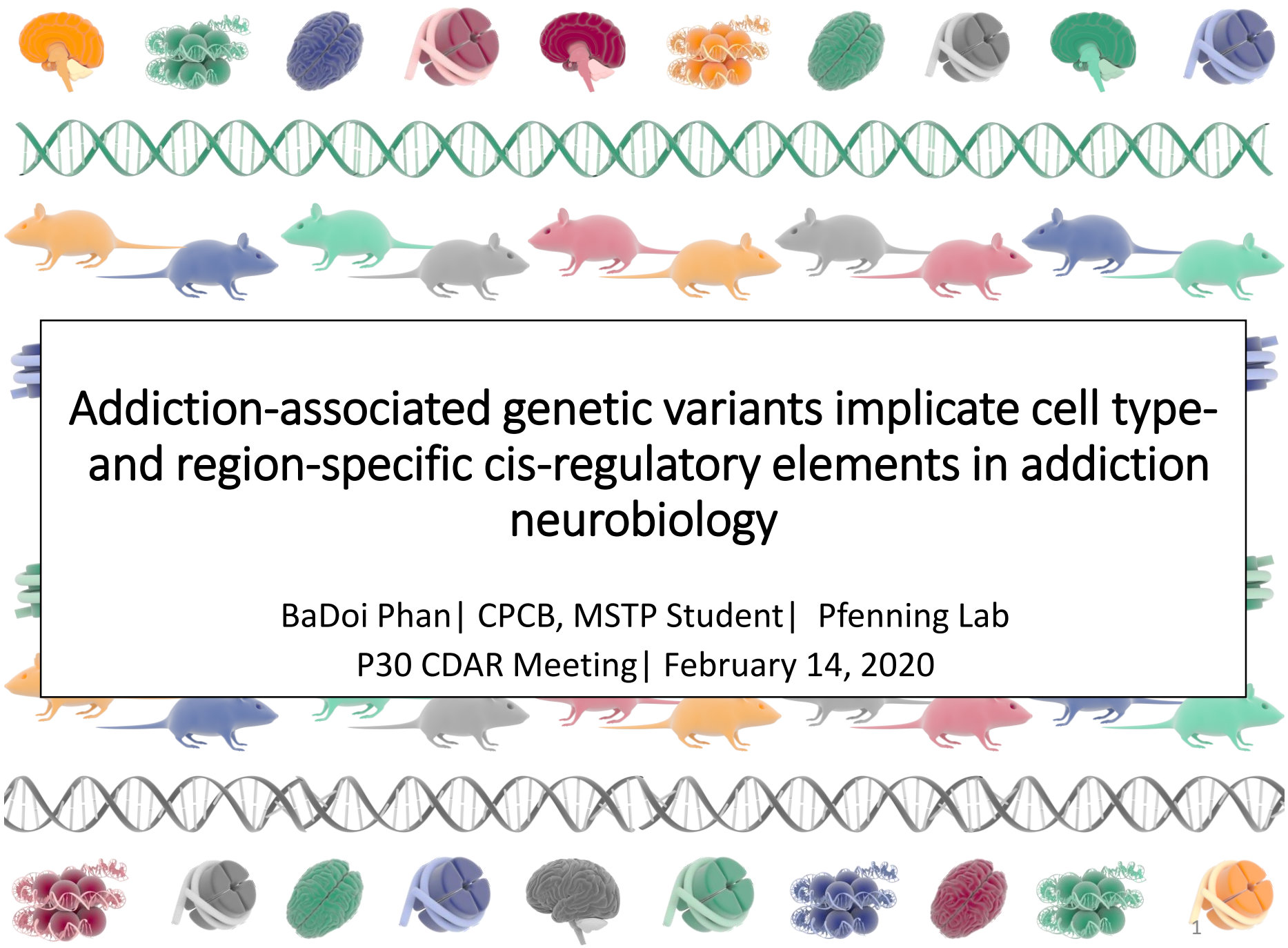
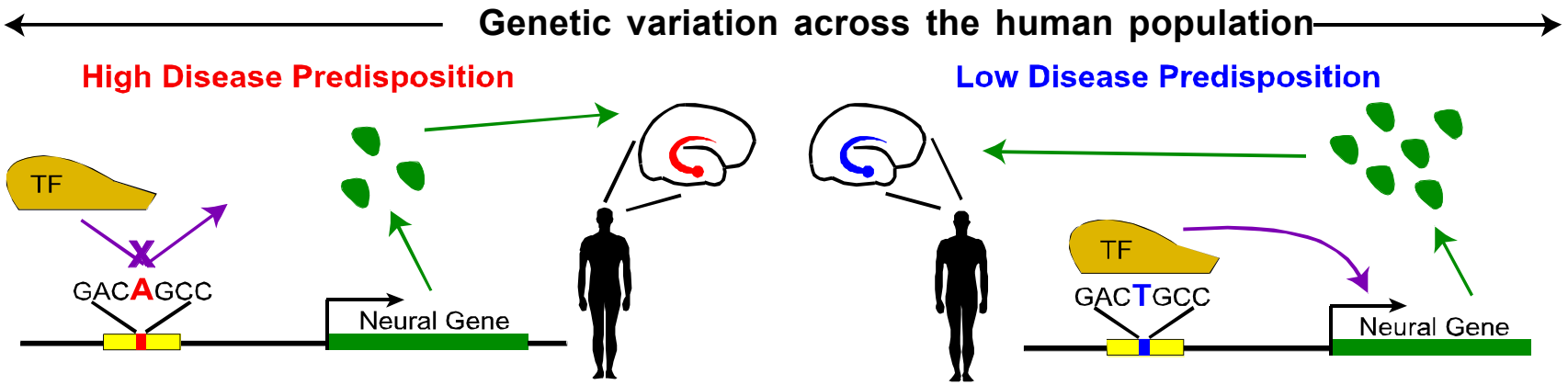


Addiction-associated genetic variants implicate cell type- and region-specific cis-regulatory elements in addiction neurobiology

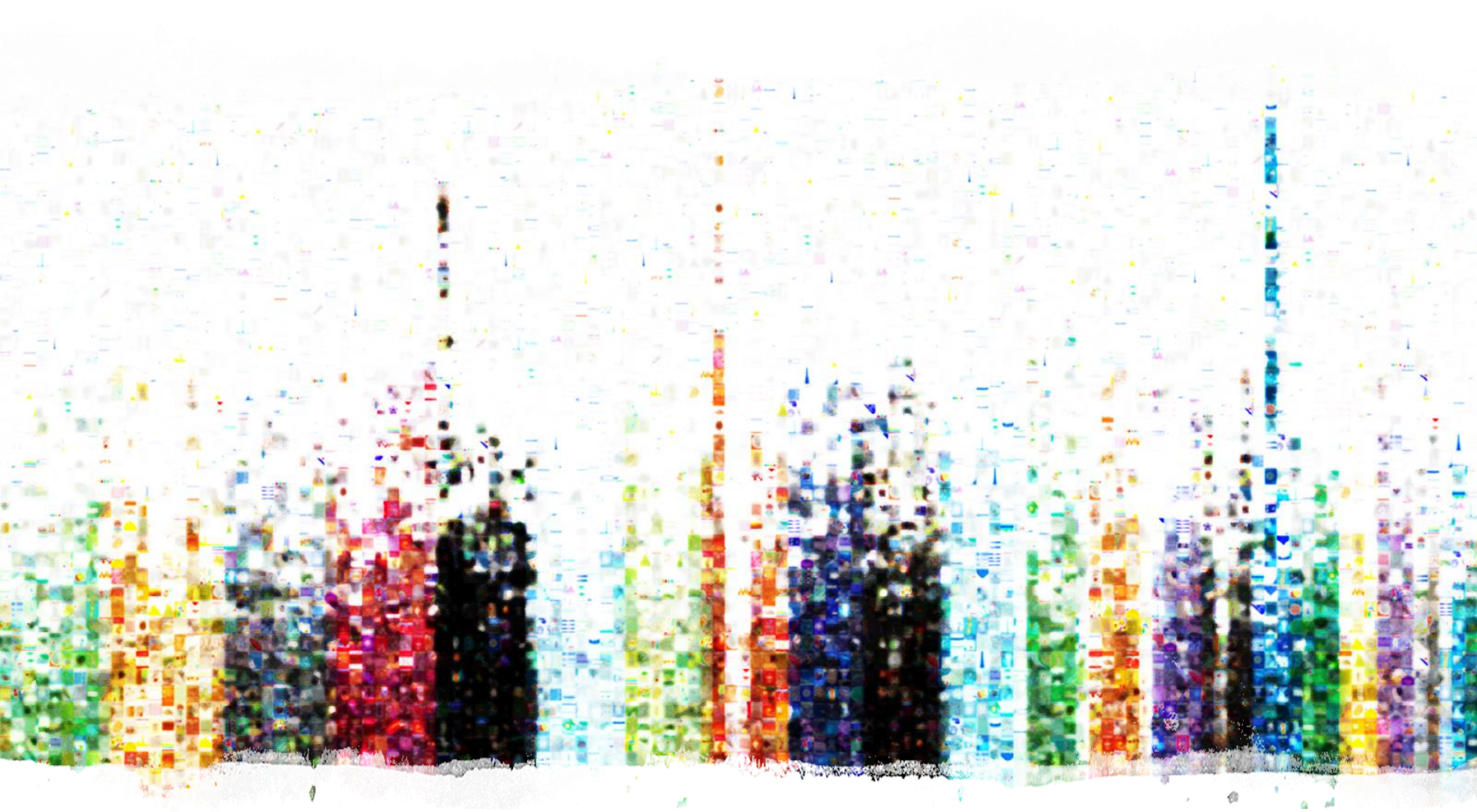
BaDoi Phan| CPCB, MSTP Student| Pfenning Lab

P30 CDAR Meeting| February 14, 2020

How do genetic variants influence the predisposition to substance use disorders?



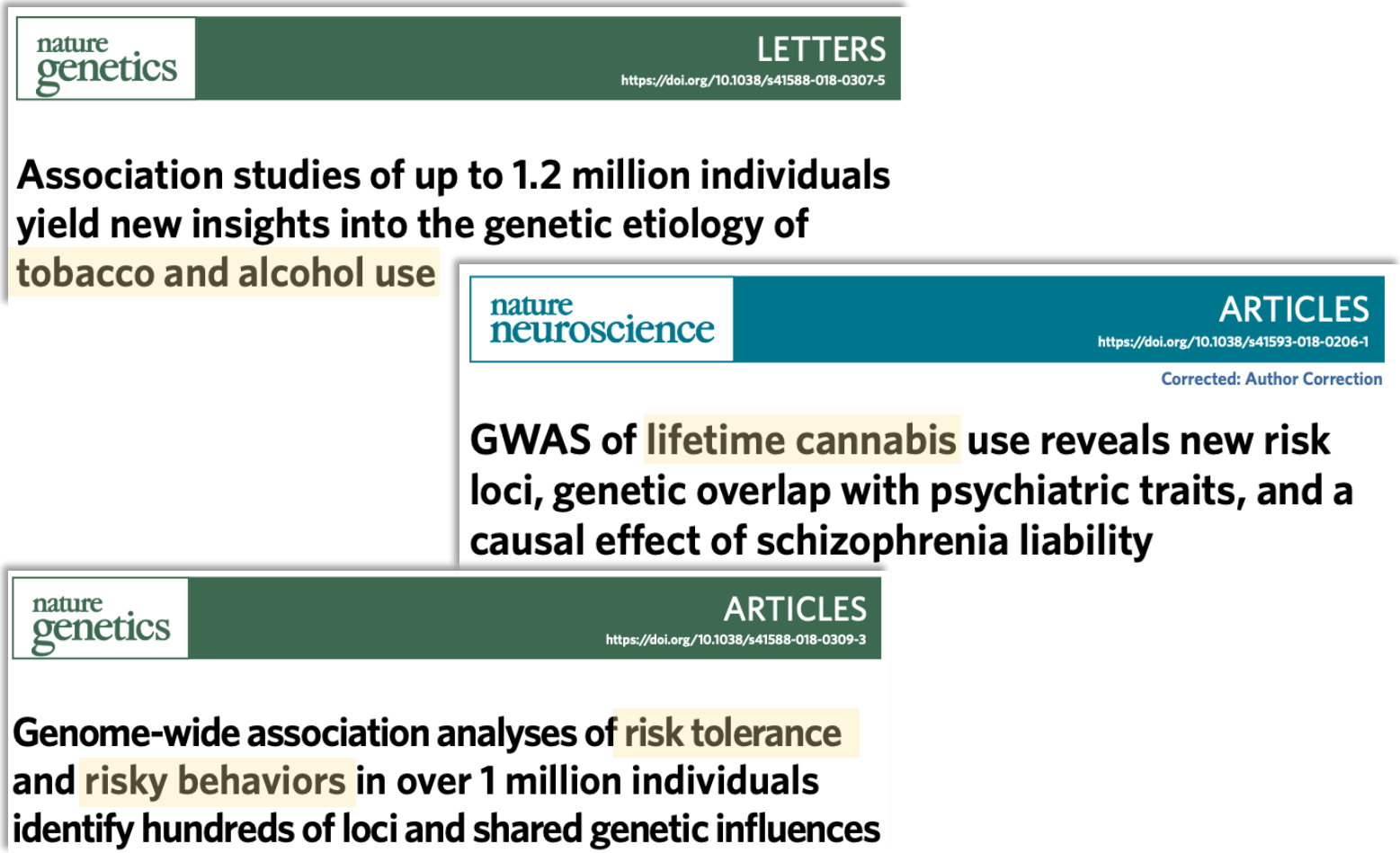
We use computational models and experimental genomic tools to predict how a mutation is going to impact the neural circuit



Exploring genetic signal of substance use phenotypes.

1. Addiction GWAS encode cell type-specific reward pathways

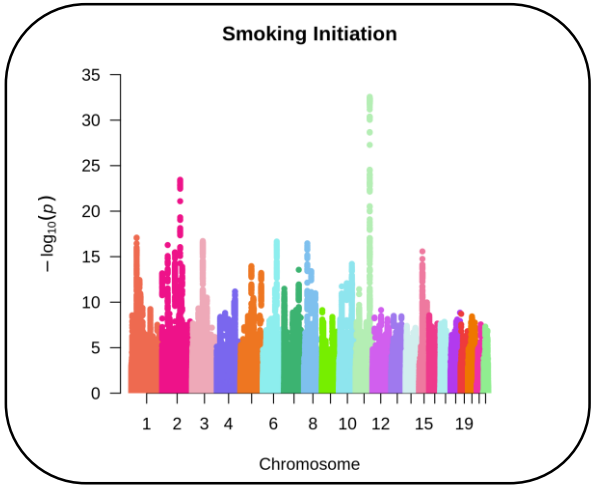
Interpreting multiple large, confident addiction-associated risk loci



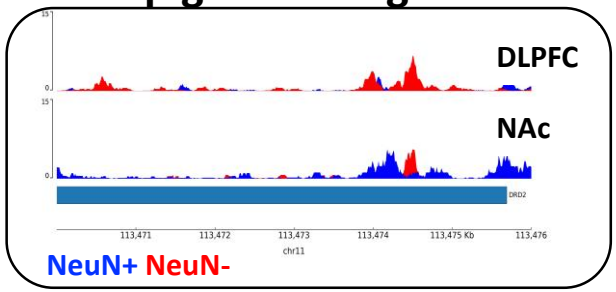
Liu et al. Nat Genetics 2019; Pasman et al. Nat Neuroscience 2018; Linner et al. Nat Genetics 2019

LD score regression identifies cell type-specific enhancers enriched in addiction-associated GWAS

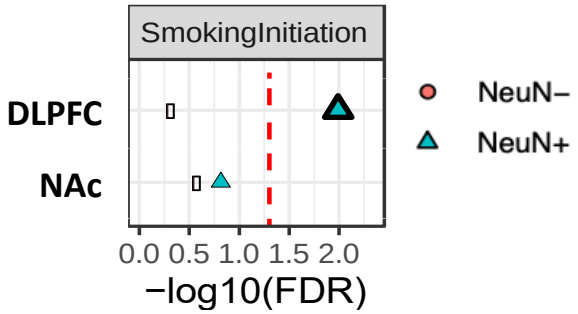
GWAS Summary statistics



Epigenetic Regions



LD score regression

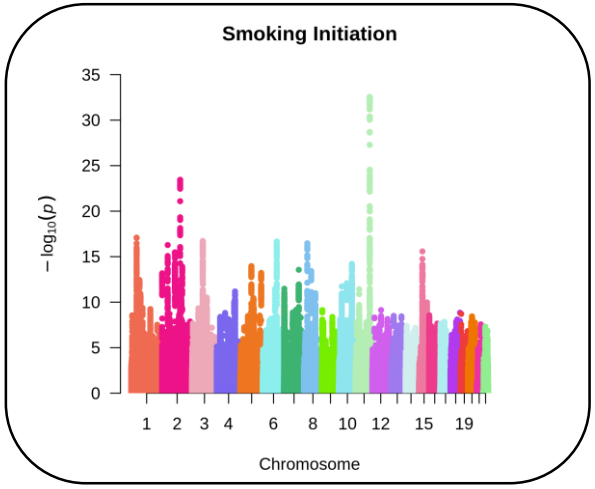


- Partition GWAS variants by cell-type specific epigenetic regions
- Cell type-specific enrichments informs the neurobiology of substance use addiction

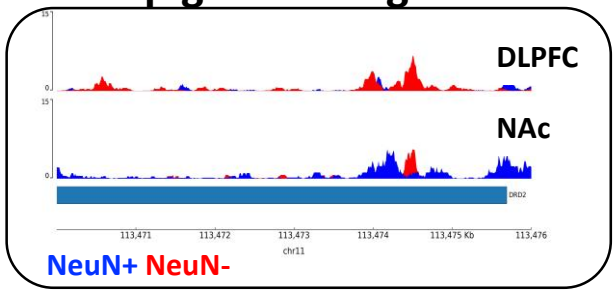
Finucane et al., Nat Genetics 2015

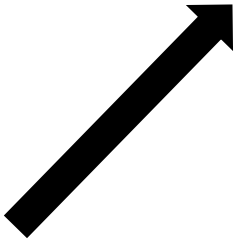
LD score regression identifies cell type-specific enhancers enriched in addiction-associated GWAS

GWAS Summary statistics

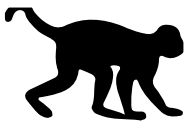


Epigenetic Regions





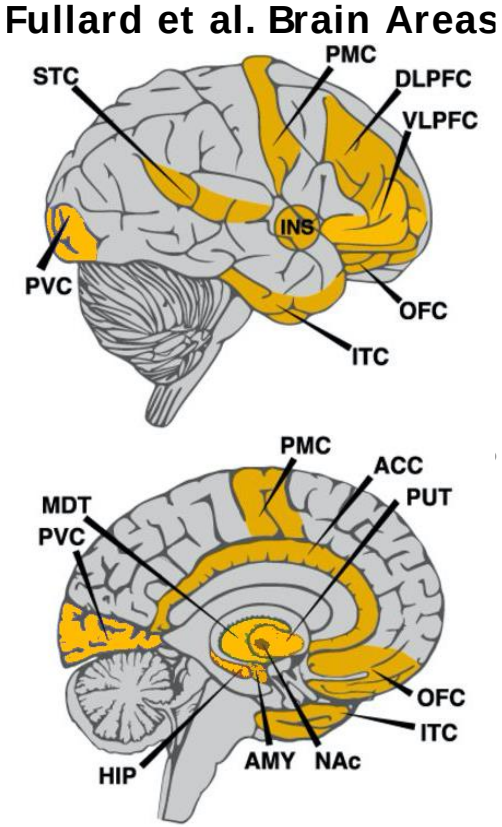
Brain Regions
Occipital Cortex Cell Types

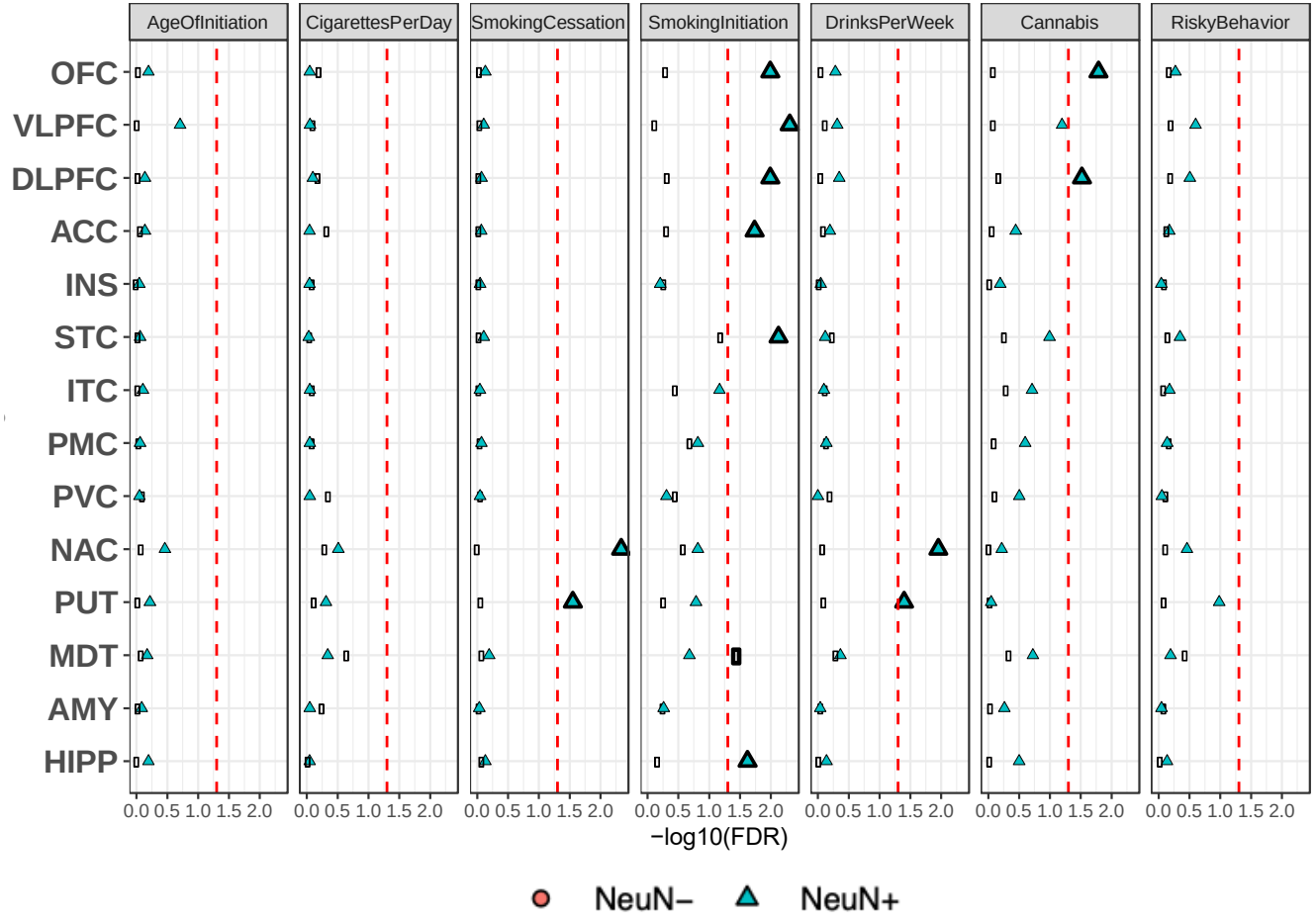


Striatum Cell Types

Targeted Epigenetics

Addiction-associated GWAS variants enrich in adult human atlas of NeuN+/- ATAC-seq peaks

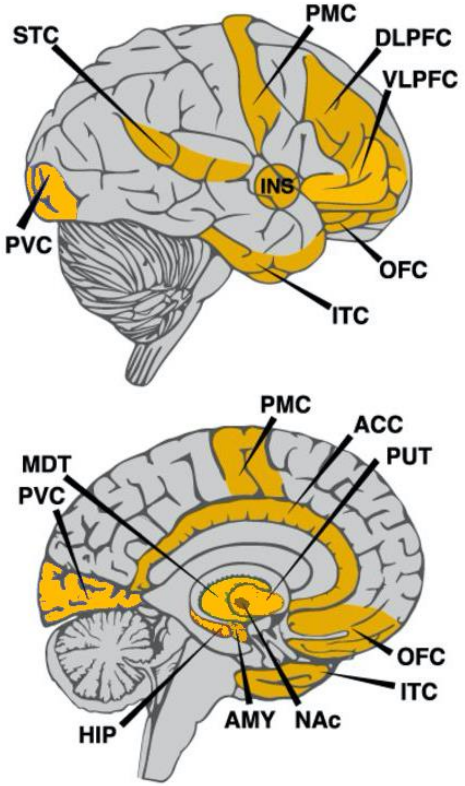




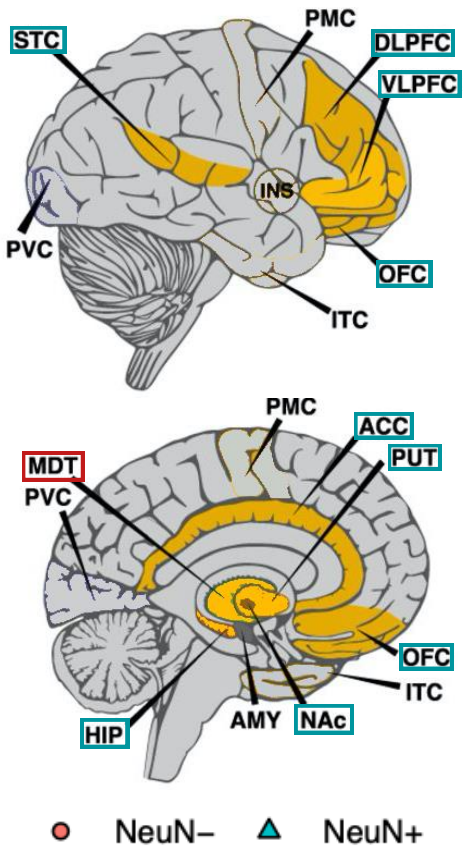
Fullard et al., Genome Res 2018

Addiction-associated GWAS variants enrich in adult human atlas of NeuN+/- ATAC-seq peaks

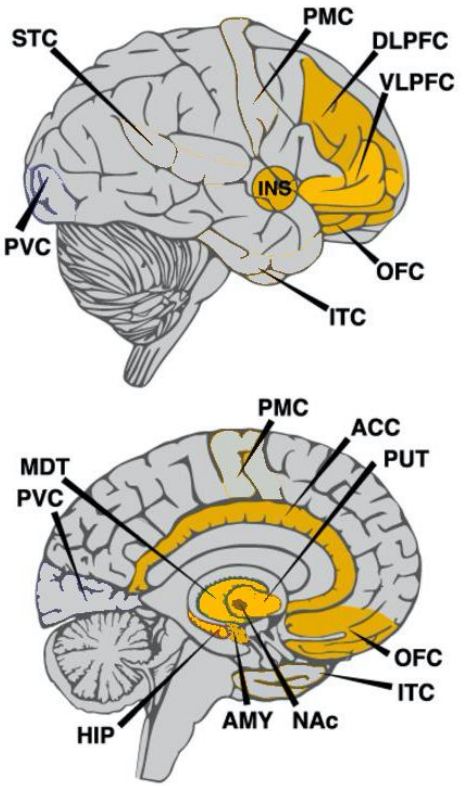
Fullard et al. Brain Areas



GWAS LDSC- Enriched Areas

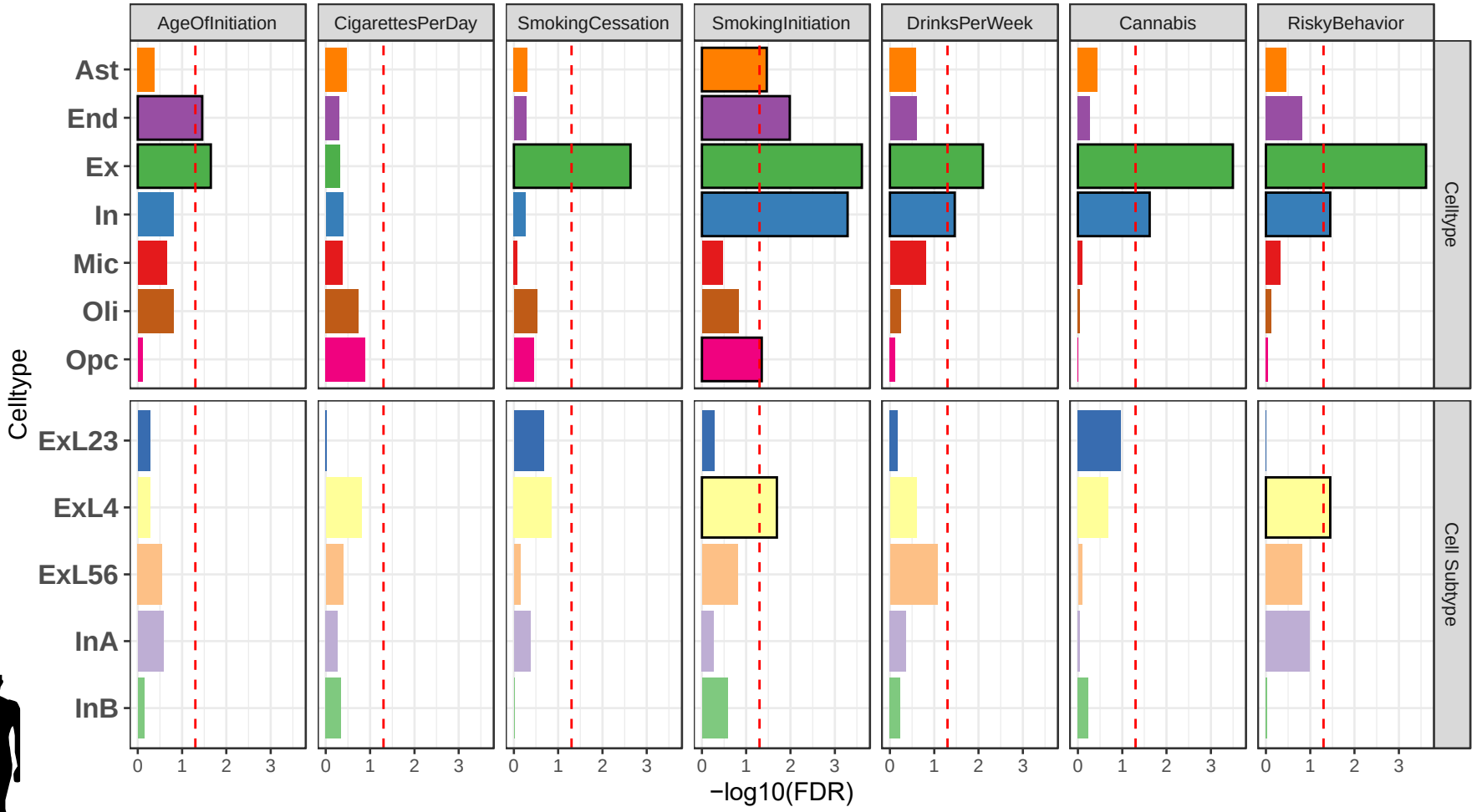


Reward- Addiction Areas



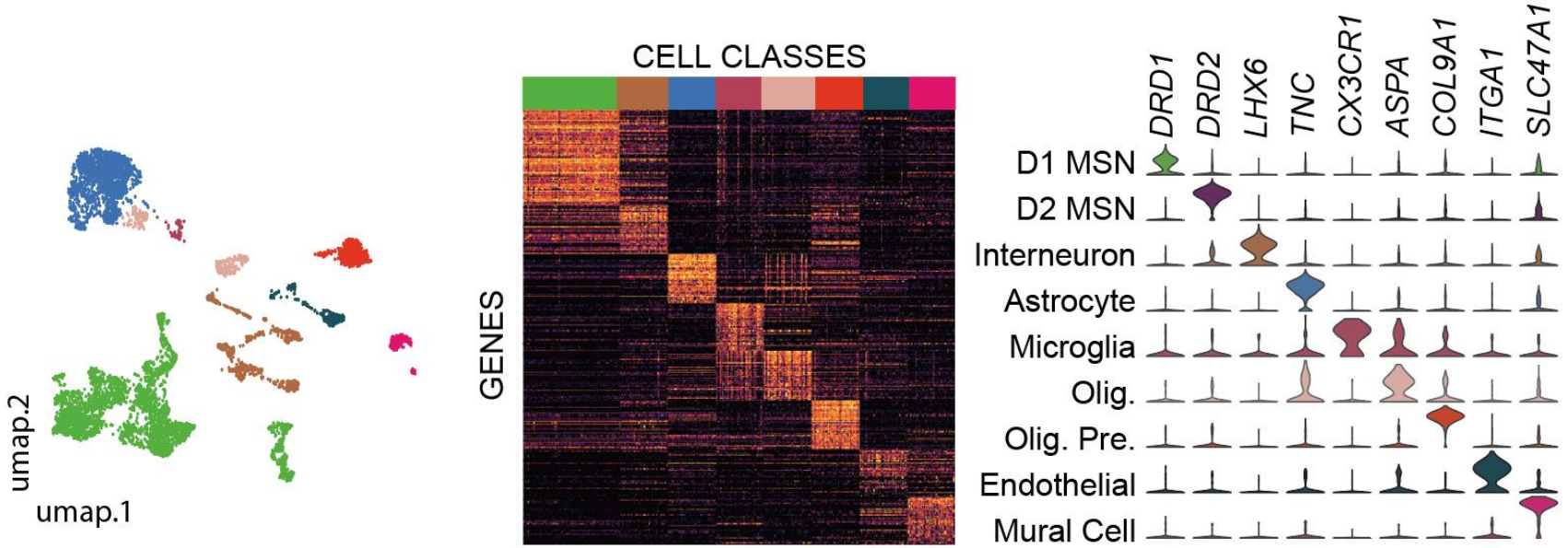
Fullard et al., Genome Res 2018; Volkow and Morales, Cell 2015

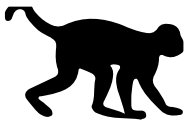
GWAS variants enrich for in single nucleus open chromatin of human occipital cortex (scTHS-seq)



Lake et al., Nat Biotech 2018

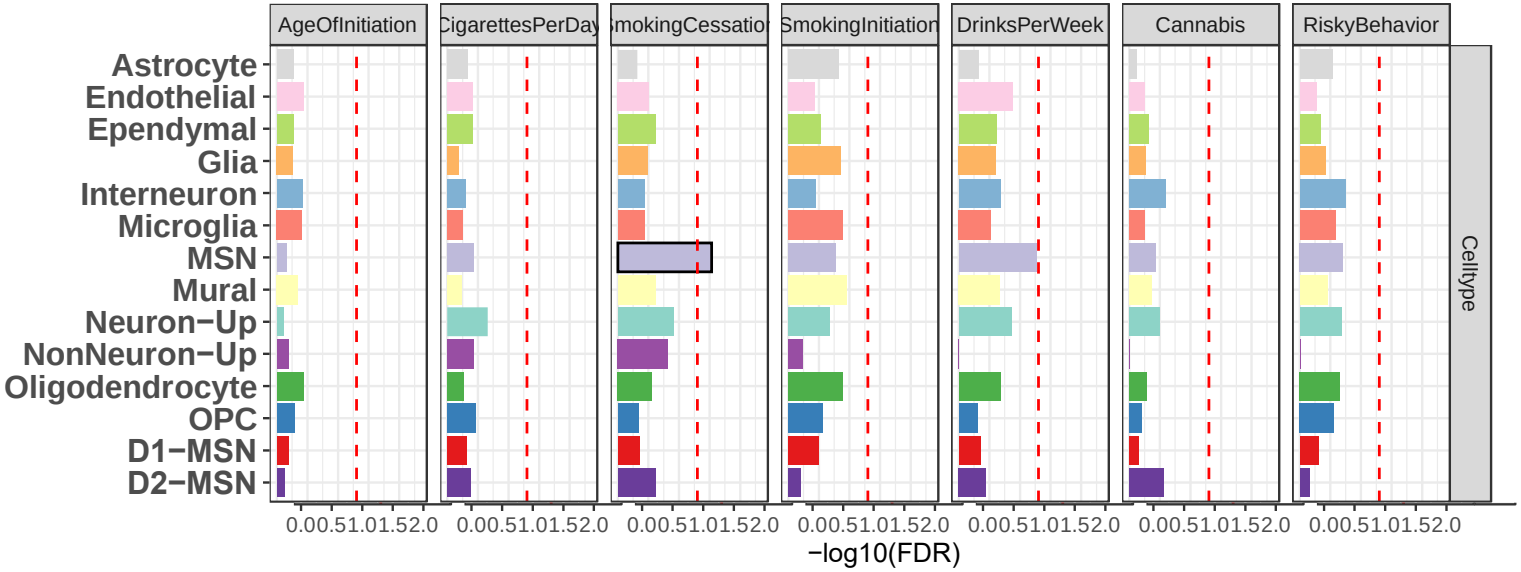
GWAS enrichment of macaque striatum cell type markers identified by single-nucleus RNA-seq

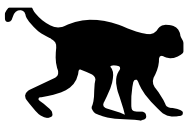




Stauffer and Byrne lab, data unpublished

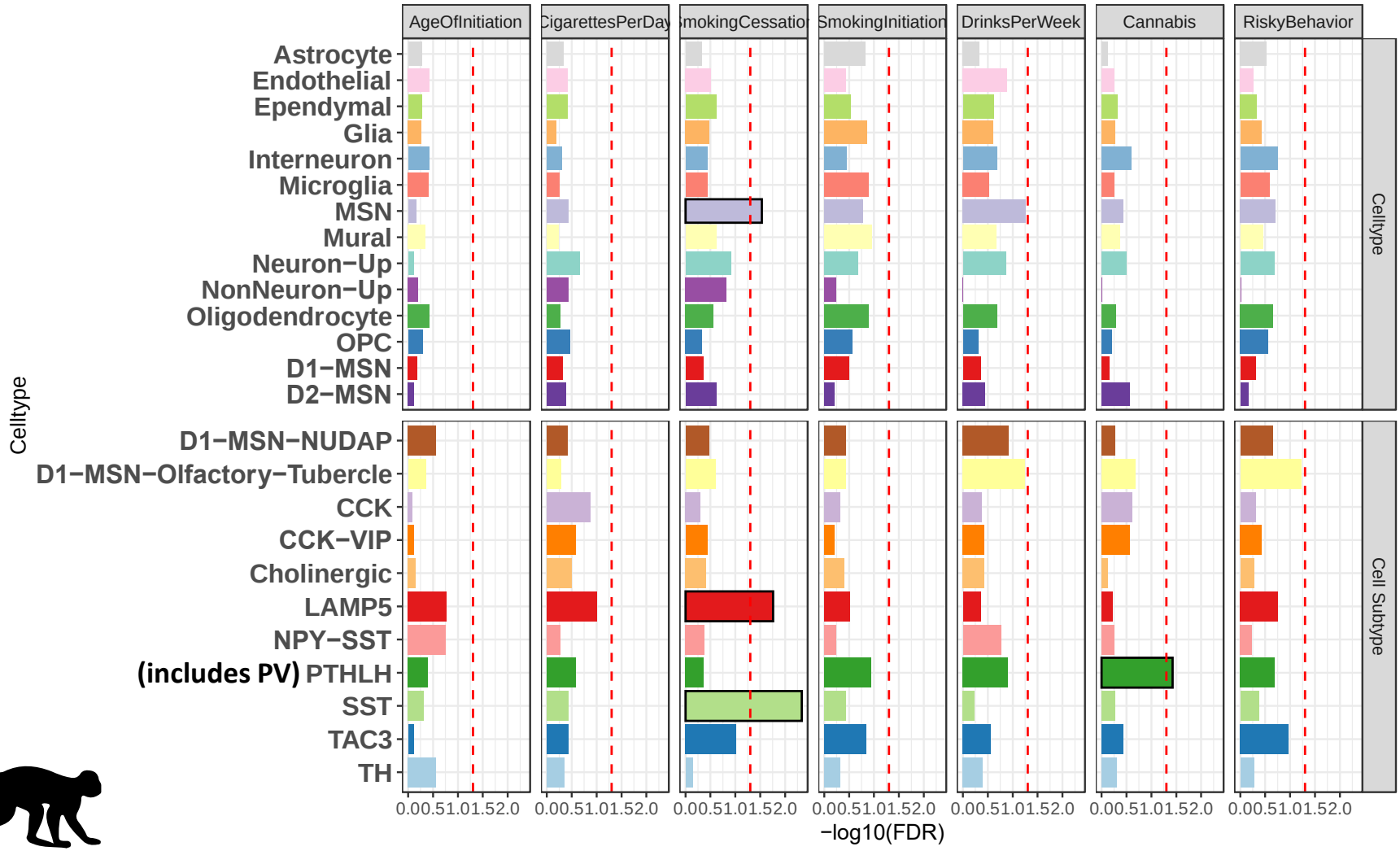
GWAS enrichment of macaque striatum cell type markers identified by single-nucleus RNA-seq





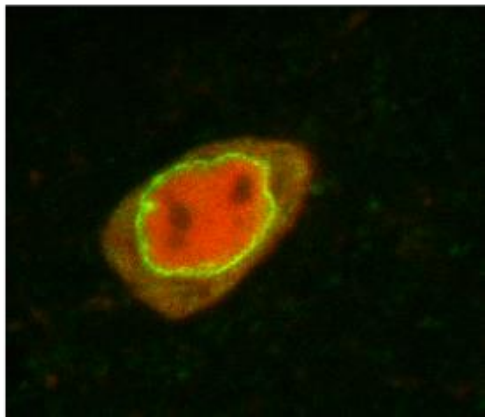
Stauffer and Byrne lab, data unpublished

GWAS enrichment of macaque striatum cell type markers identified by single-nucleus RNA-seq



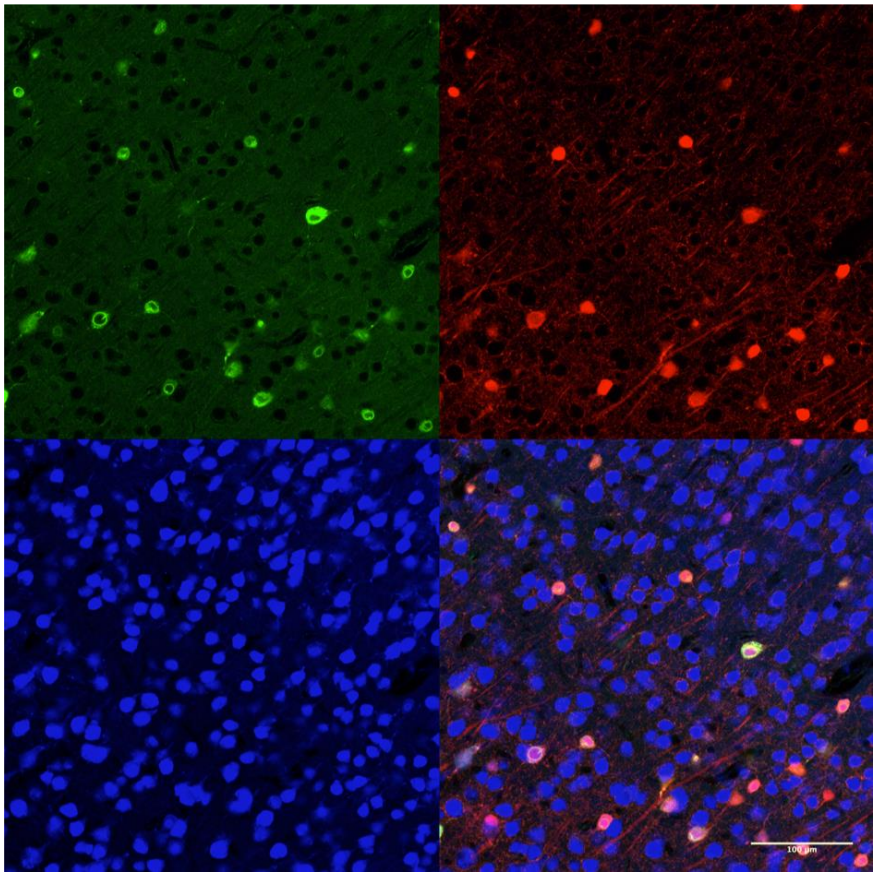
Stauffer and Byrne lab, data unpublished

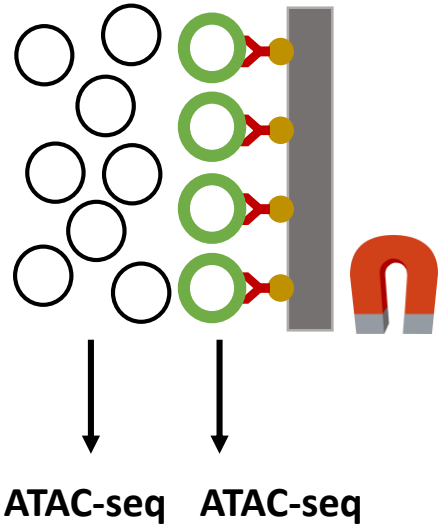
Cell type-specific *cre* mouse lines allow for ATAC-seq of PV+ nuclei using INTACT



Cellular localization of Sun1-GFP

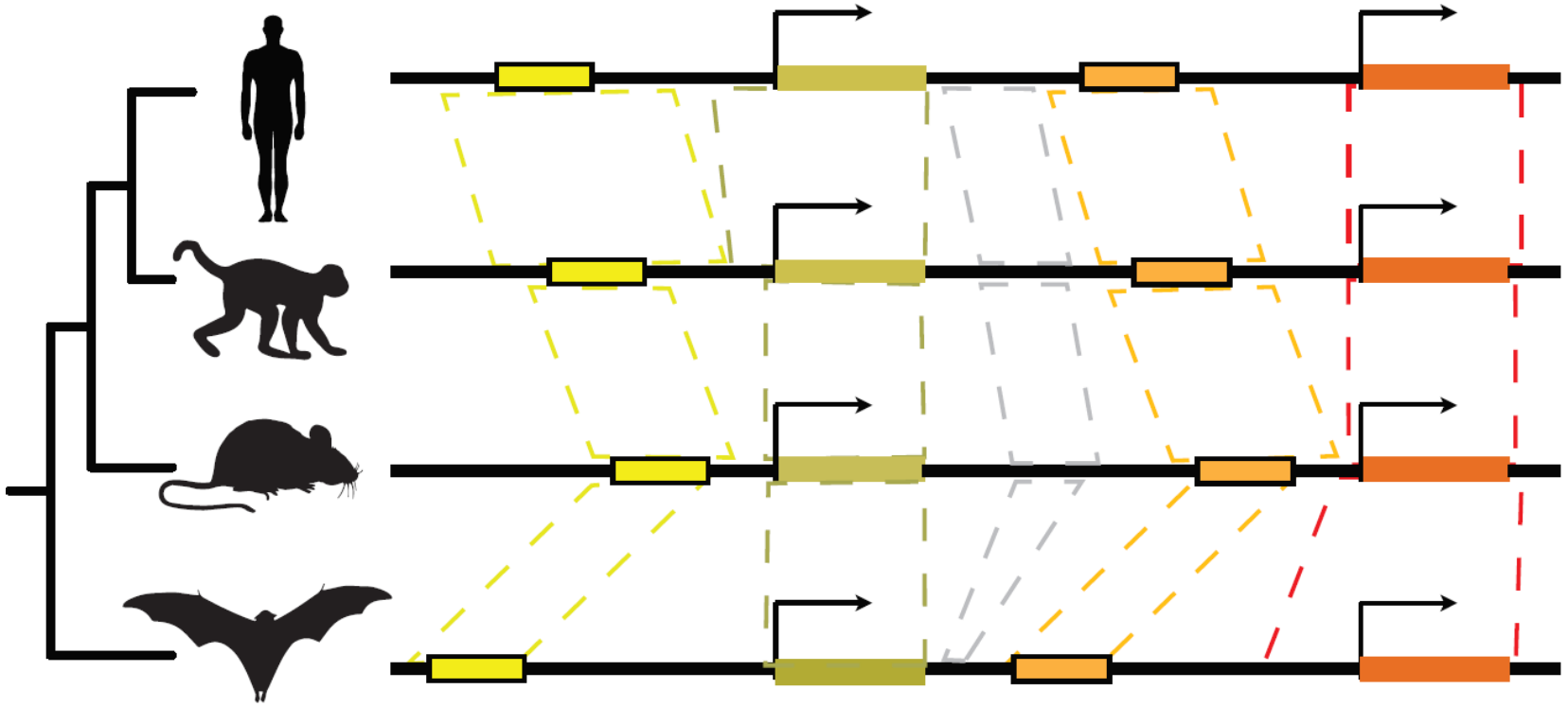
All Neurons **PV+ neurons** **Tagged cells for isolation**





Method adapted from Mo et al., Neuron, 2015

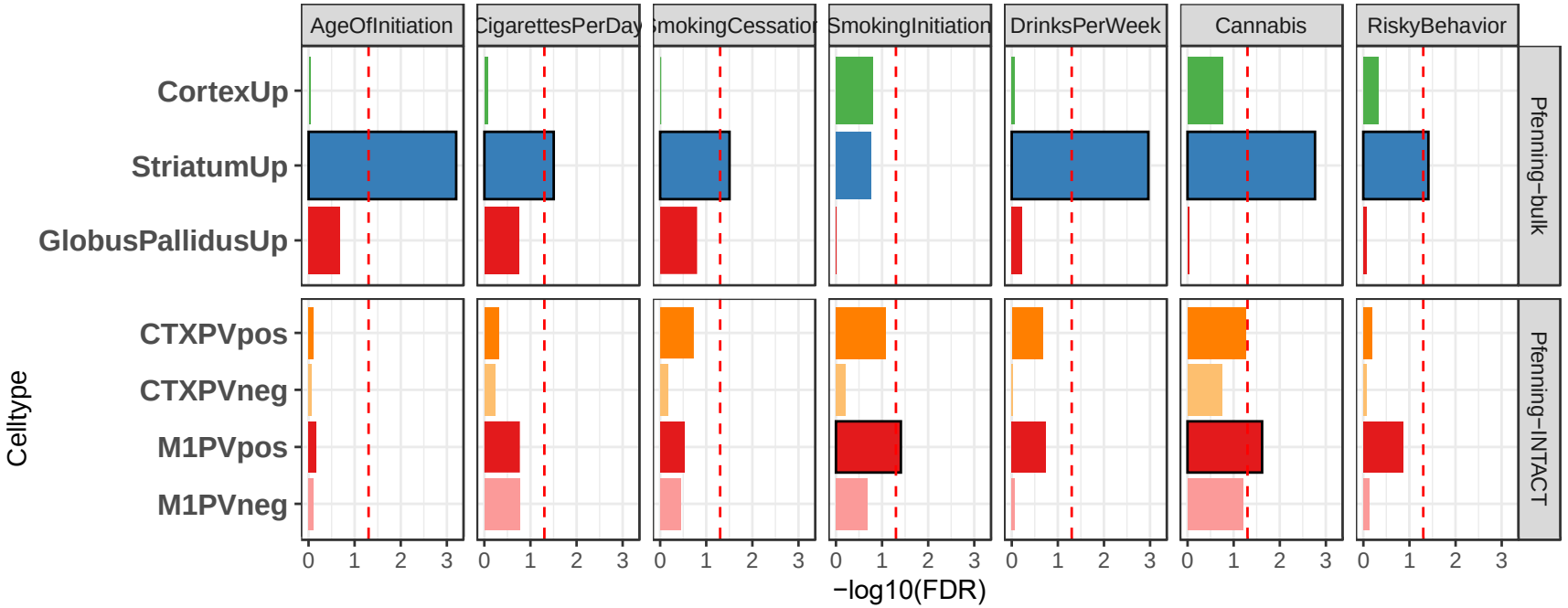
Cactus, HAL, HALPER: Multi-Species Alignments maps orthologous enhancer regions from mouse to hg38



Cactus, HAL, Hickey et al., Bioinformatics 2013

HALPER, Irene Kaplow, Erin Zhang, Morgan Wirthlin

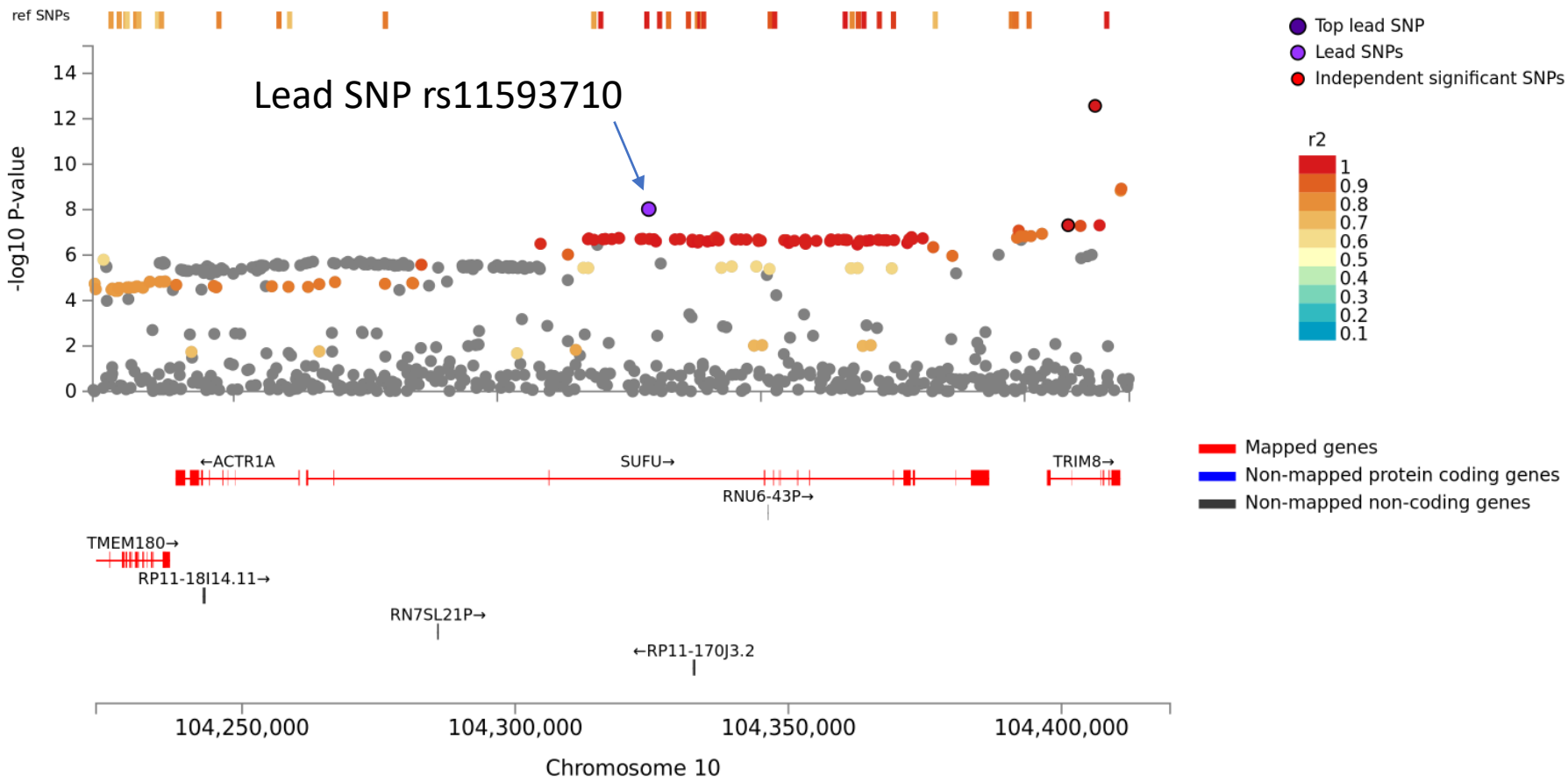
Mouse genetic tools dissect inhibitory neuronal enrichments using cell type-specific ATAC-seq



Leveraging computational tools for predicting functional SNPs.

2. From GWAS to gene regulation

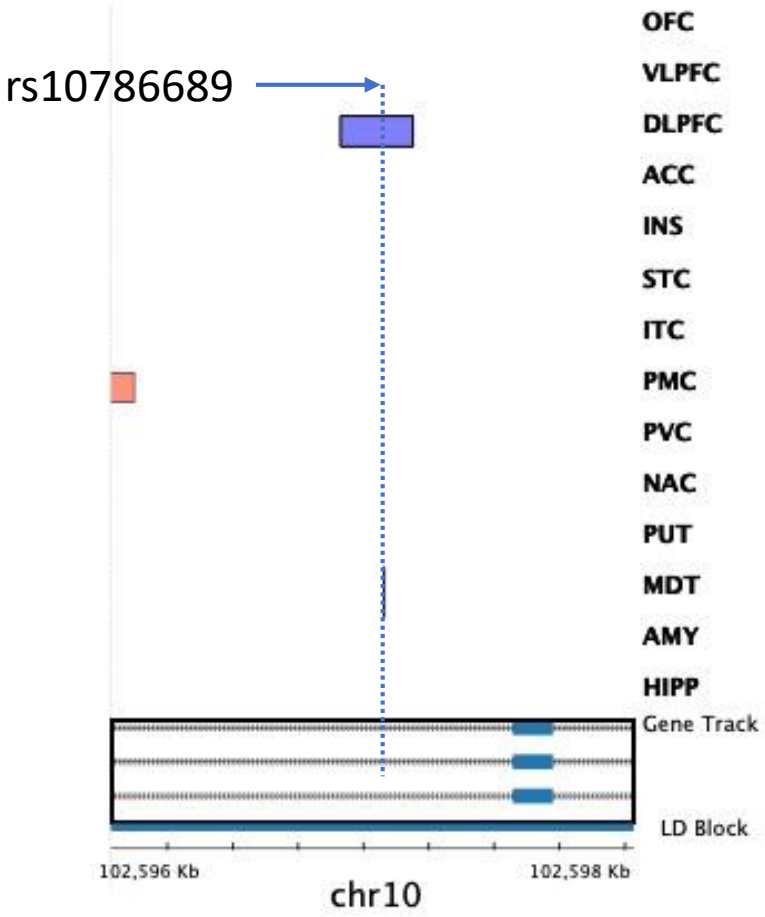
LD block of Smoking Initiation SNPs within **SUFU** gene (Suppressor of Fused)



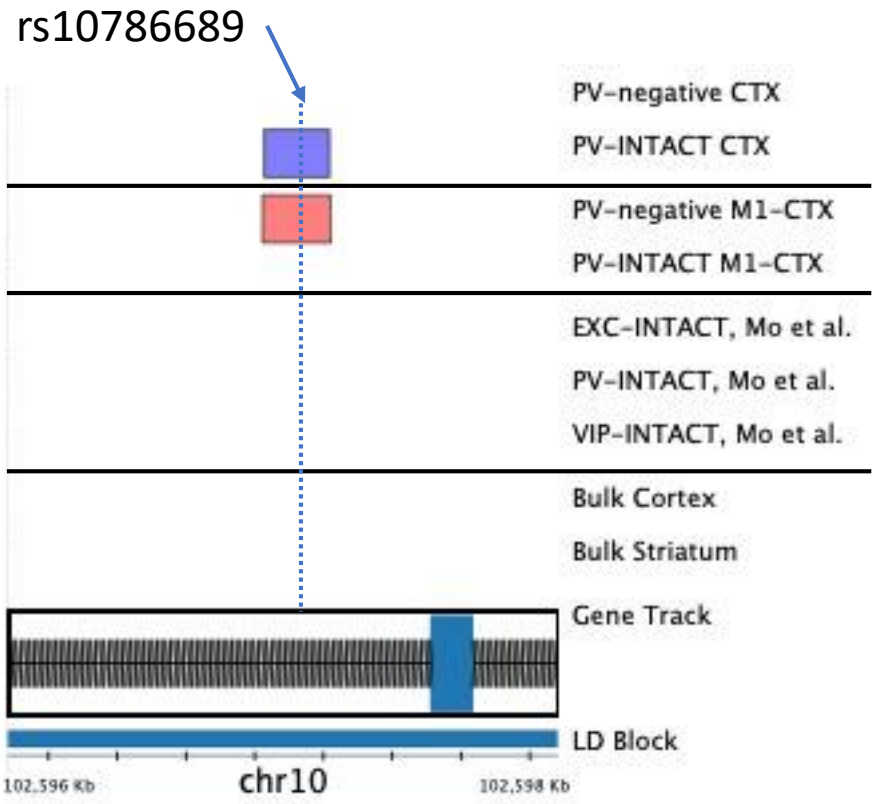
FUMA, Watanabe et al., Nat Com 2017; Liu et al, Nat Genetics 2019

SUFU, rs10786689: Fine-mapping addiction loci with cross-species cell-type specific epigenetics

Human NeuN+/NeuN- ATAC-seq



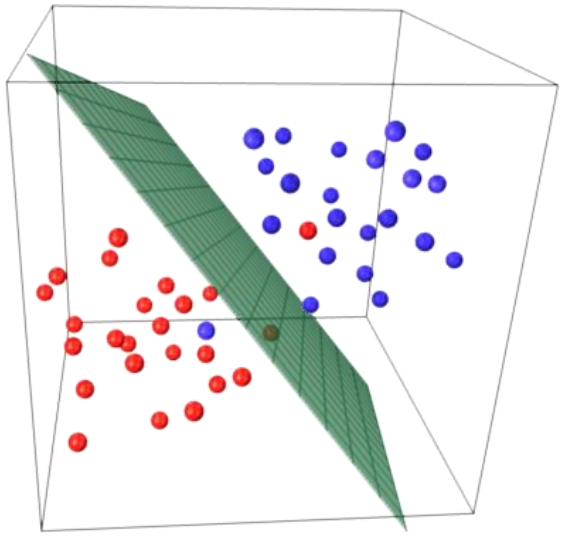
Mapped mouse PV+/PV- ATAC-seq



Fullard et al., Genome Res 2018, Mo *et al., Neuron*, 2015

Can we we predict the impact that the SNP is going to have on PV enhancer function?

TCAGATTAGTTA…



Active regulatory element in PV+ neurons

OR

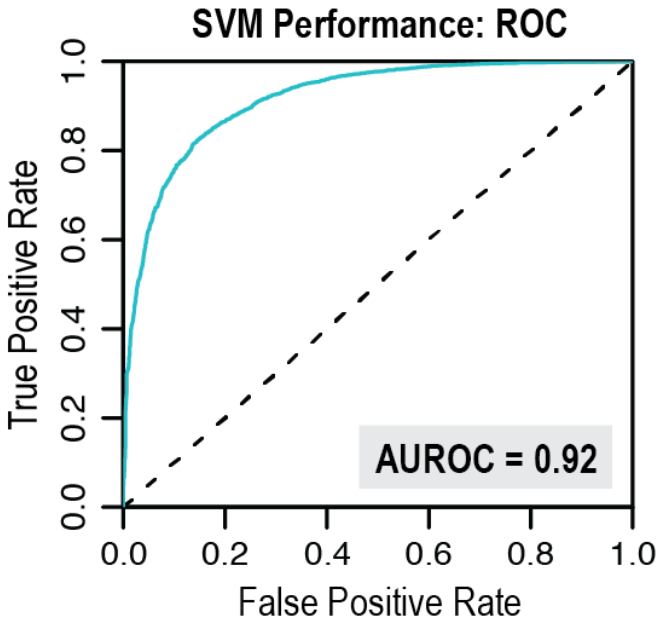
Active regulatory element in PV- cells

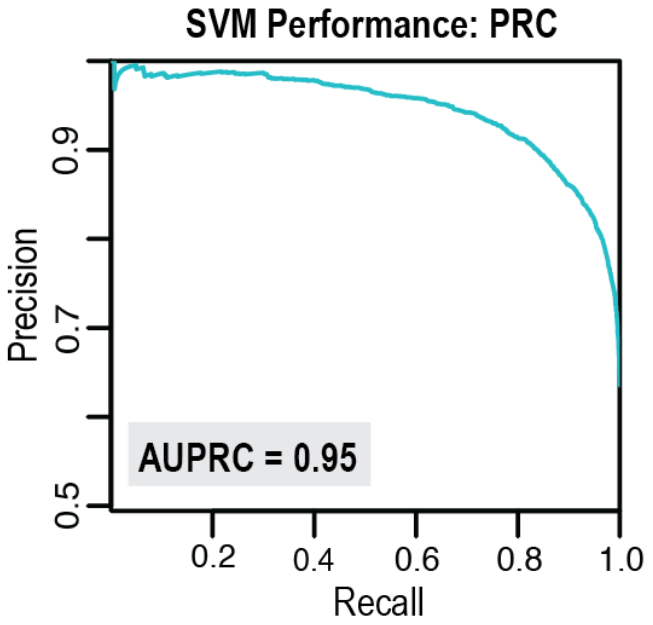
Differential regions of open chromatin

SVM Classifier

Prediction of regulatory activity

LS-GKM predicts PV+/- enhancers





	Accuracy	Precision	Recall
Pfenning Validation Data	84.1%	87.8%	87.1%
Mo *et al.* Data	76.7%	73.1%	80.0%

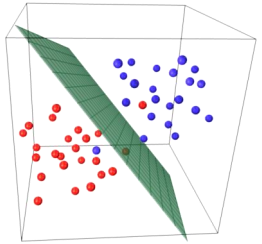
LS-GKM: Ghandi, Lee *et al.*, *PLOS Comp Biol*, 2014
Mo *et al.*, *Neuron*, 2015

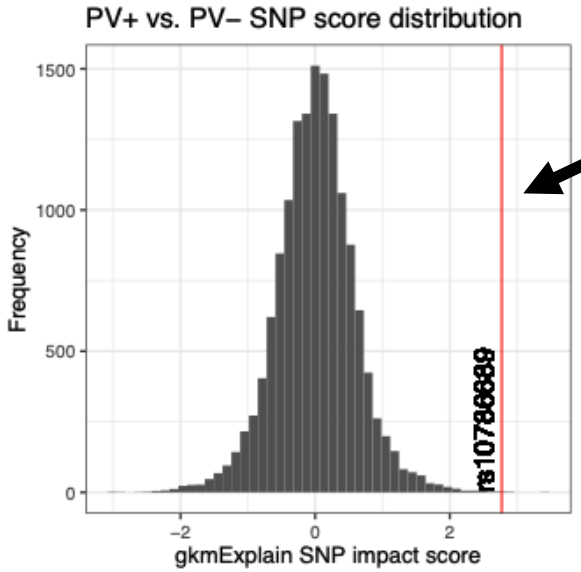
Alyssa Lawler

SUFU, rs10786689: cell type-specific ML model scores impact of SNP on PV-enhancer activity

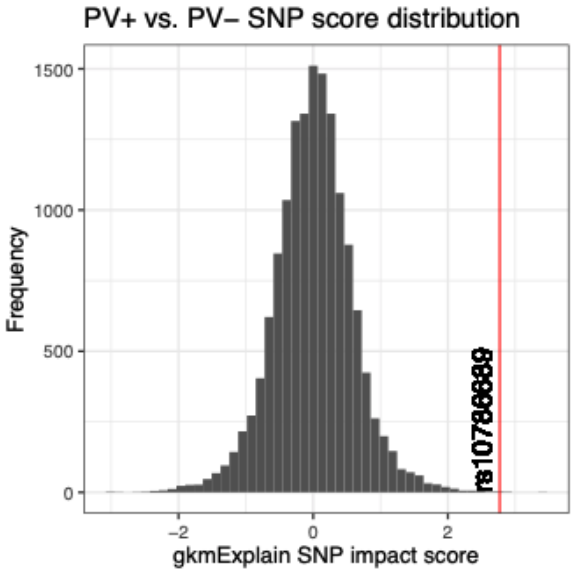
CTAGGAAGTCCAGCCTTTG **T** TCTCCAGTGGCTCTTCCCTG

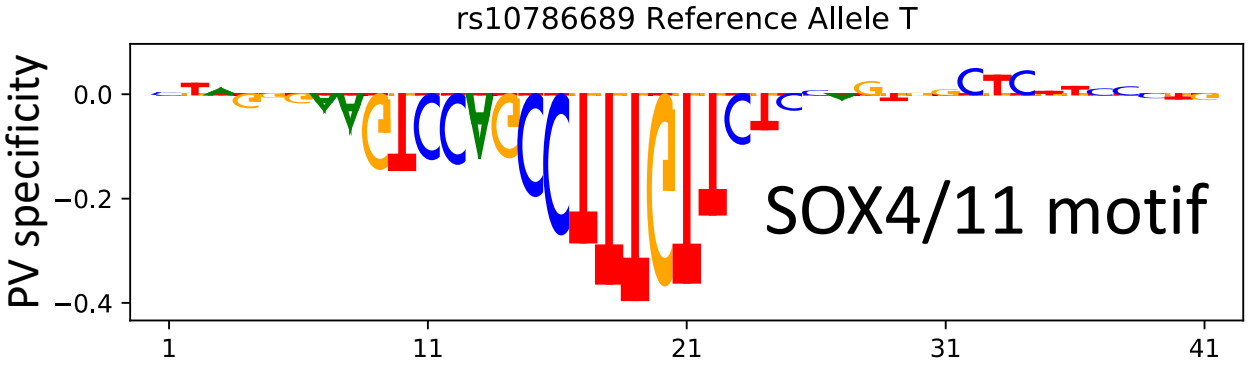
CTAGGAAGTCCAGCCTTTG **C** TCTCCAGTGGCTCTTCCCTG



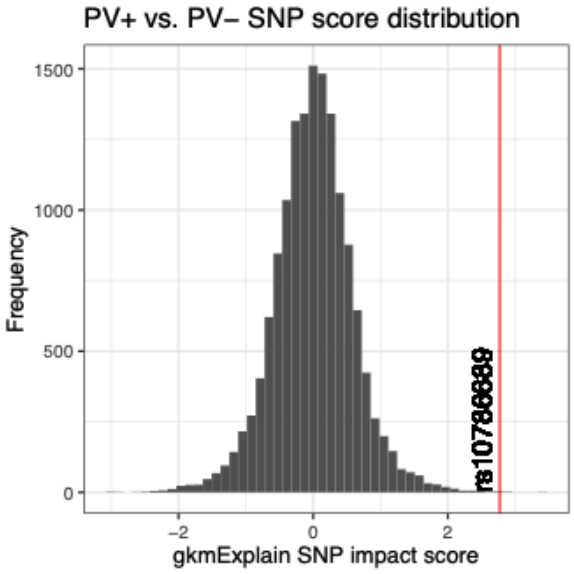


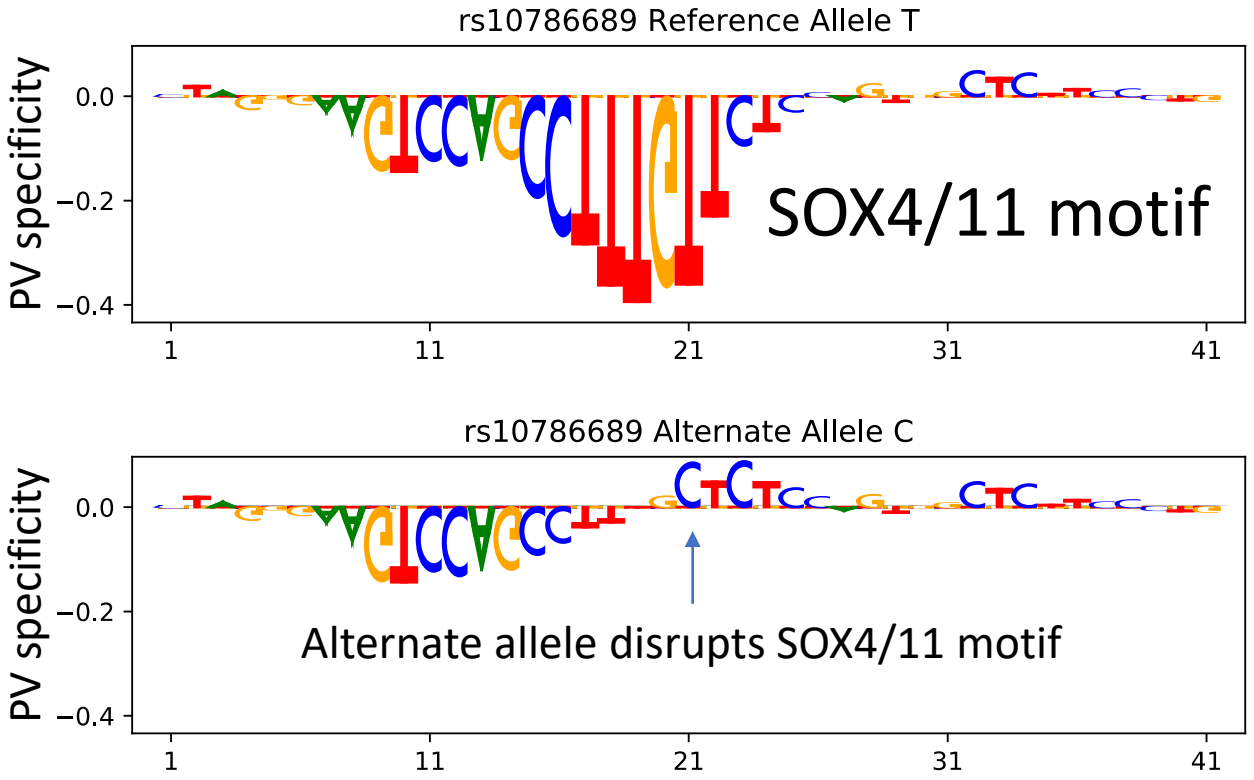
SUFU, rs10786689: cell type-specific ML model scores impact of SNP on PV-enhancer activity





SUFU, rs10786689: cell type-specific ML model scores impact of SNP on PV-enhancer activity



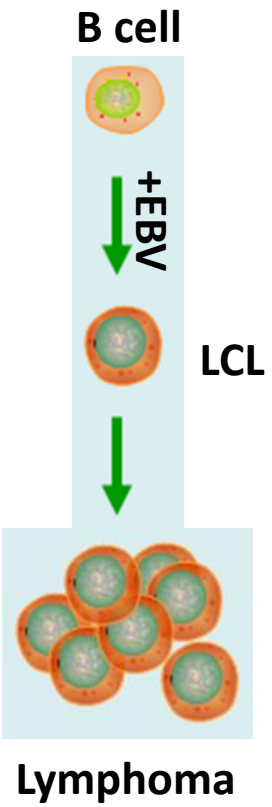


Assessing ML models on SNP functional effects with MPRA

3. Validating ML SNP predictions

Lymphoblastoid cell lines (LCL)

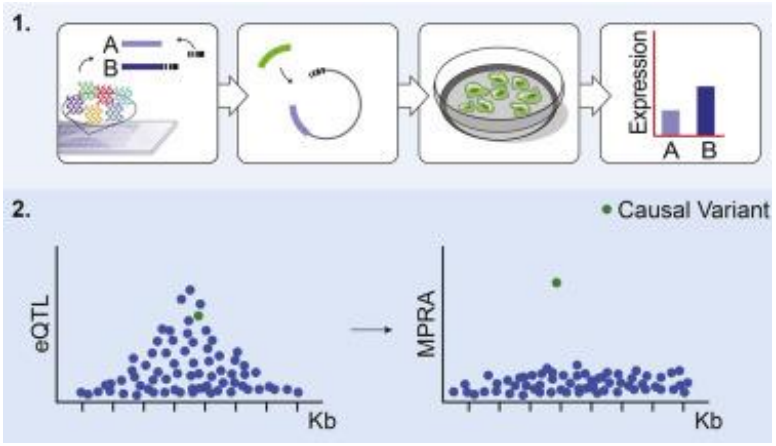
- Human GM12878
  - Mormon Utah mother (HapMap project)
  - B-lymphocyte + Epstein-Barr Virus
  - Benchmark for genomic assays & population genetics
- Mouse CH12.LX
  - Immortalized B-cell lymphoma
- ENCODE DNase hypersensitivity
  - Stam & Crawford labs (GM12878), Stam lab (CH12.LX)
  - ATAC-seq pipeline (w/ DNase setting)



ENCODE project PMID: 22955616, 29126249,

LCL for addiction genomics?

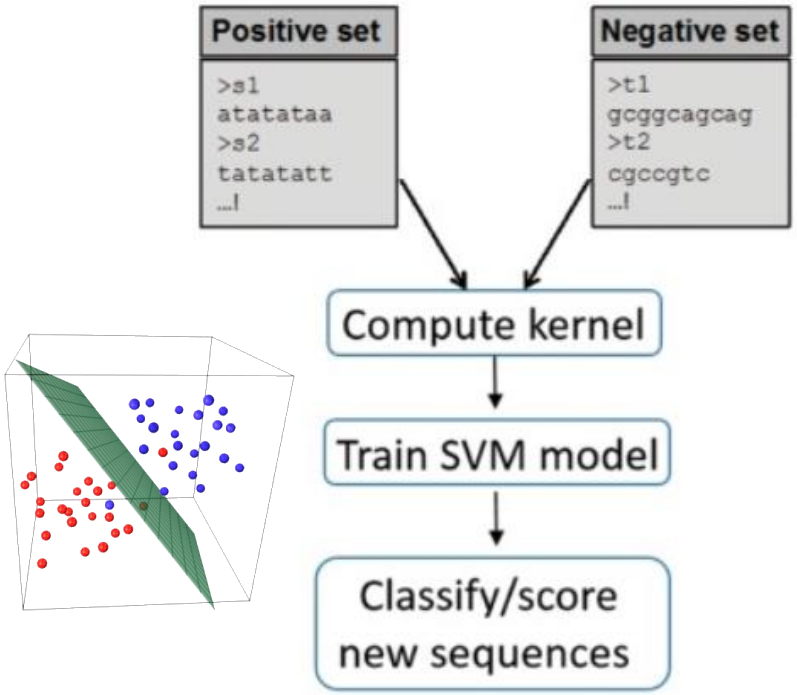
1. How well do ML methods predict enhancer activity?
2. Do models from mouse LCL predict human SNP function?
3. How well do ML models predict MPRA SNP function?



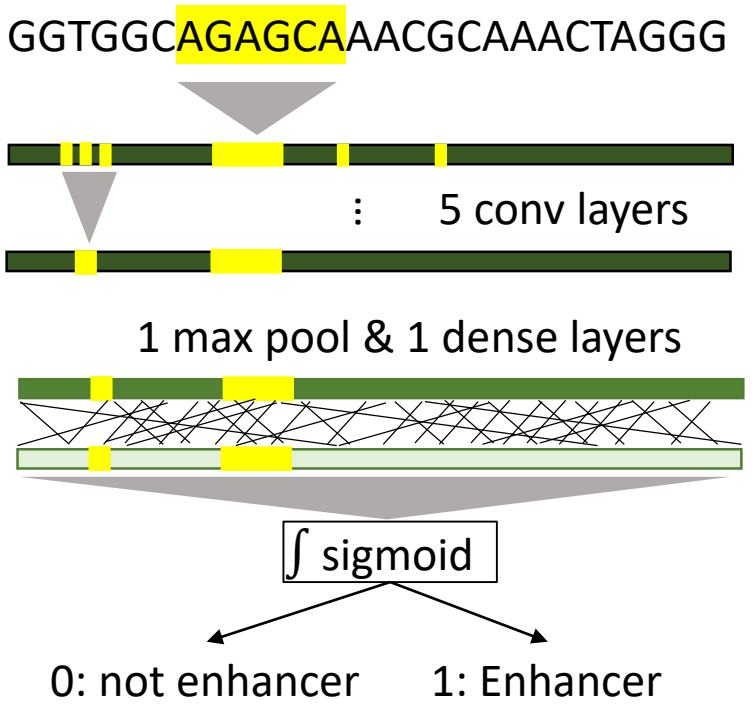
Tewhey et al., Cell 2016

ML models for enhancer activity

Gapped k-mer SVM



Convolutional Neural Network



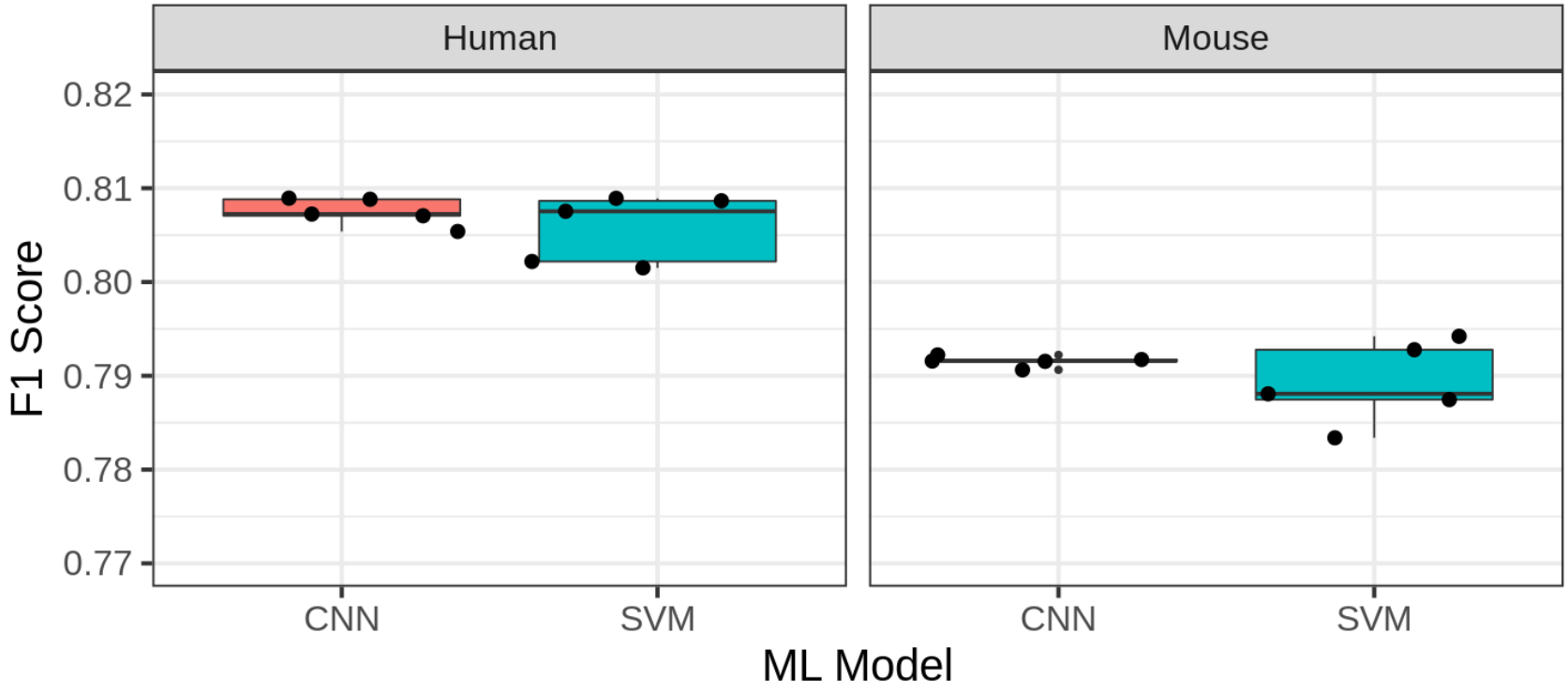
Ghandi, Lee et al., PlOS Comp Bio 2014; Lee, D. Bioinformatics 2016

Positives & Negatives

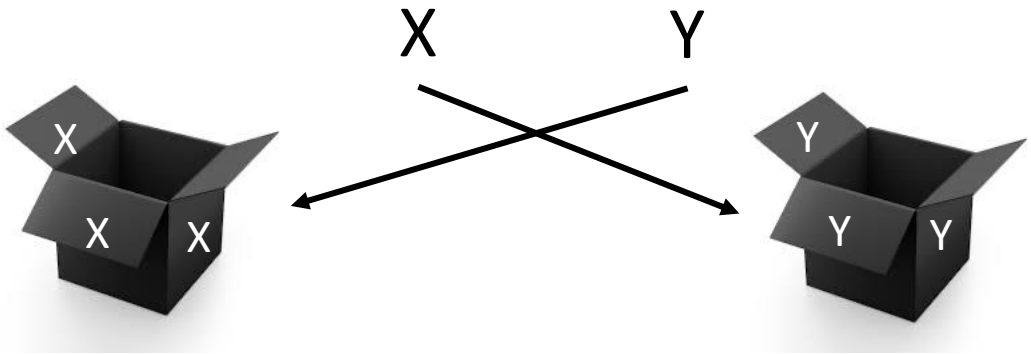
- Positive set
  - DNA-seq open chromatin peaks
  - Train **167bp** enhancer sequences
  - Summit centered—most informative
- Negative set
  - GC & repeat matched genomic sequences
- Parameter tuning
  - Learning rate, convolution filter size, k-mer length

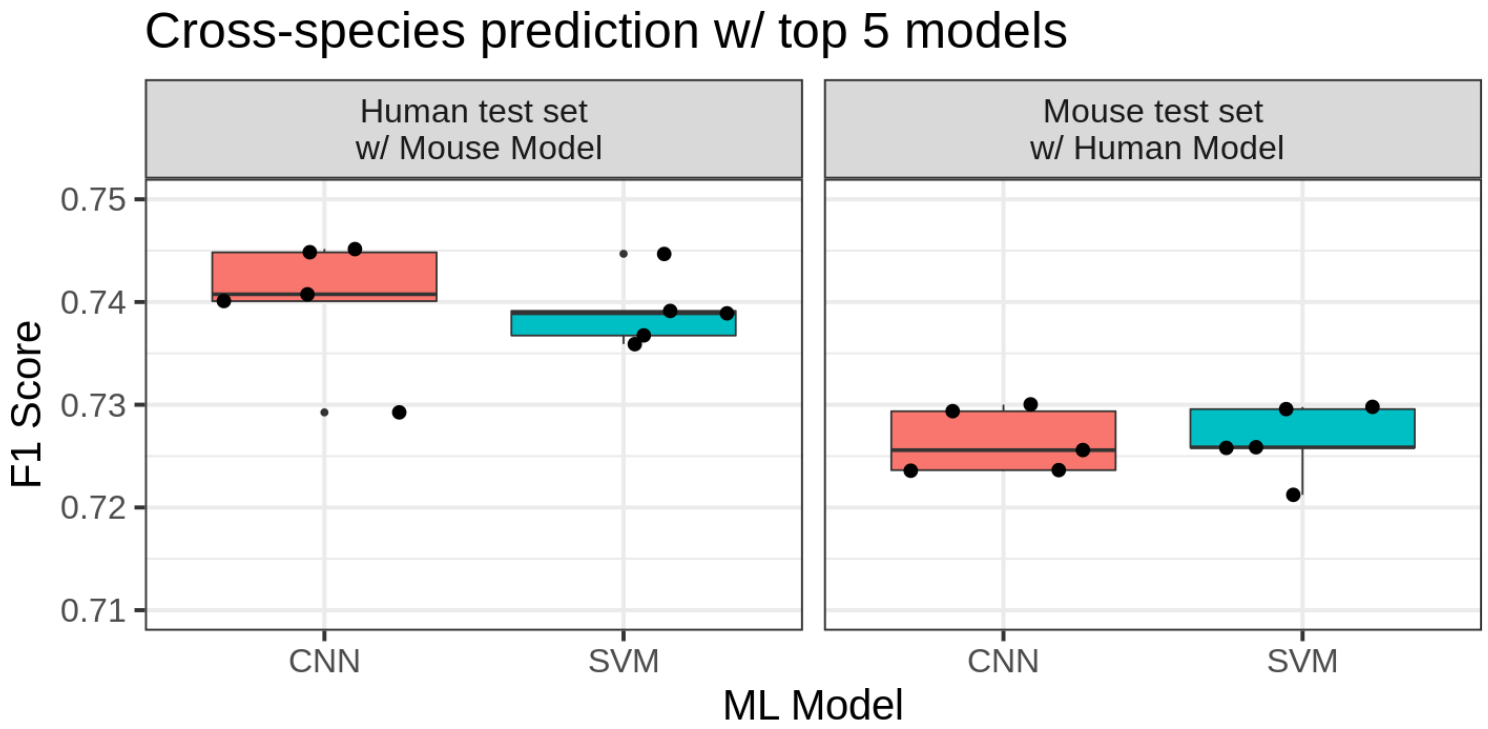
Which ML methods are better?

Top models trained within each sample, ranked by F1-score.

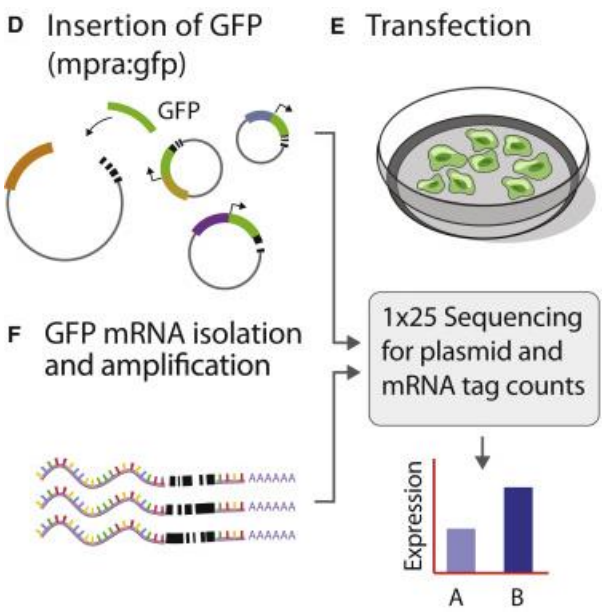


How well does model trained in {Species X} predict test set from {Species Y}?

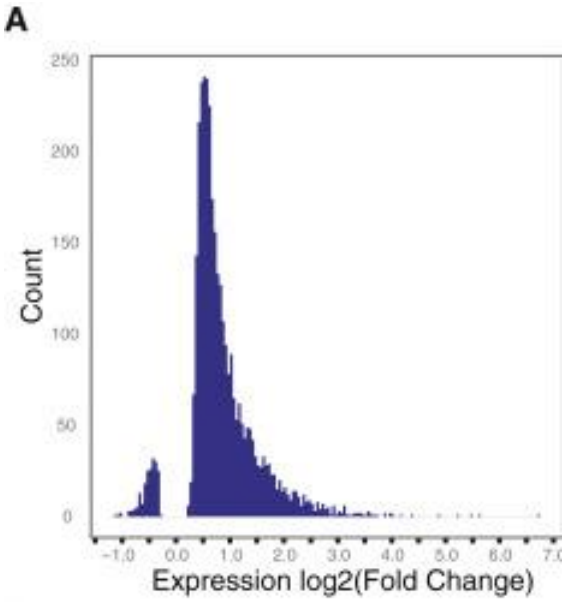




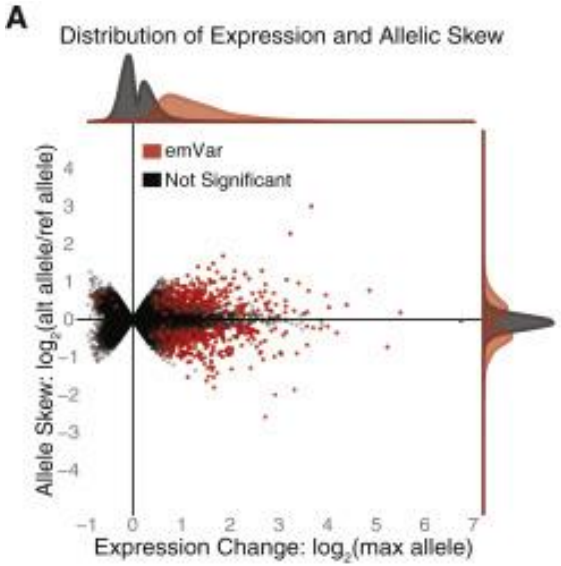
GM12878 SNP-MPRA design



~30,000 SNP pairs tested in 180bp MPRA



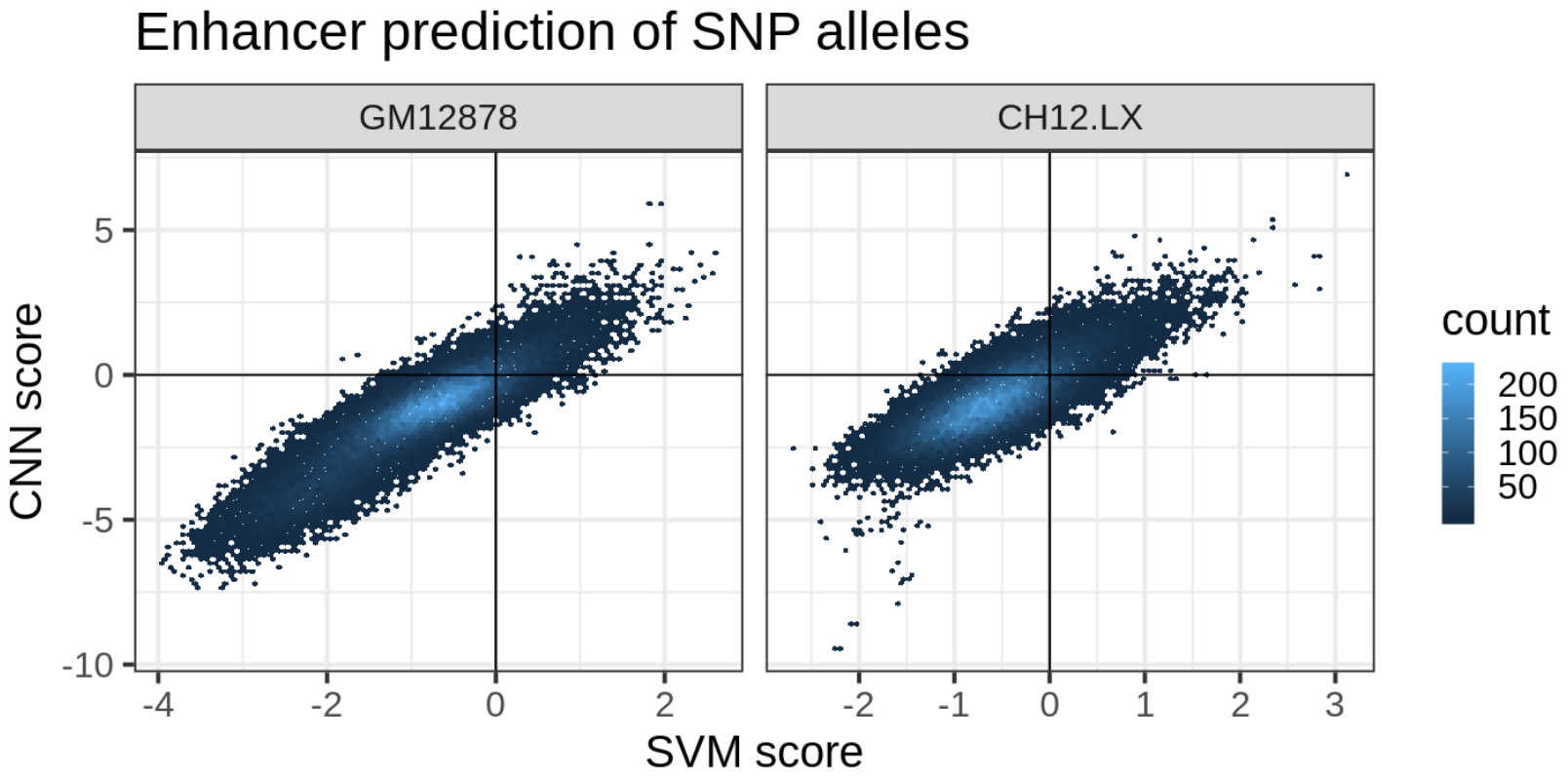
Many sequences increase GFP expression log2(RNA/DNA)



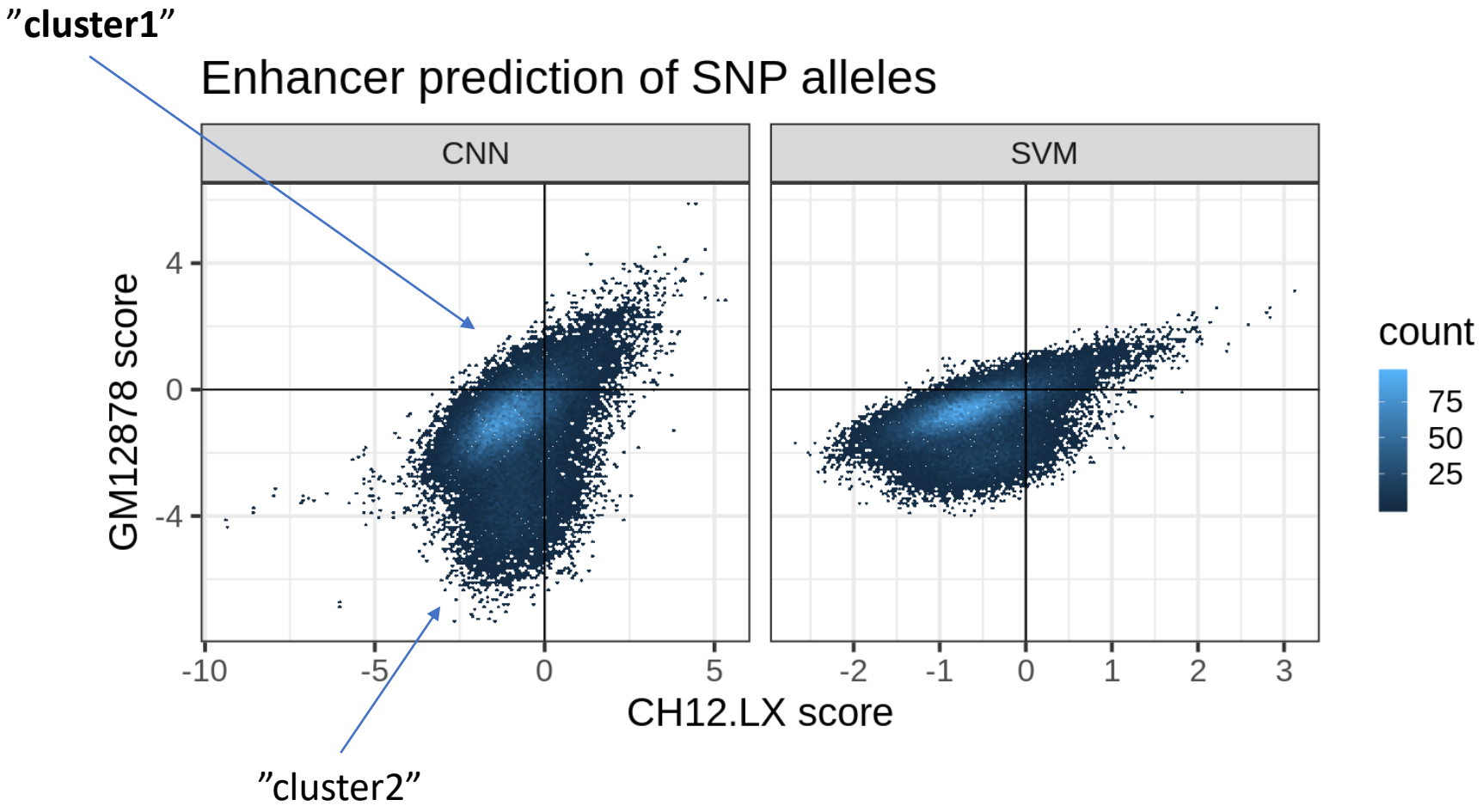
~842 SNPs are expression modulating variants (emVAR) Alt allele vs. Ref allele

Tewhey et al., Cell 2016, from figures 1, 3, 4

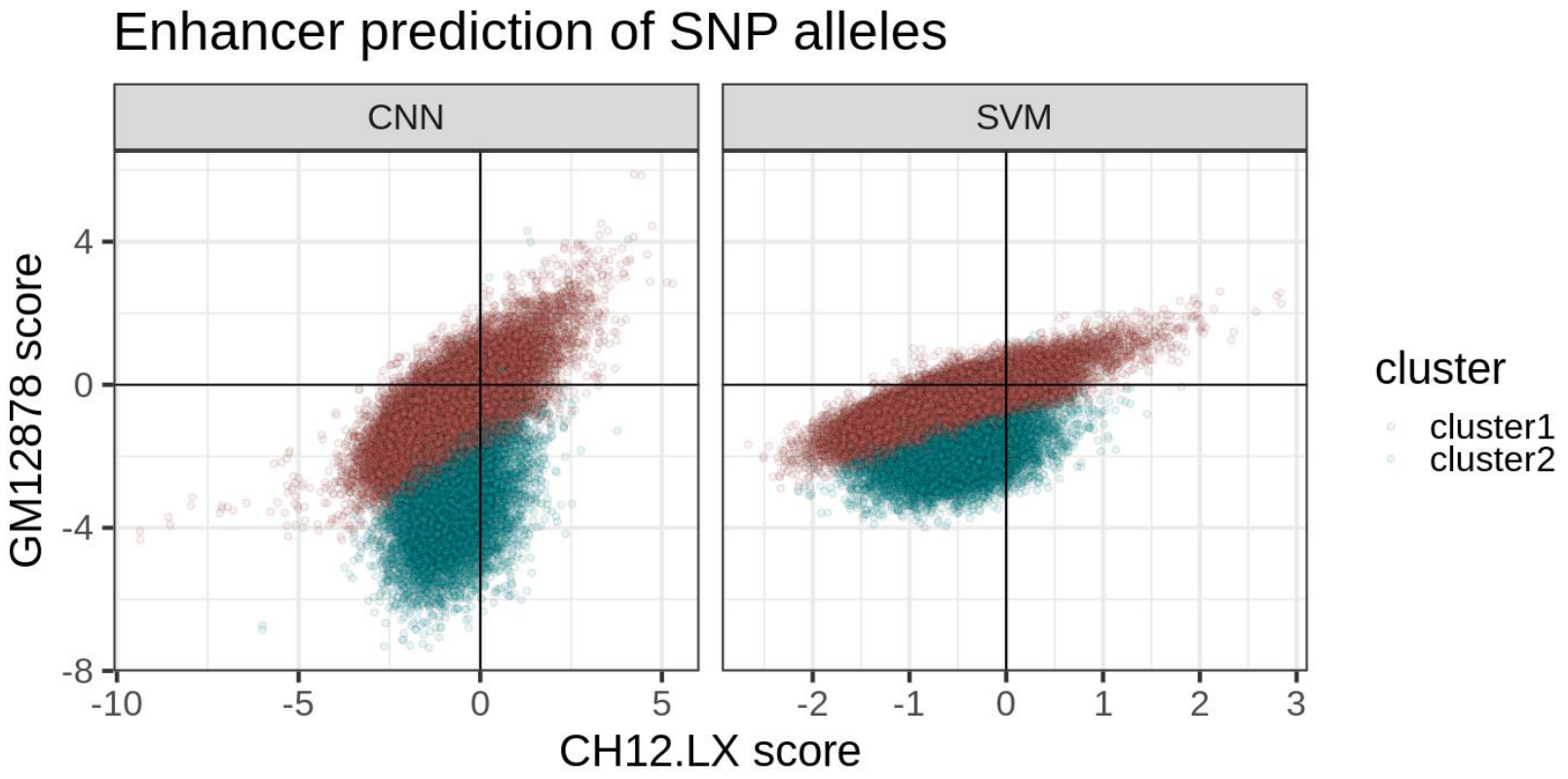
Both models make similar enhancer activity predictions



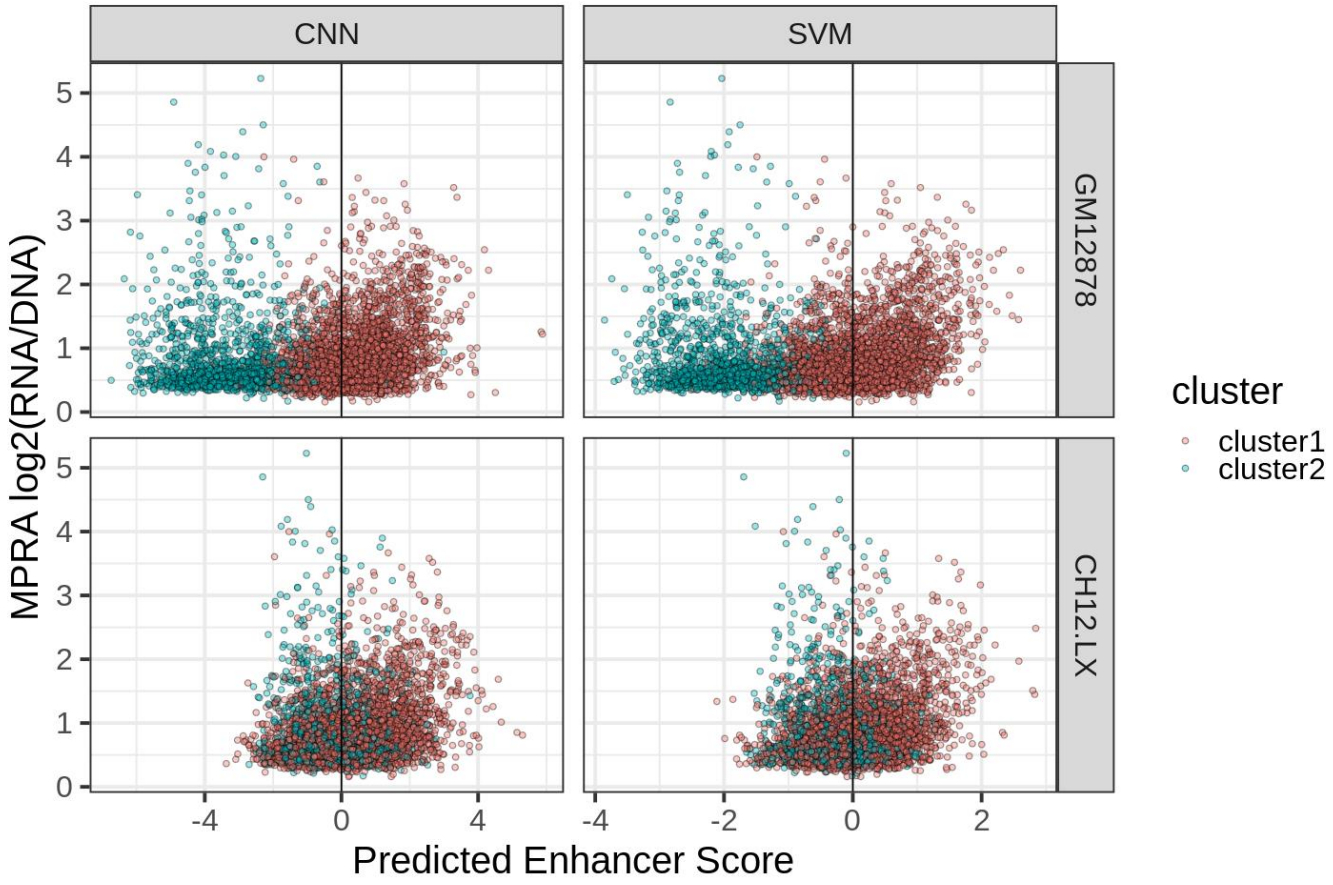
GM12878 models learn different sequences of enhancers from CH12.LX



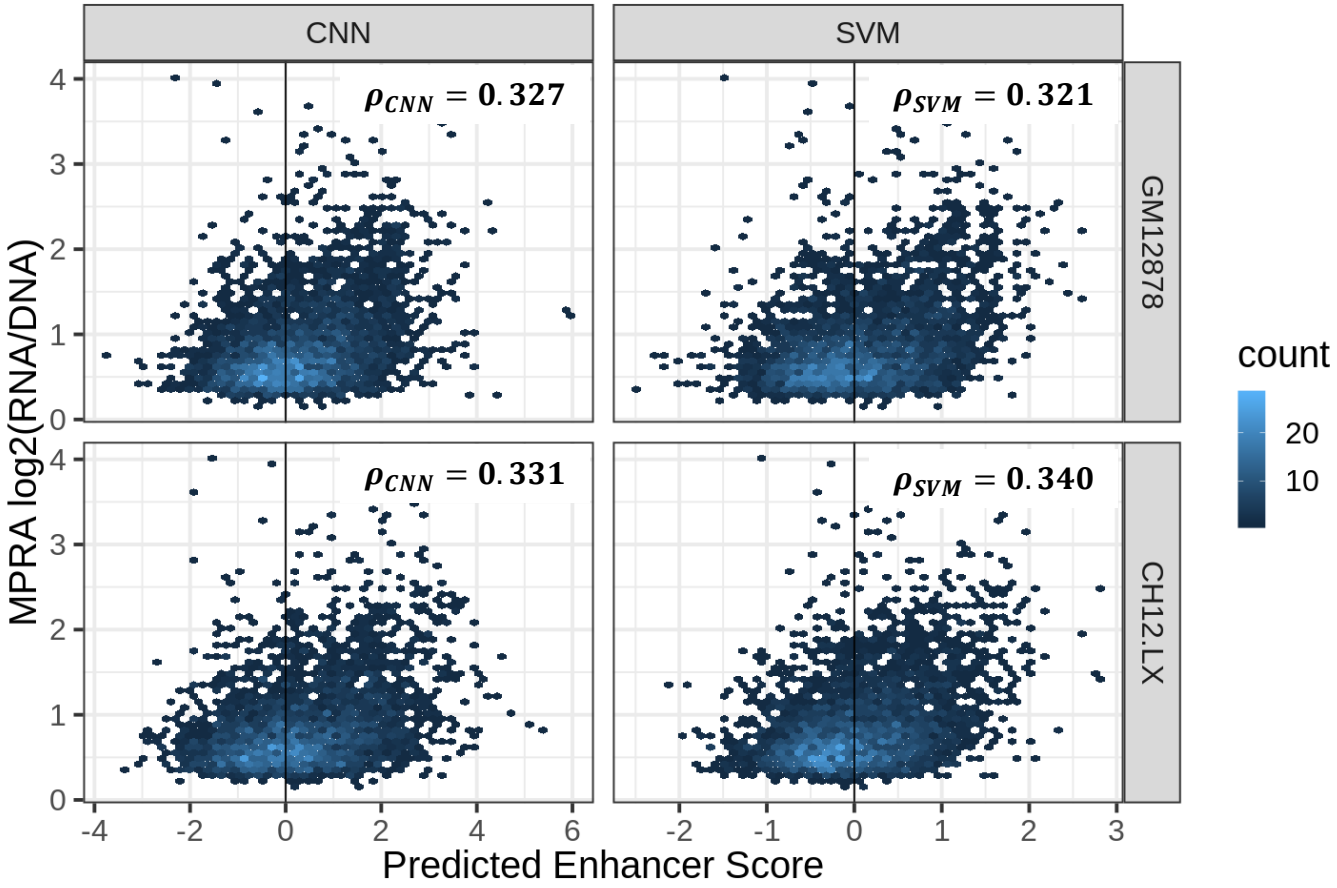
GM12878 models learn different sequences of enhancers from CH12.LX



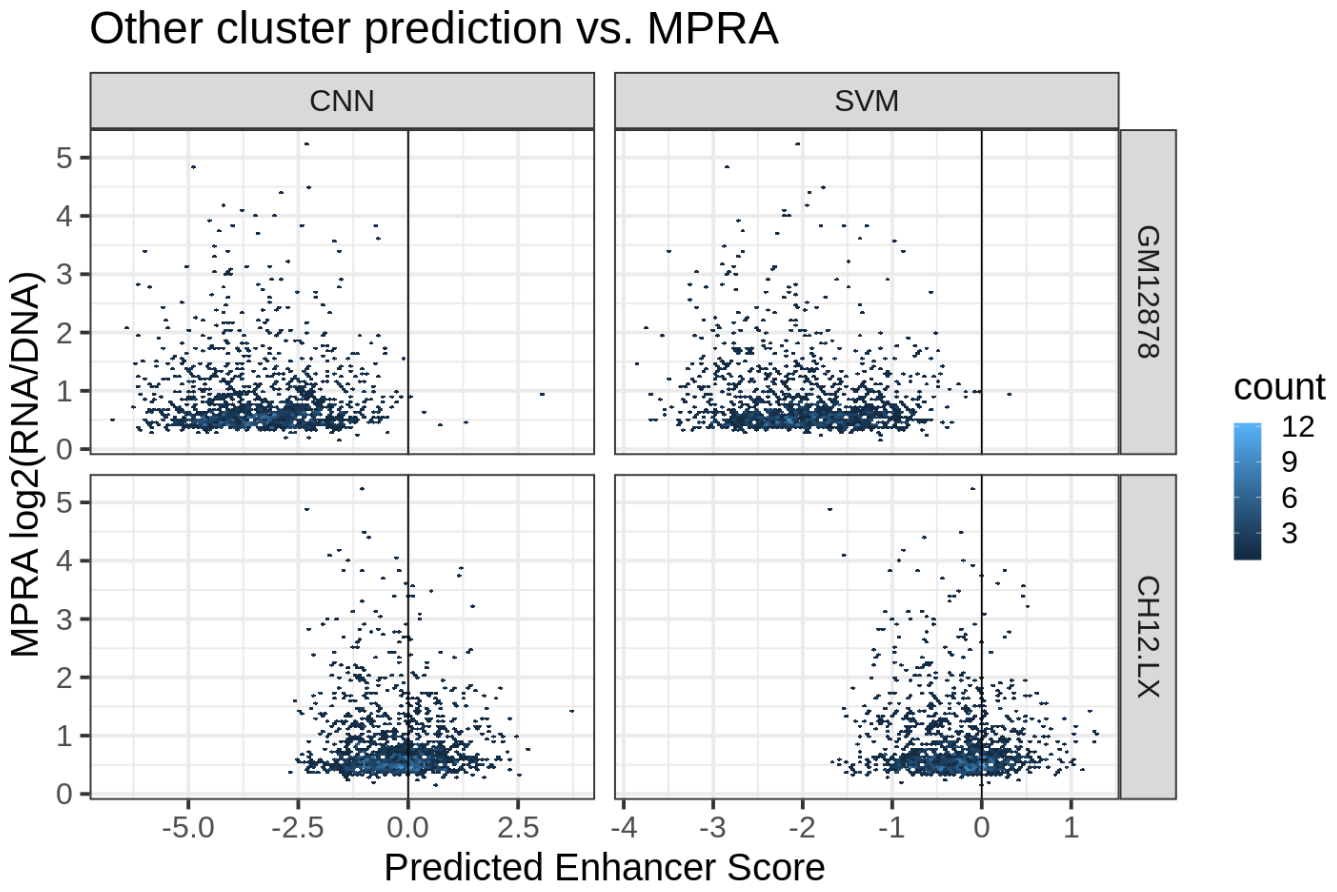
Do models predict enhancer activity? log2FC(RNA/DNA)



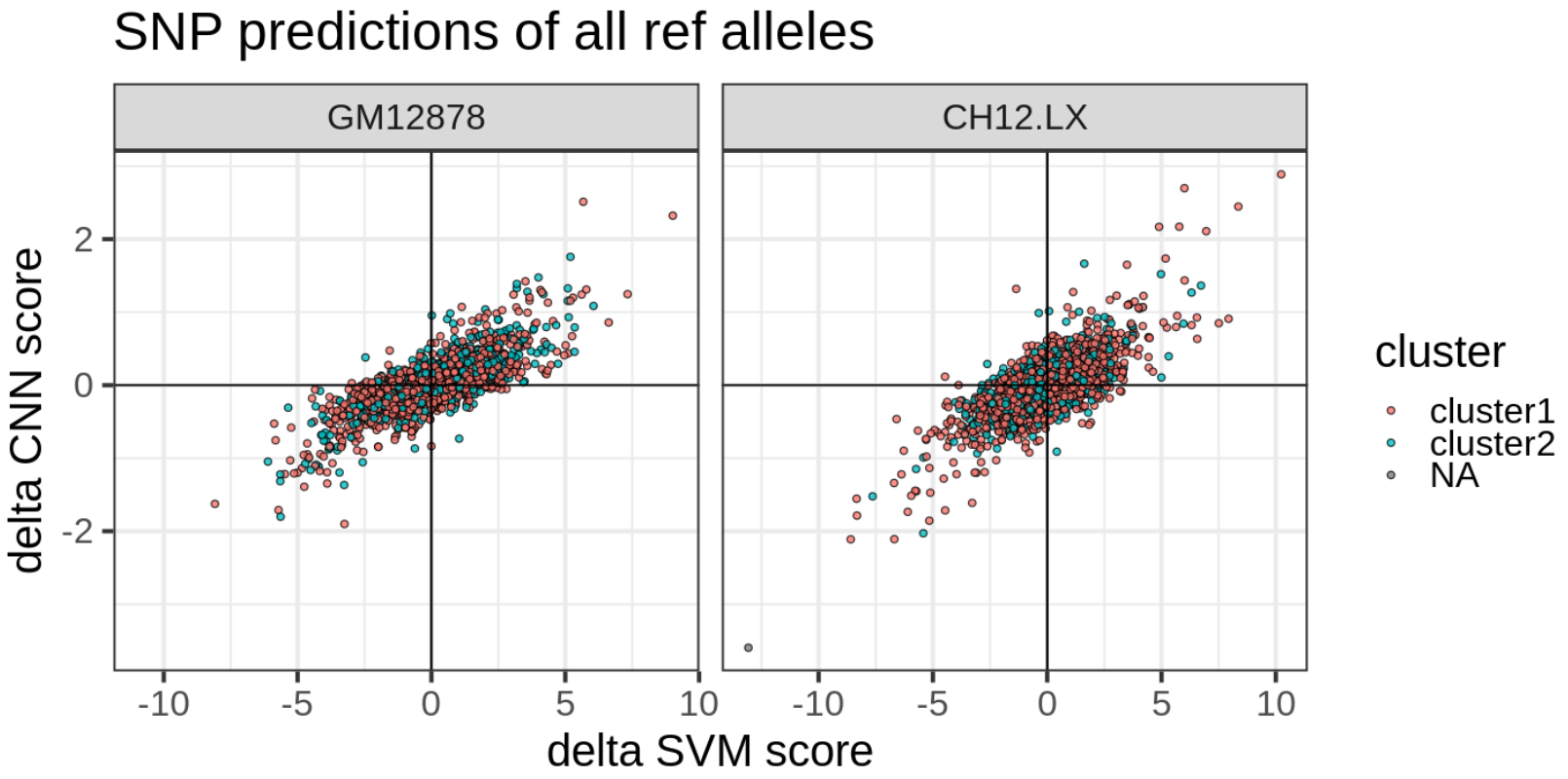
Both species predict cluster 1 SNP enhancer activity well



GM12878 models poorly predicts enhancer activity of cluster 2 SNPs

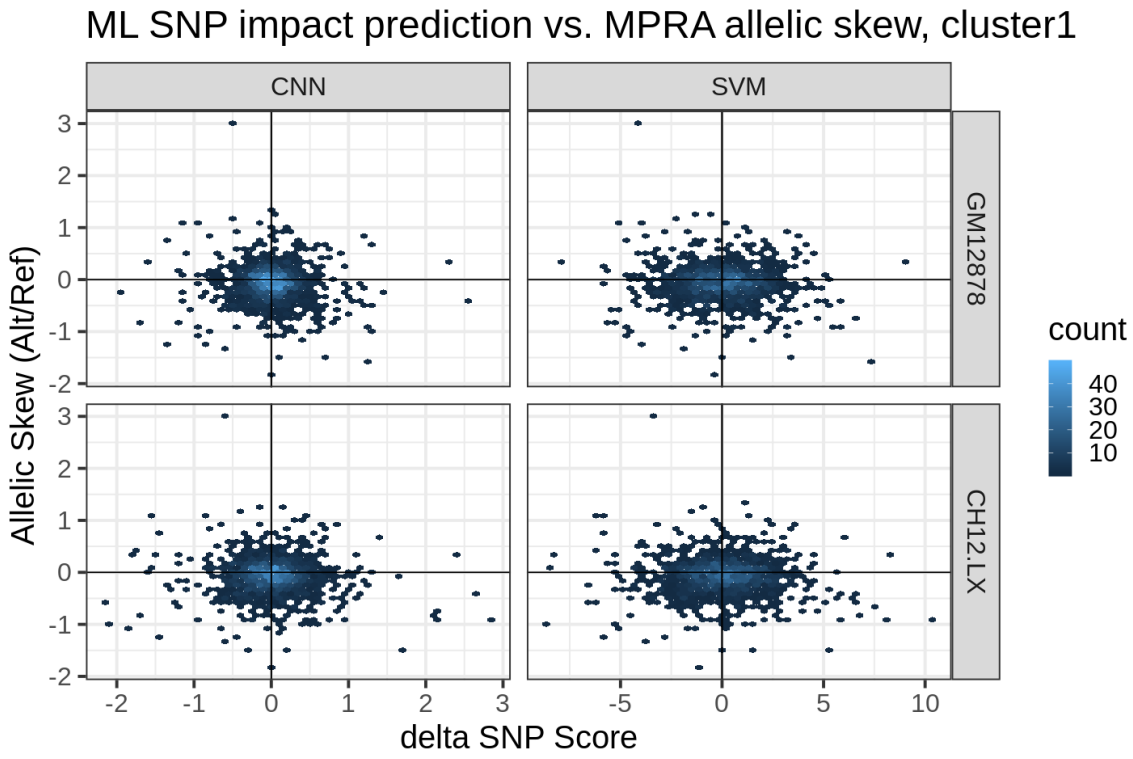


Both models predict similar SNP effects



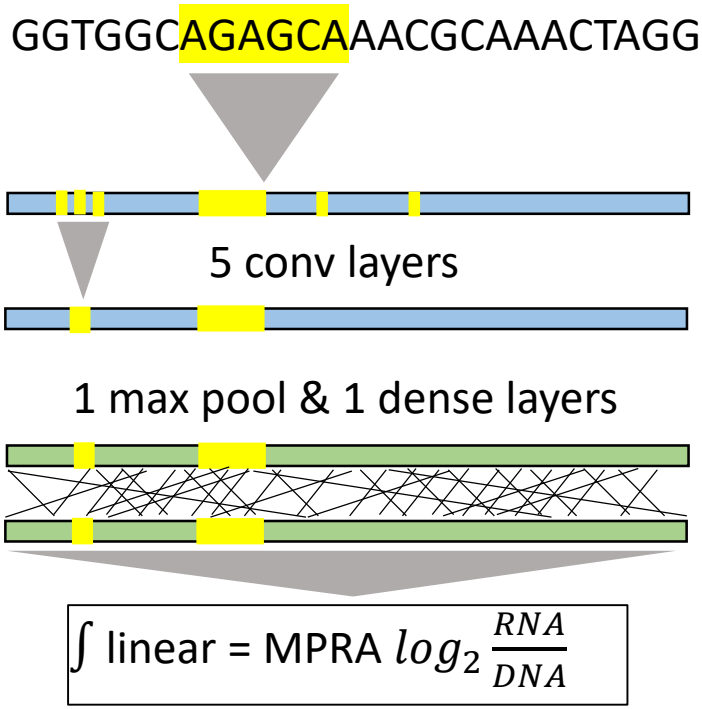
CNN: $score_{Alt} - score_{ref}$
SVM: w/ deltaSVM method

Do ML models predict allelic skew?

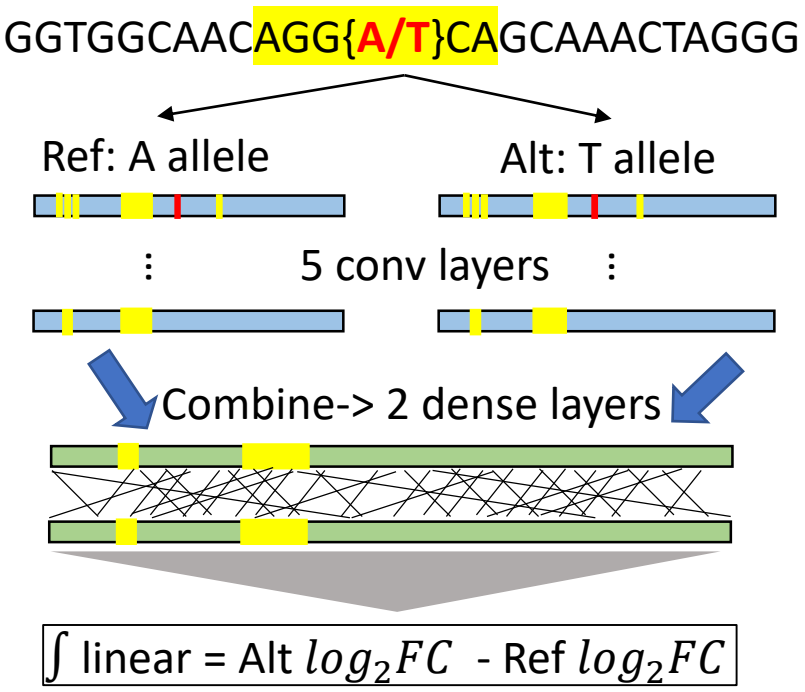


Fine-tune CNN regression models

CNN-log2FC model to predict MRPA activity



CNN-skew model to predict SNP impact on MPRA activity

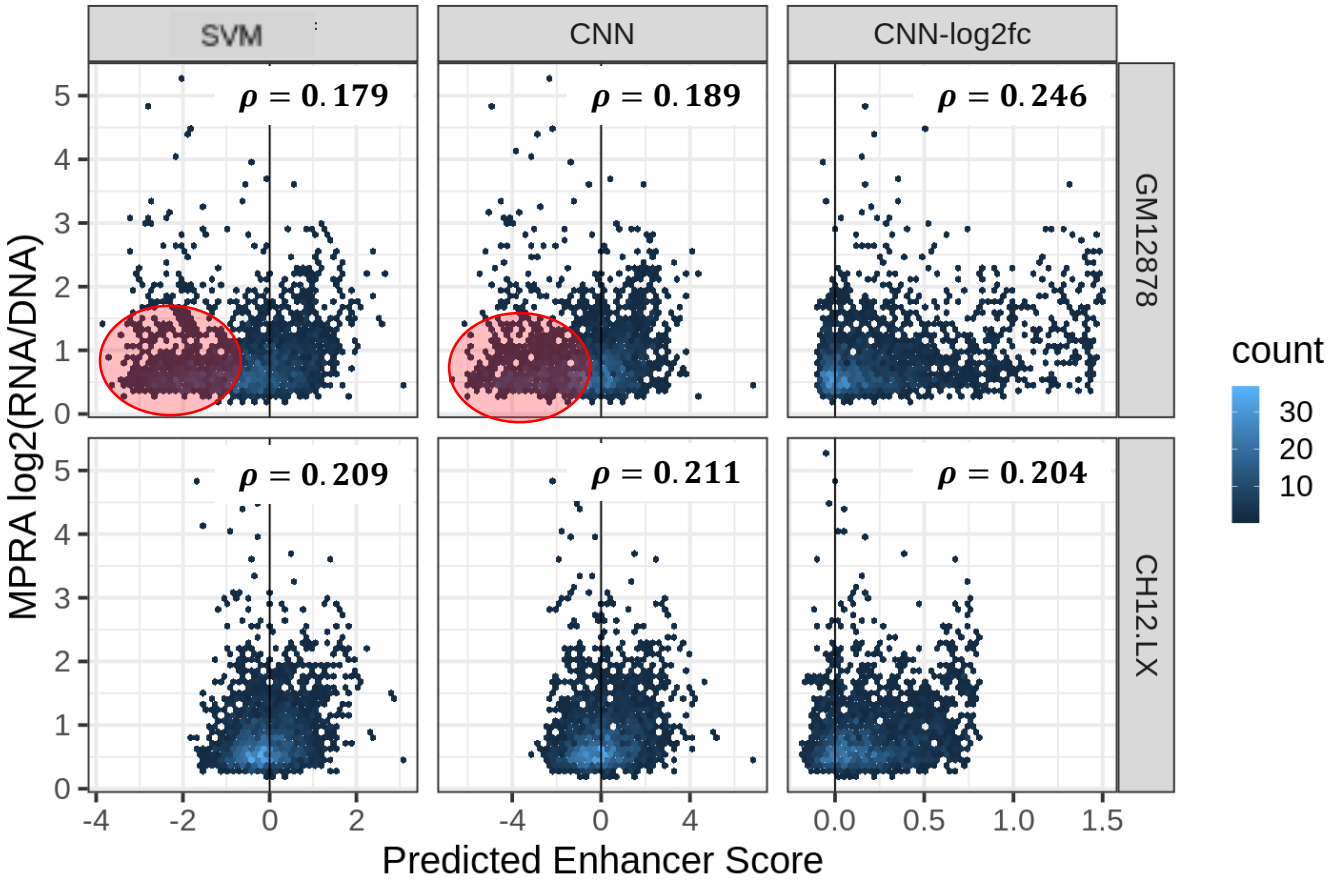


Frozen Param

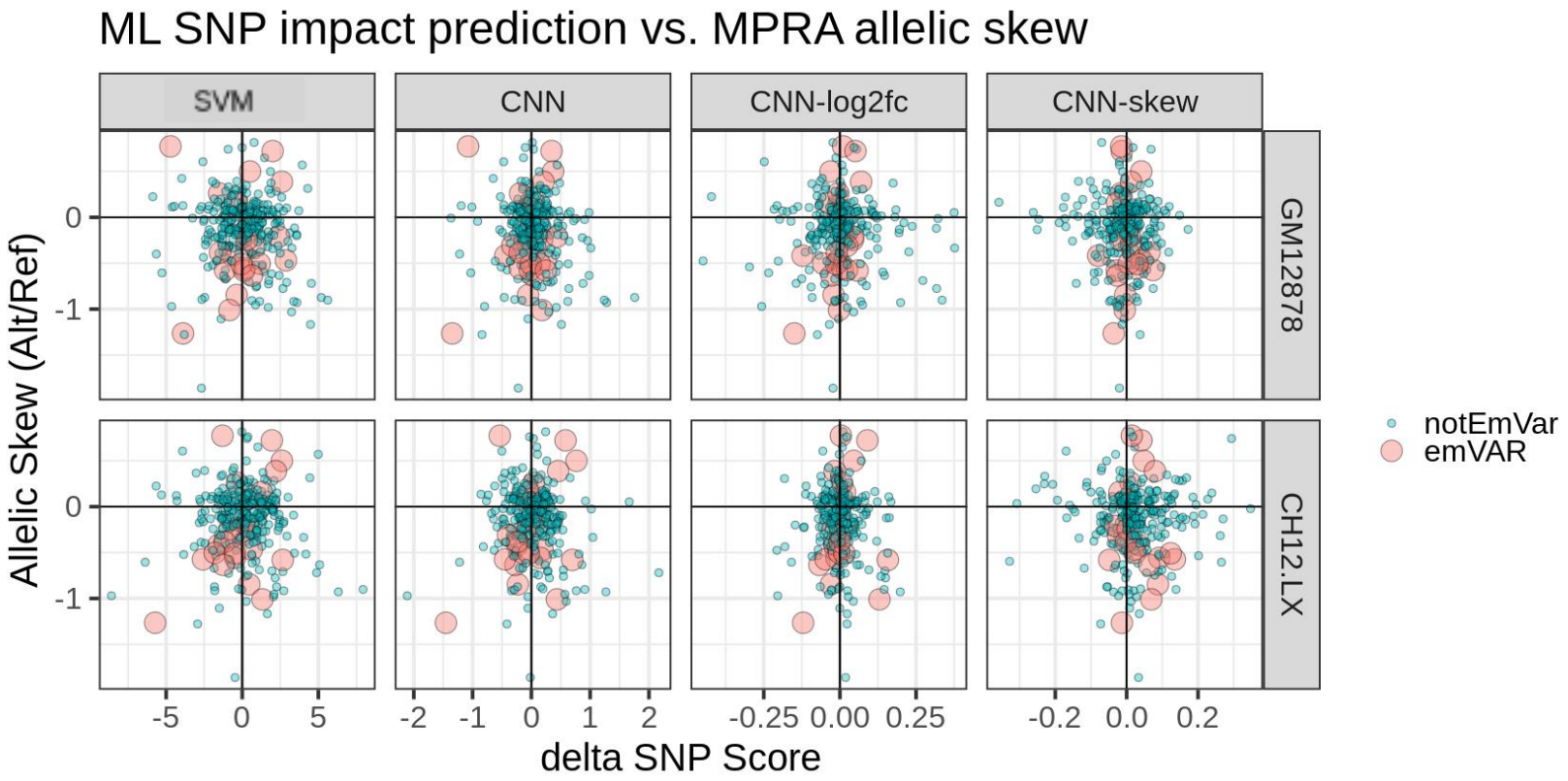
Learnable Param

Split Tewhey MPRA SNPs 50/50 for training & testing.

Fine-tuned regression models predict enhancer activity better than classifiers



Fine-tuned regression models aren't great at predicting allelic skew either



Conclusions

1. Addiction-associated GWAS SNPs encode the reward epigenome

2. Addiction-associated variants enrich for human, macaque, and mouse cell type-specific enhancers

3. ML models trained on mouse cell-type specific epigenomes may infer enhancer activity at SNPs

Acknowledgments

Pfenning Lab

Chaitanya Srinivasan
Irene Kaplow
Morgan Wirthlin
Ashley Brown
Grant Fox
Alyssa Lawler
Easwaran Ramamurthy
Cathy Su
Michael Kleyman
Nitinram Velraj
Noelle Toong
Snigdha Agarwal
Erin Zhang
Byungsoo Kang
Yeonju Kim
Emma Jin
Amulya Garimella

Collaborators

Leah Byrne
William Stauffer

University of Pittsburgh

Carnegie Mellon University