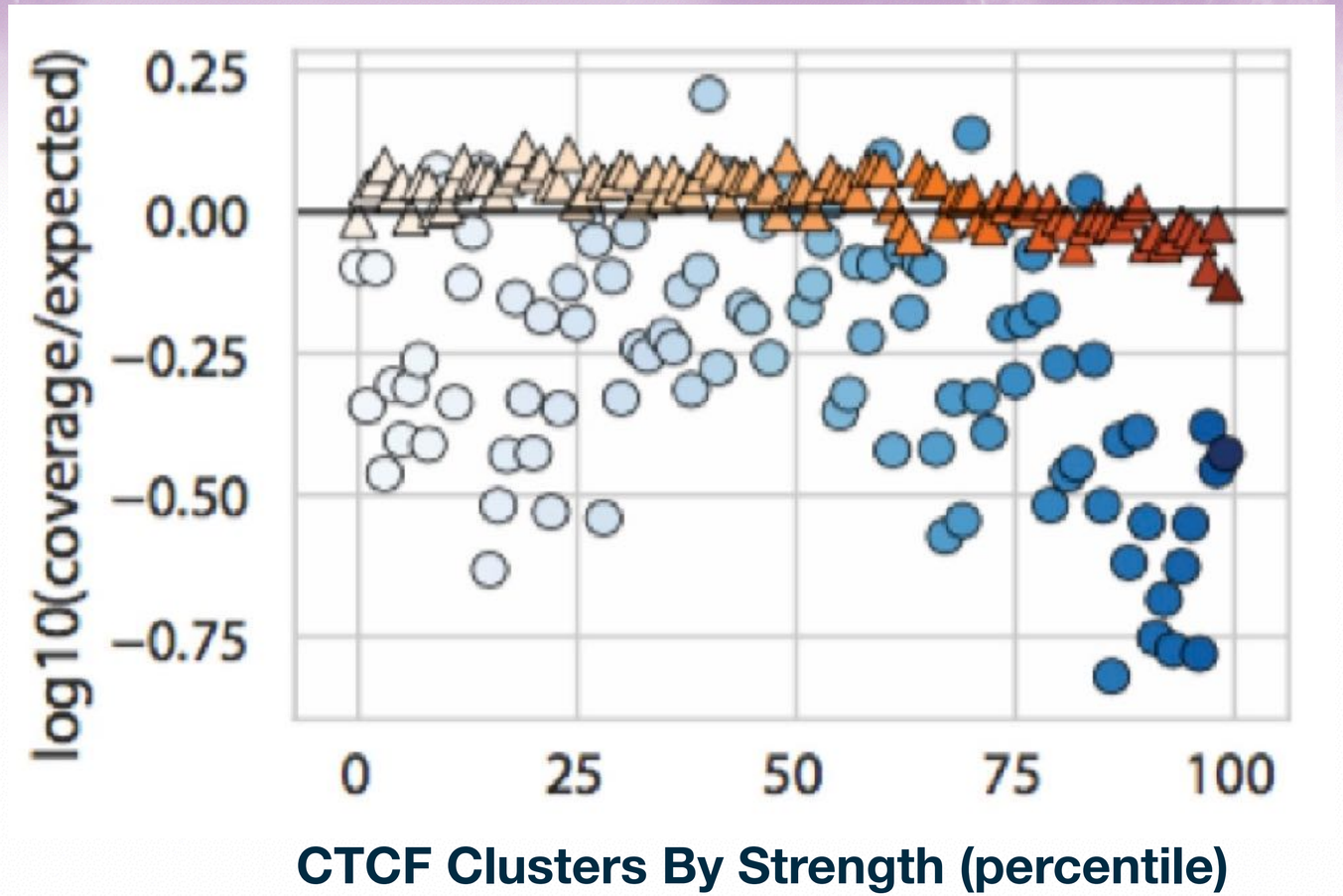
Selection correlates with function

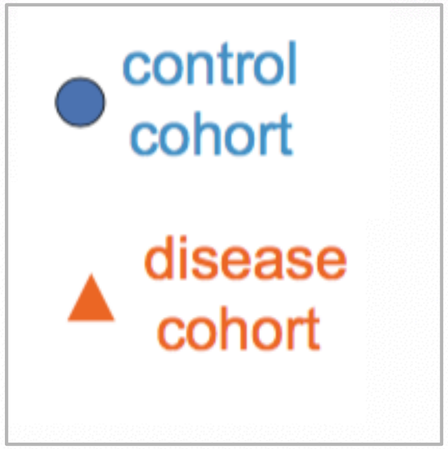


Hi-C Data: Rao et al 2015
Expression Data: GTeX

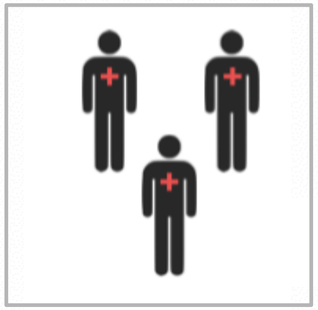
Genomic Element Data: ENCODE, Epigenomics Roadmap
Structural Variant Data: Apes: Sudmant 2013; Controls, Cases: Coe, 2014

But not in autism patients…





vs.

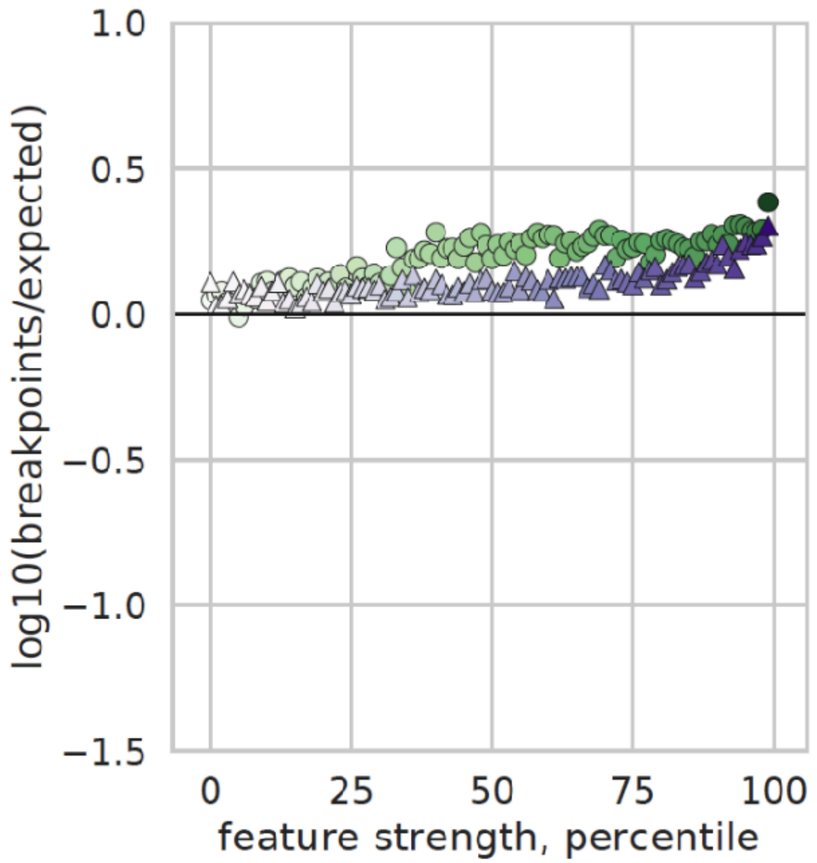


Hi-C Data: Rao et al 2015
Expression Data: GTeX

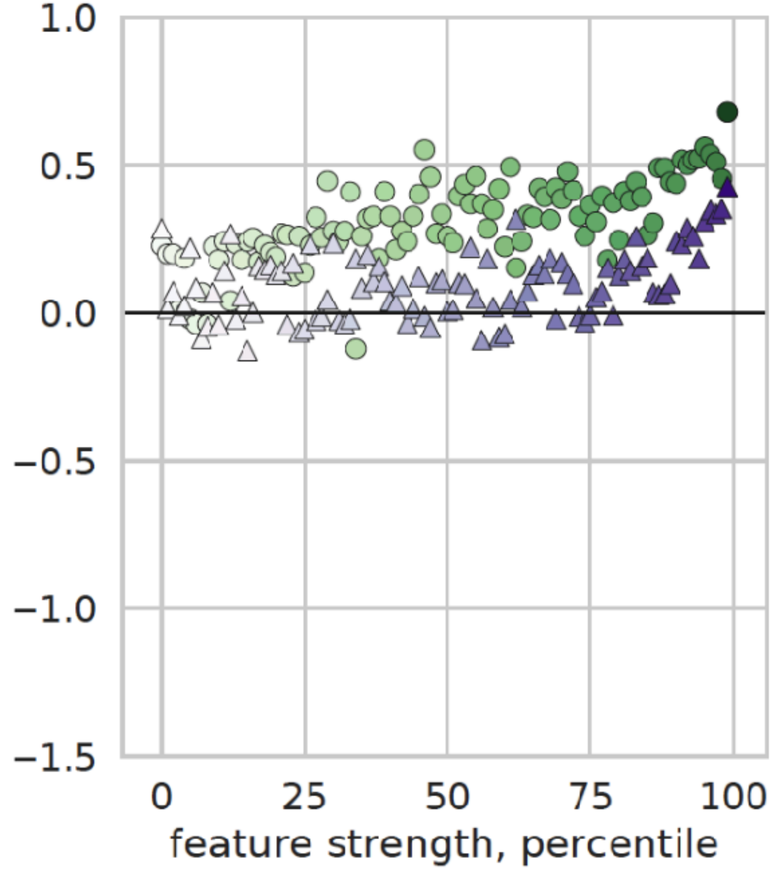
Genomic Element Data: ENCODE, Epigenomics Roadmap
Structural Variant Data: Apes: Sudmant 2013; Controls, Cases: Coe, 2014

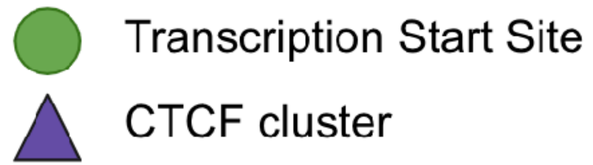
Deletions enriched in cancer, CHD?

Tumor Genomes



Heart Defect Patients





Cancer Structural Variant Data: COSMIC
Congenital Heart Defect Variant Data: PCGC
Genomic Element Data: ENCODE, Epigenomics Roadmap

Hi-C Data: Rao et al 2015
Expression Data: GTeX

Idea: predict how mutations change chromatin interaction maps

Training

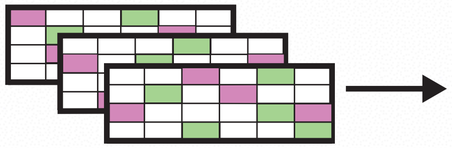
genomic sequences (inputs)

ATCAGGGAAATTTCCACTAGTTTAGGGTAATAA

} x 75,000

ATCCACTAGTTTAGGGTAATAACAGGGAAATTT

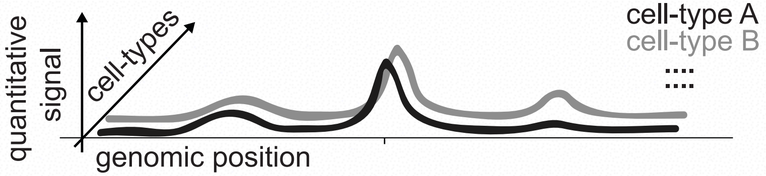
convolution layers, pooling



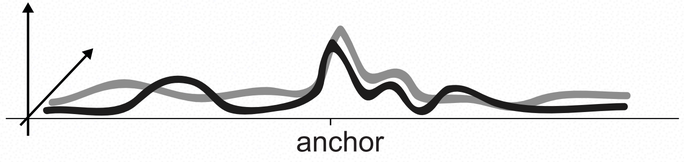
genomic position

dense layers, LSTM layers

experimental 4C profiles (targets)



} x 75,000 x # cell-types



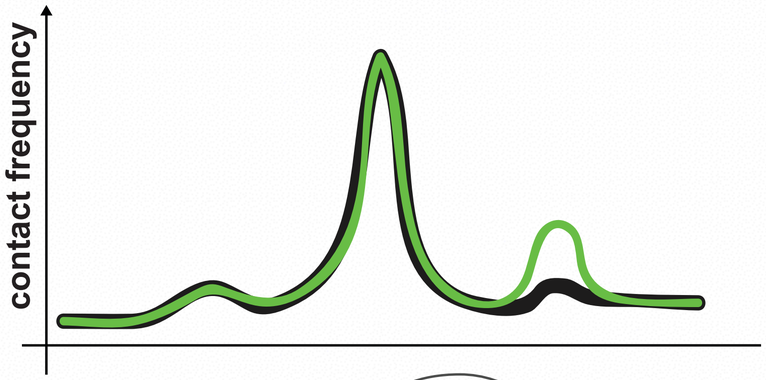
Predictions

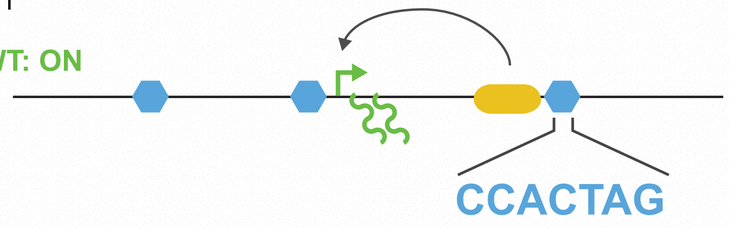
genomic sequences (inputs)

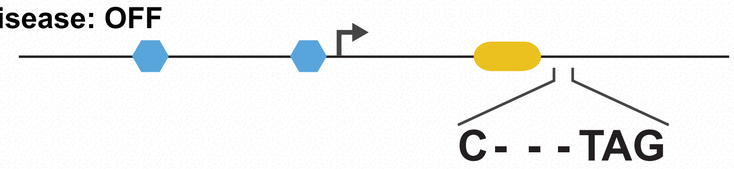
WT: CTCF motif ..... CCACTAG .....

disease: deletion ..... C- - -TAG .....

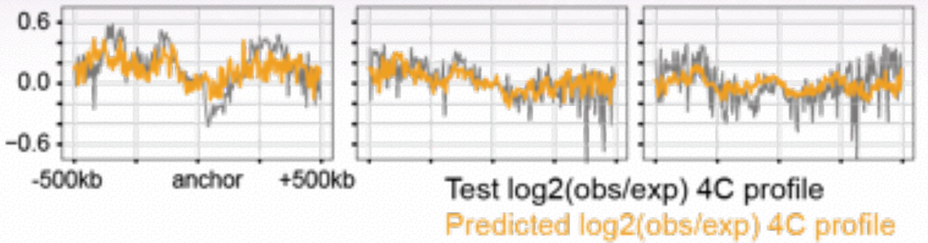
predicted impact of variant

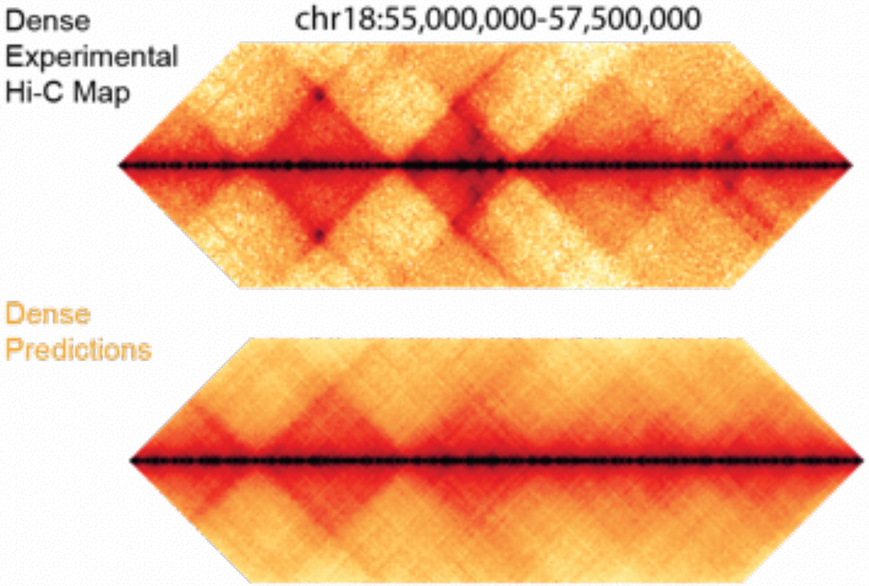






Preliminary Results





Conclusions

- Mutations that delete TAD boundaries are strongly selected against in primates and healthy people, but not patients, suggesting a broad role for enhancer hijacking in disease.
- Non-coding mutation scoring tools should be TAD aware.
- Preliminary results suggest that effects of variants on chromatin interaction maps can be predicted from epigenetic data and potentially from sequence alone.
- If so, this opens the door to identifying causal variants that function by changing chromatin structure.

Chromatin Structure Meets Population Genetics

What is the relationship between genetic variation and chromatin structure?

1. Open chromatin in the human developing telencephalon sheds light on non-coding mutations in autism.

   Eirene Markenscoff-Papadimitriou, Pawel Przytycki & Sean Whalen

2. Chromatin boundaries are under strong negative selection.

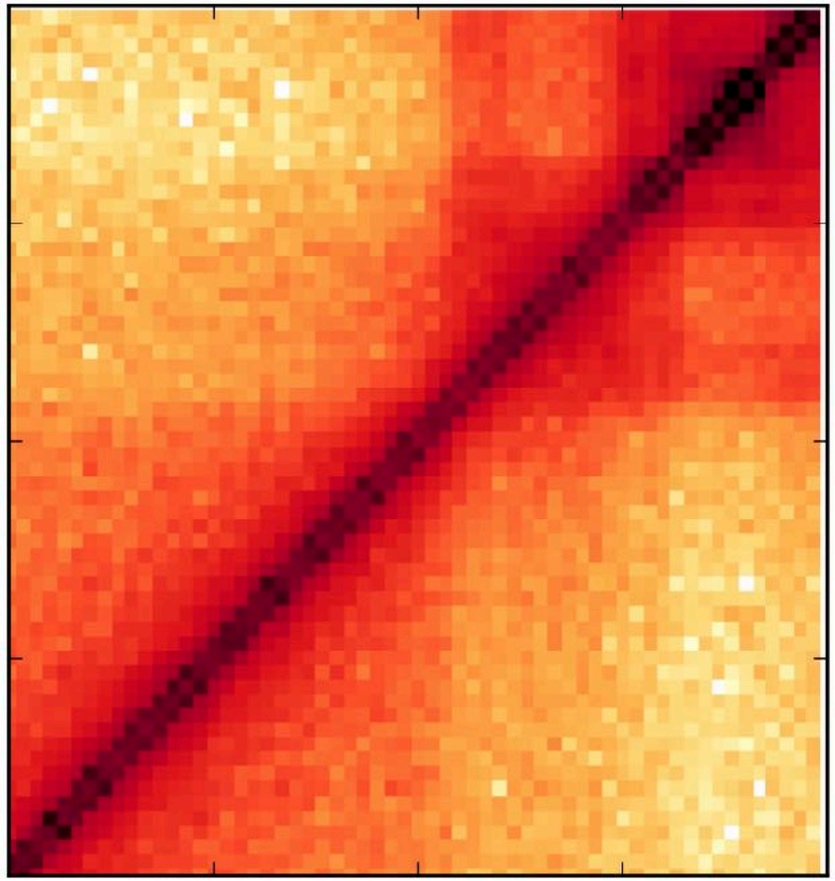
   Geoff Fudenberg

3. Chromatin interactions and linkage disequilibrium are uncorrelated along the human genome.

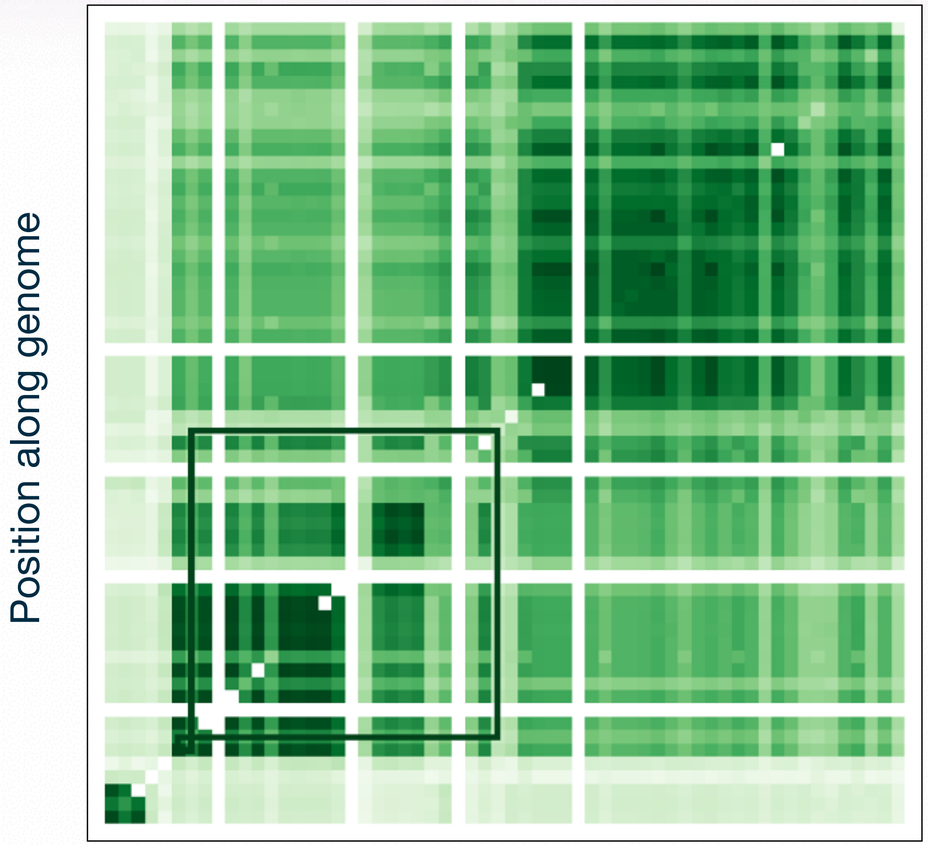
   Sean Whalen

Chromatin & genetic interaction maps both have block structure

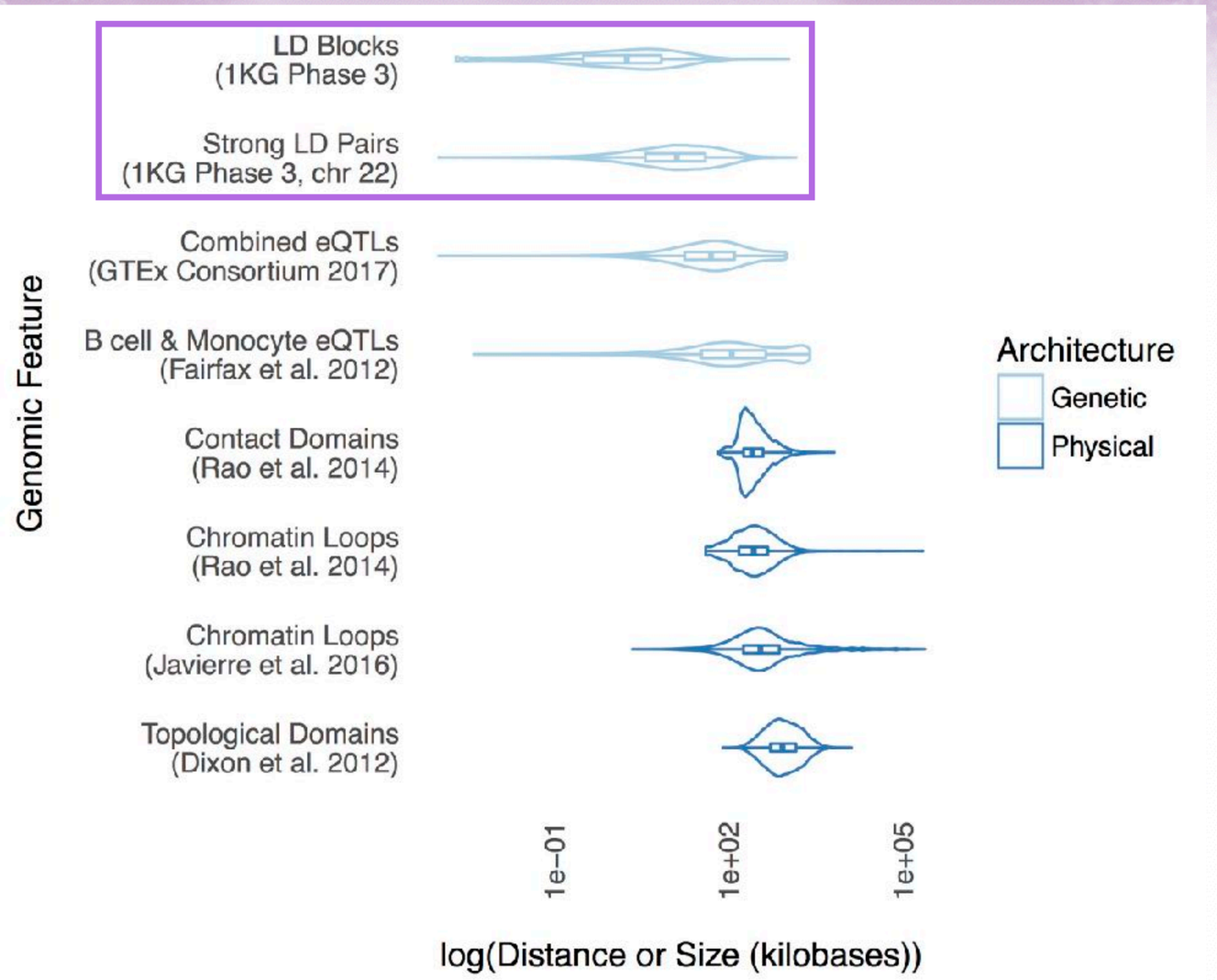
Chromatin Contact Frequency



Linkage Disequilibrium (LD)

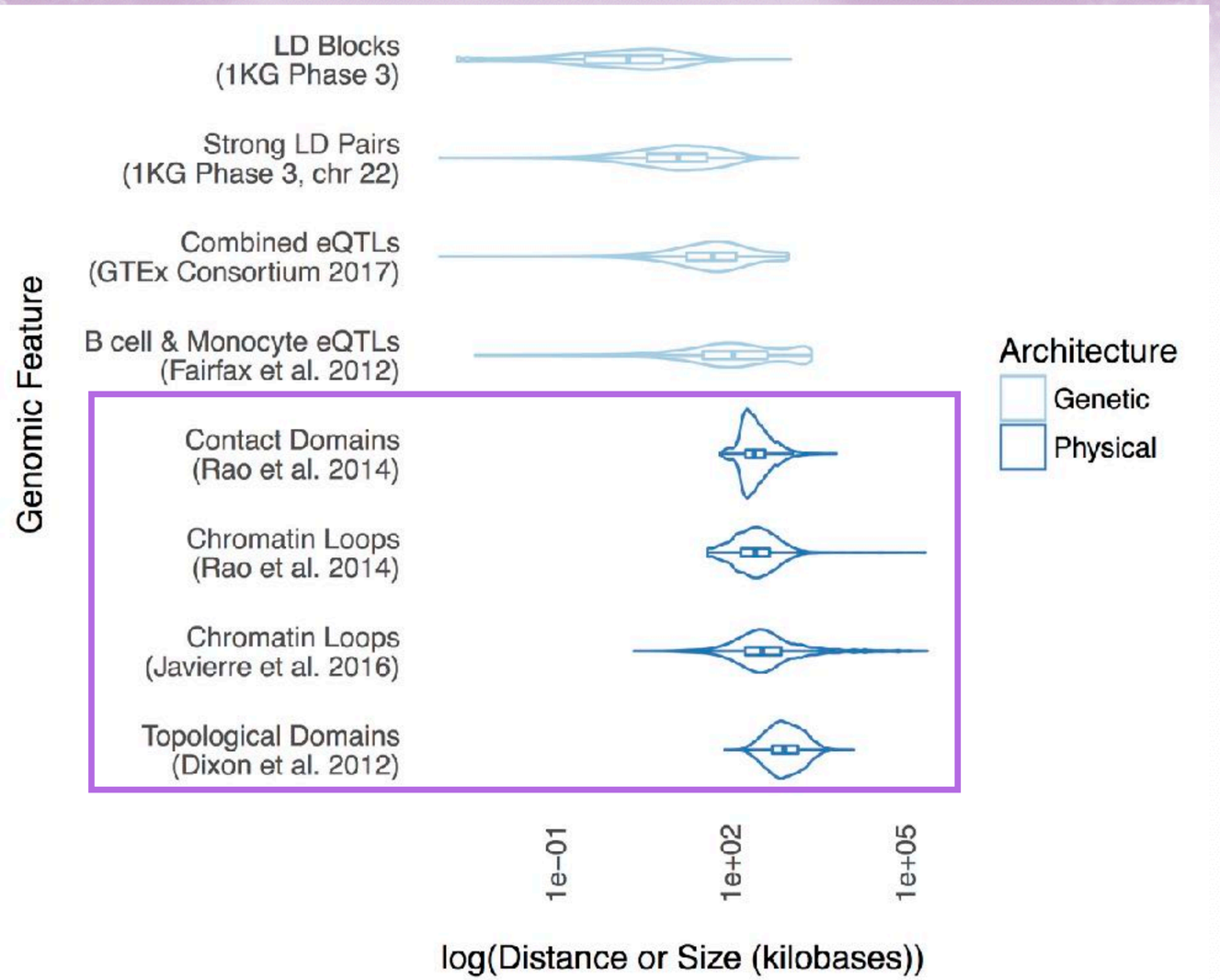


Genetic and physical interaction maps are uncorrelated and have different scales



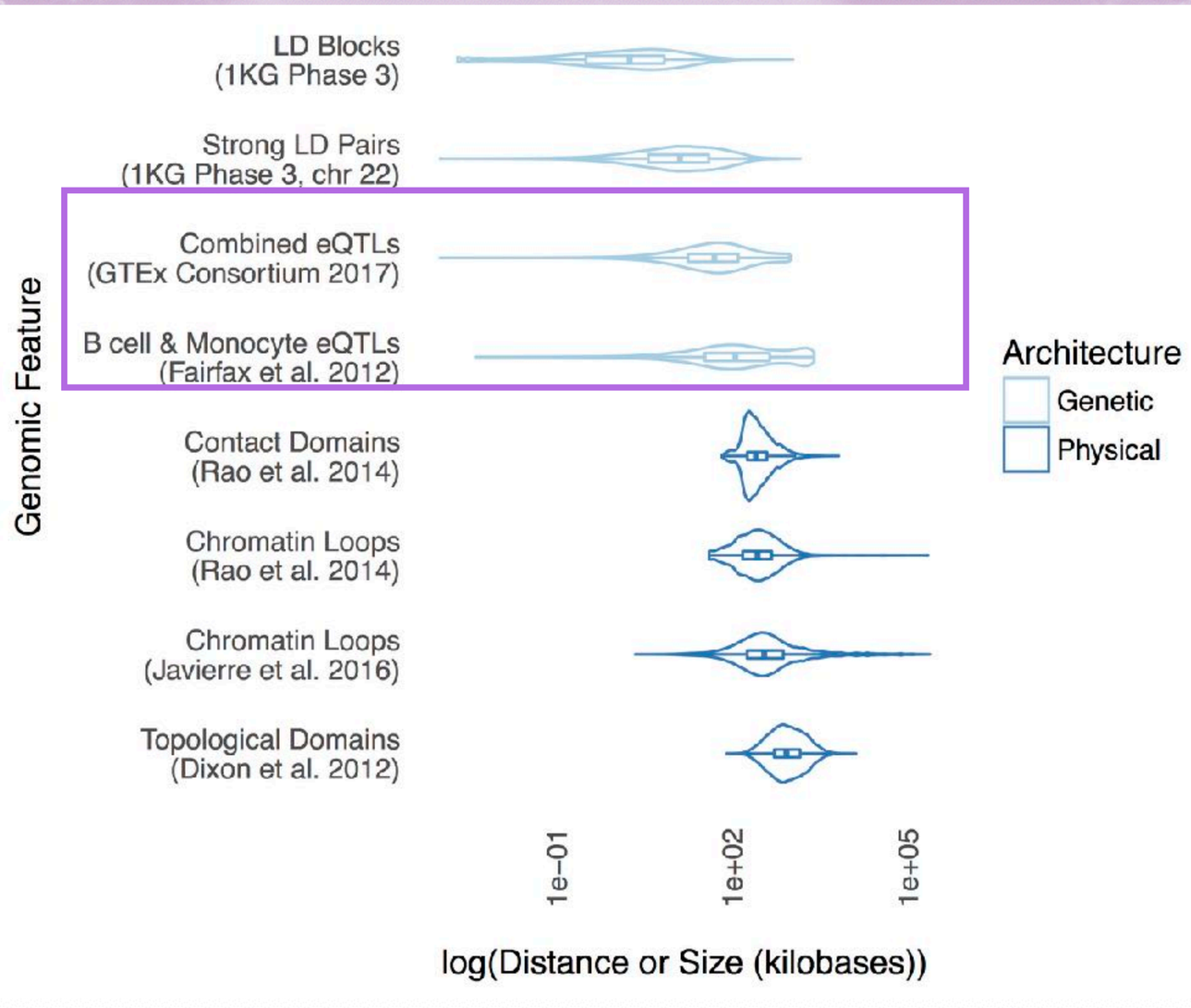
27 billion SNP pairs (1KGP),
1.6 million LD blocks,
3.1 million chromatin interactions
across 22 cell types (Hi-C, PCHi-C)

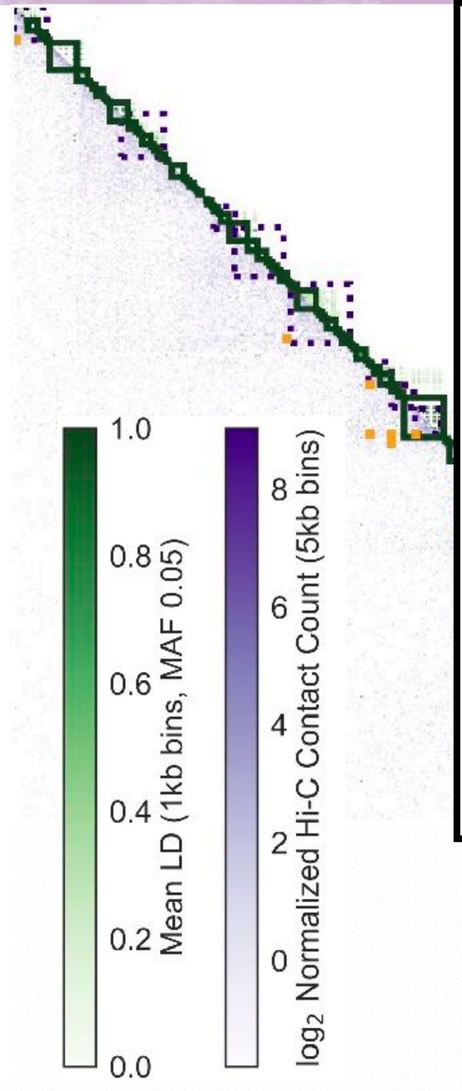
Genetic and physical interaction maps are uncorrelated and have different scales

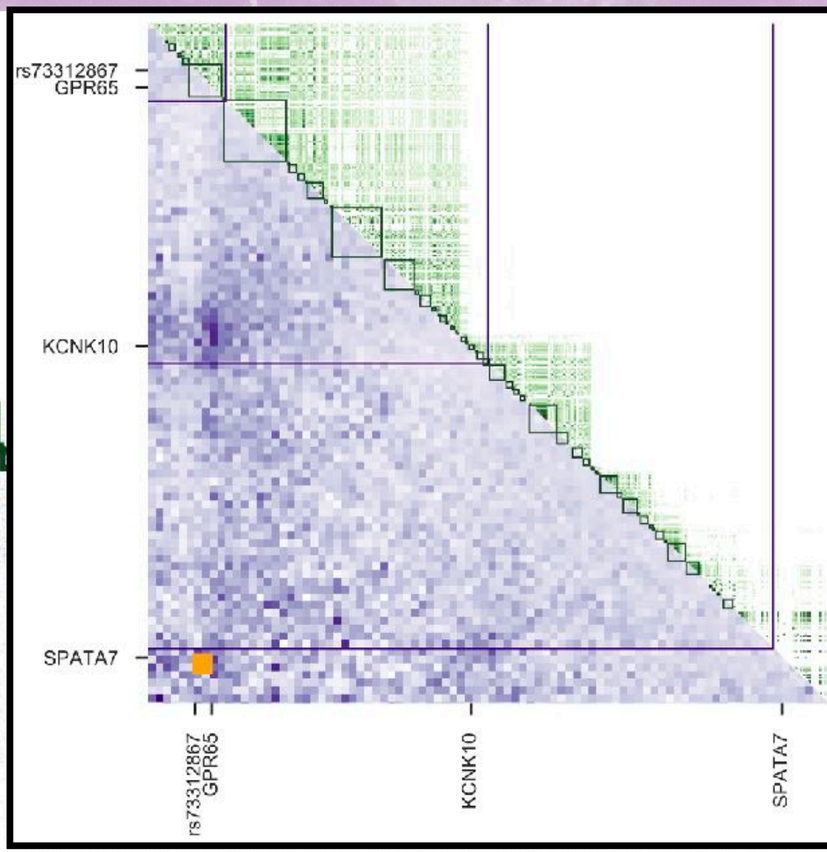


27 billion SNP pairs (1KGP),
1.6 million LD blocks,
3.1 million chromatin interactions
across 22 cell types (Hi-C, PCHi-C)

Genetic and physical interaction maps are uncorrelated and have different scales

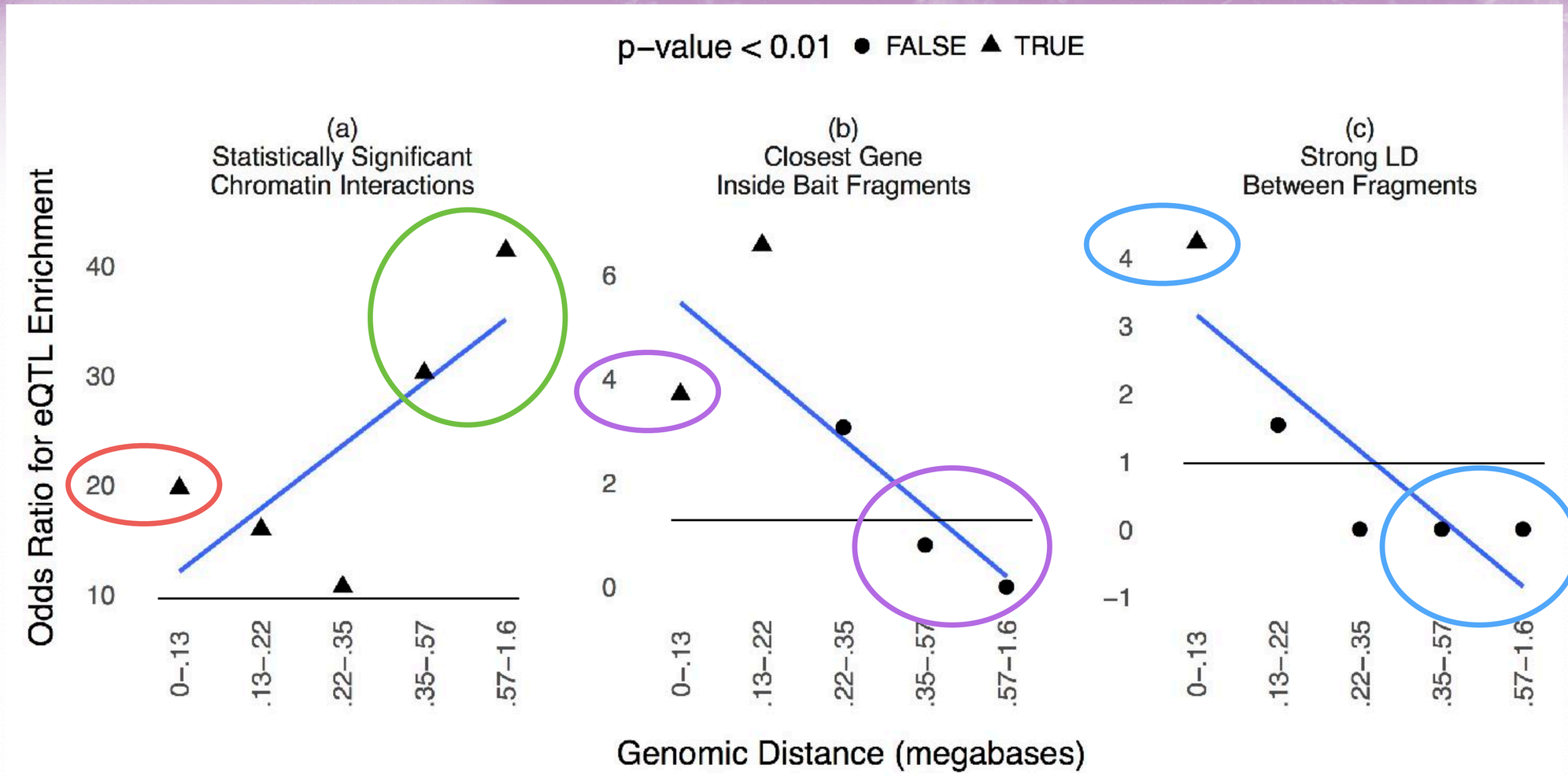






27 billion SNP pairs (1KGP),
1.6 million LD blocks,
3.1 million chromatin interactions
across 22 cell types (Hi-C, PCHi-C)

Chromatin interactions are more enriched for eQTLs than are closest gene or LD



eQTLs: Fairfax et al. 2012 (B-cells)
PCHi-C: Javierre et al. 2016 (B-cells)

Interacts ~10% of time

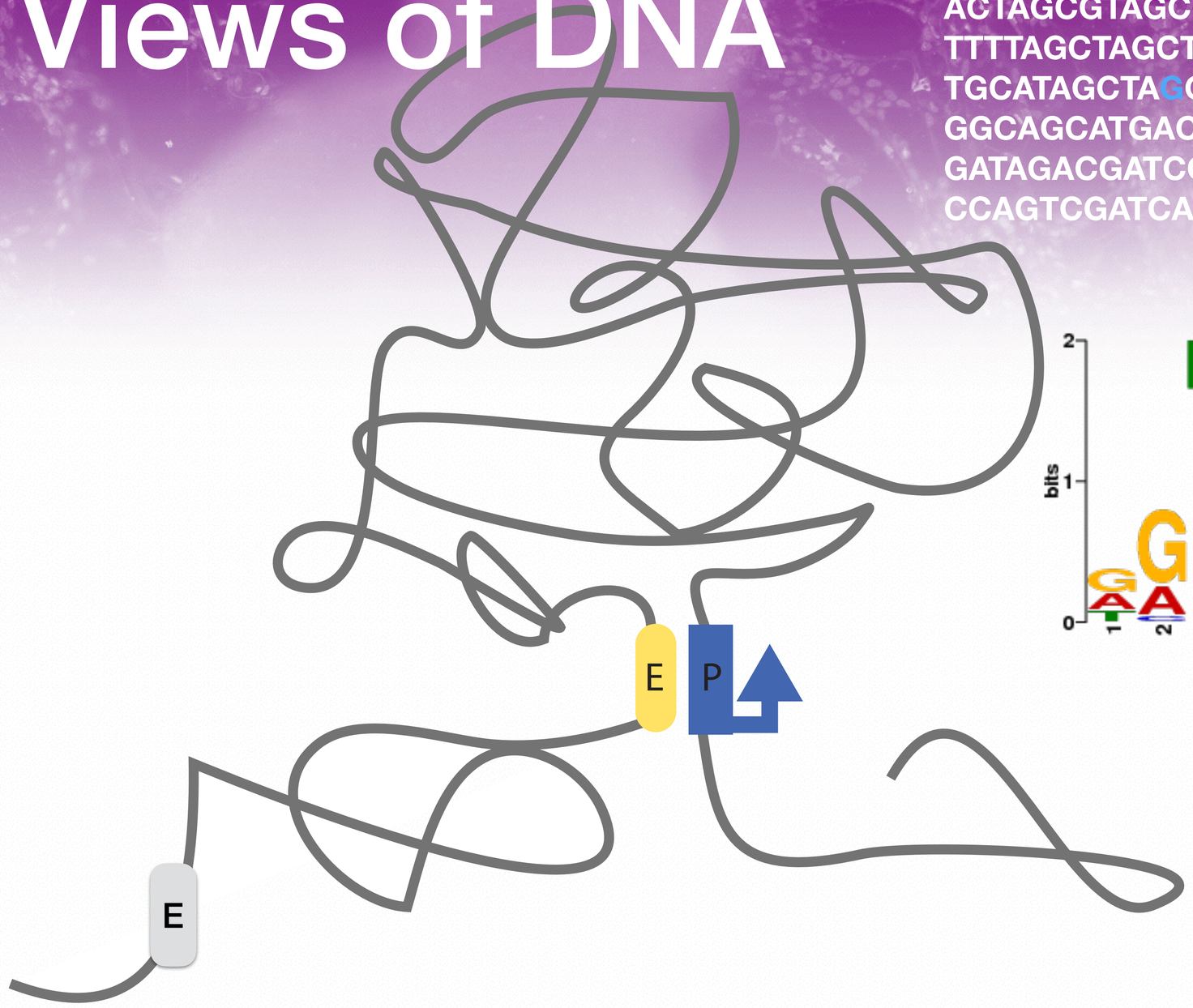
Interacts ~1-7% of time

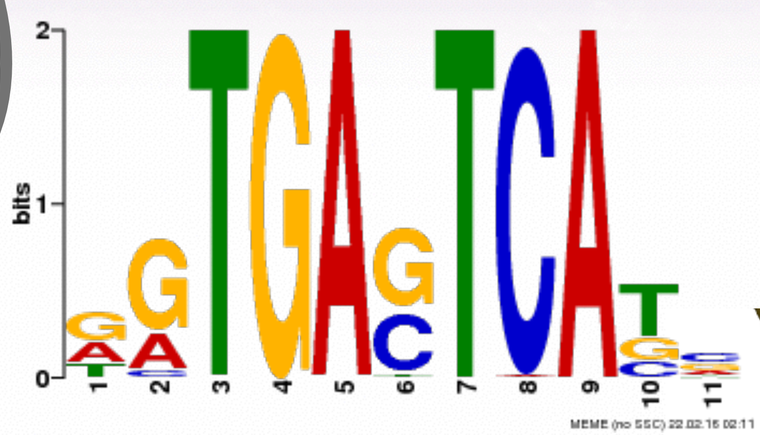
Conclusions

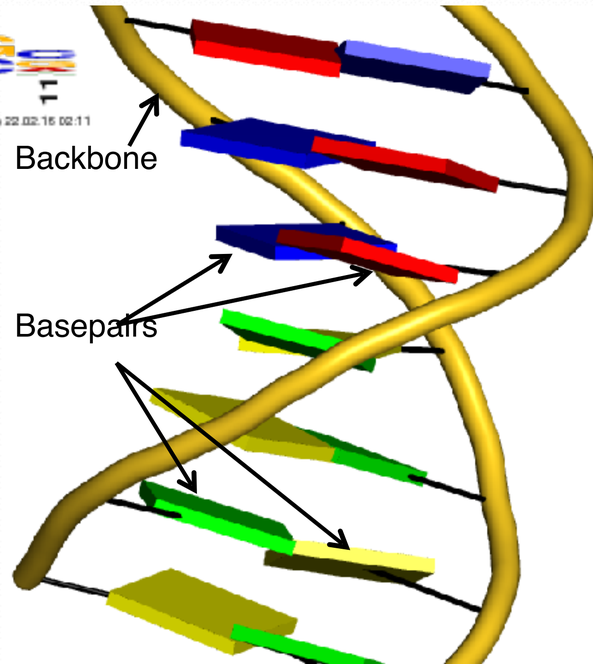
- Chromatin interactions and genetic interactions both have nested block structures in the human genome. BUT these are completely uncorrelated at scales >5Kb for interphase Hi-C.
- Most distal (>5Kb) non-coding variants do not target the closest expressed gene, and they are not in LD with their target genes.
- Linked SNPs can be in different chromatin domains. eQTLs and their target genes are often in the same one, but have LD=0.
- While TAD locations and gene content are conserved across evolutionary time, recombination is low at BEs and breaks up linkage within TADs as they segregate in human populations.
- Ongoing work: What about meiotic Hi-C maps?

Views of DNA

ACTAGCGTAGCTAGCGATATCTAGGGCGATCGATGCTACGTATCGAGC
TTTTAGCTAGCTAGCTAGCATCGATGCATCGATCGTACGATCGATCGTA
TGCATAGCTAGCTAGCATGCATGCATCGATCGAATCGATATTAGCTAGC
GGCAGCATGACTAGTCAGATATCGTACGATGTCGAAAACTGATCAGTC
GATAGACGATCGATCGATCGATCGAGGCGCATCGATCGATGCTAGCAT
CCAGTCGATCAGTCGATCGATCGATCGATCGATCGATCGACTAGATCG







Acknowledgements

Collaborators: Eirene Markenscoff-Papadimitriou, John Rubenstein, Tom Nowakowsi, Stephan Sanders, Stephanie Schalbetter, Jonathan Baxter, Matt Neale

pollard lab

Funded by NIH, NSF, Gladstone, UCSF, Moore Foundation, Chan-Zuckerberg Biohub