



# Single-cell Epigenomic Analysis

Yang (Eric) Li

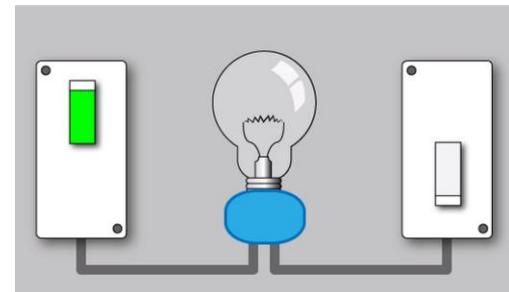
2024 CSHL seqtec course

Nov 21, 2024

Yang (Eric) Li Lab  
Website: [yelilab.wustl.edu](http://yelilab.wustl.edu)  
Email: [yeli@wustl.edu](mailto:yeli@wustl.edu)

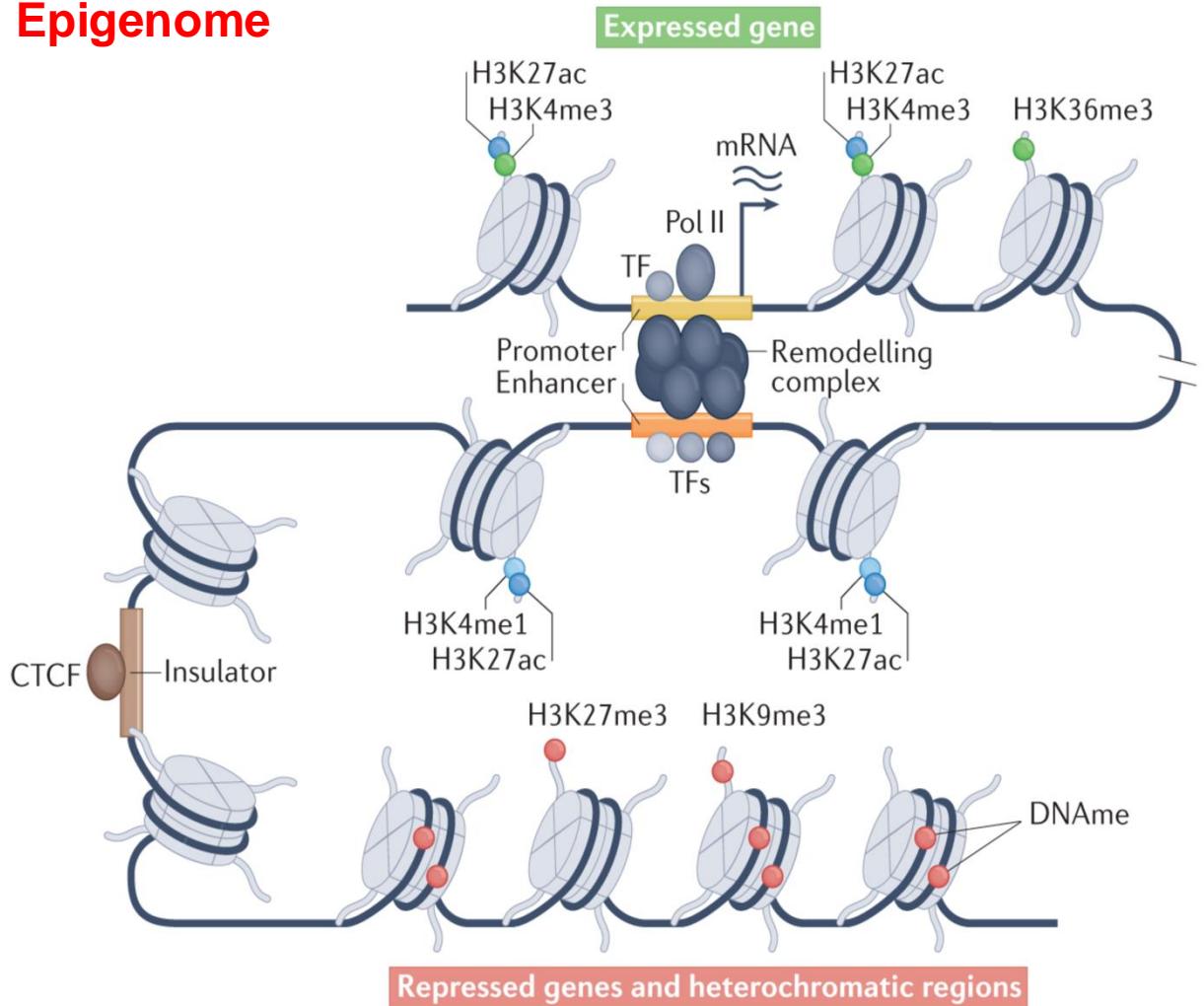


# What's epigenome?



Light is “regulated” by switches.

## Epigenome

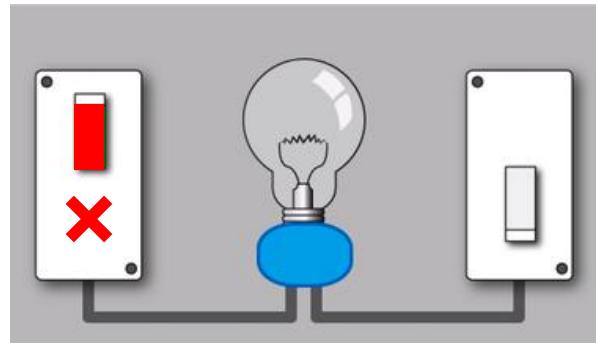


**The epigenome is the collection of all of the epigenetic marks on the DNA in a single cell.**

The epigenomic marks different between different cell types.

Preissl, Gaulton & Ren, Nat Rev Genetics, 2022

# The epigenome and gene regulation is the key to complex functions



Schizophrenia



Brain tumor



Depression



Bipolar Disorder



Alzheimer's Disease

# The burden of brain tumor and mental disorders



Brain tumor



Schizophrenia



Bipolar Disorder



Depression



Alzheimer's Disease

In 2021, ~**83,500** individuals were diagnosed with brain and other CNS tumors in the US.

Average monthly cost: ~\$8,000

Hard to treat and deliver drugs

In 2019, **1 in every 8 people**, or 970 million people around the world were living with a mental disorder.

Top 1 Burden of Diseases in the US.

20 years lived with disability.

Mental disorder, WHO, 2022

US Burden of Disease Collaborators. JAMA. 2013

# Genetic variants and somatic mutations can lead to brain tumor and multiple mental disorders



Brain tumor



Schizophrenia



Bipolar Disorder



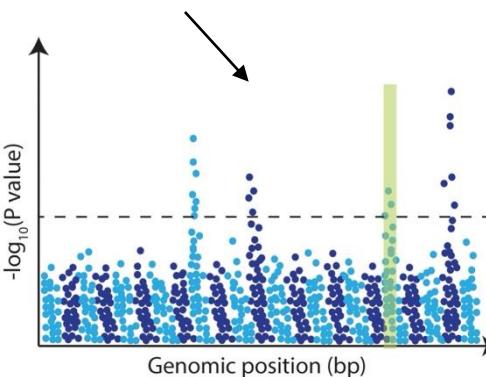
Depression



Alzheimer's Disease

Genome-wide association study  
(GWAS)

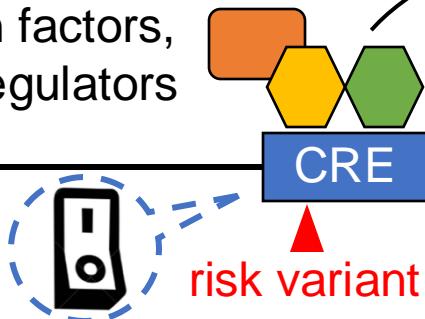
>90% of the variants are non-coding and yet to be functionally interpreted



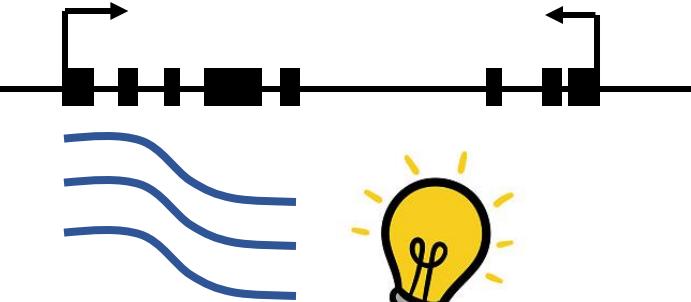
# A paradigm for understanding the noncoding disease risk variants, and the key barriers

How do risk variants affect transcription factor binding, and which ones?

Transcription factors,  
Chromatin regulators



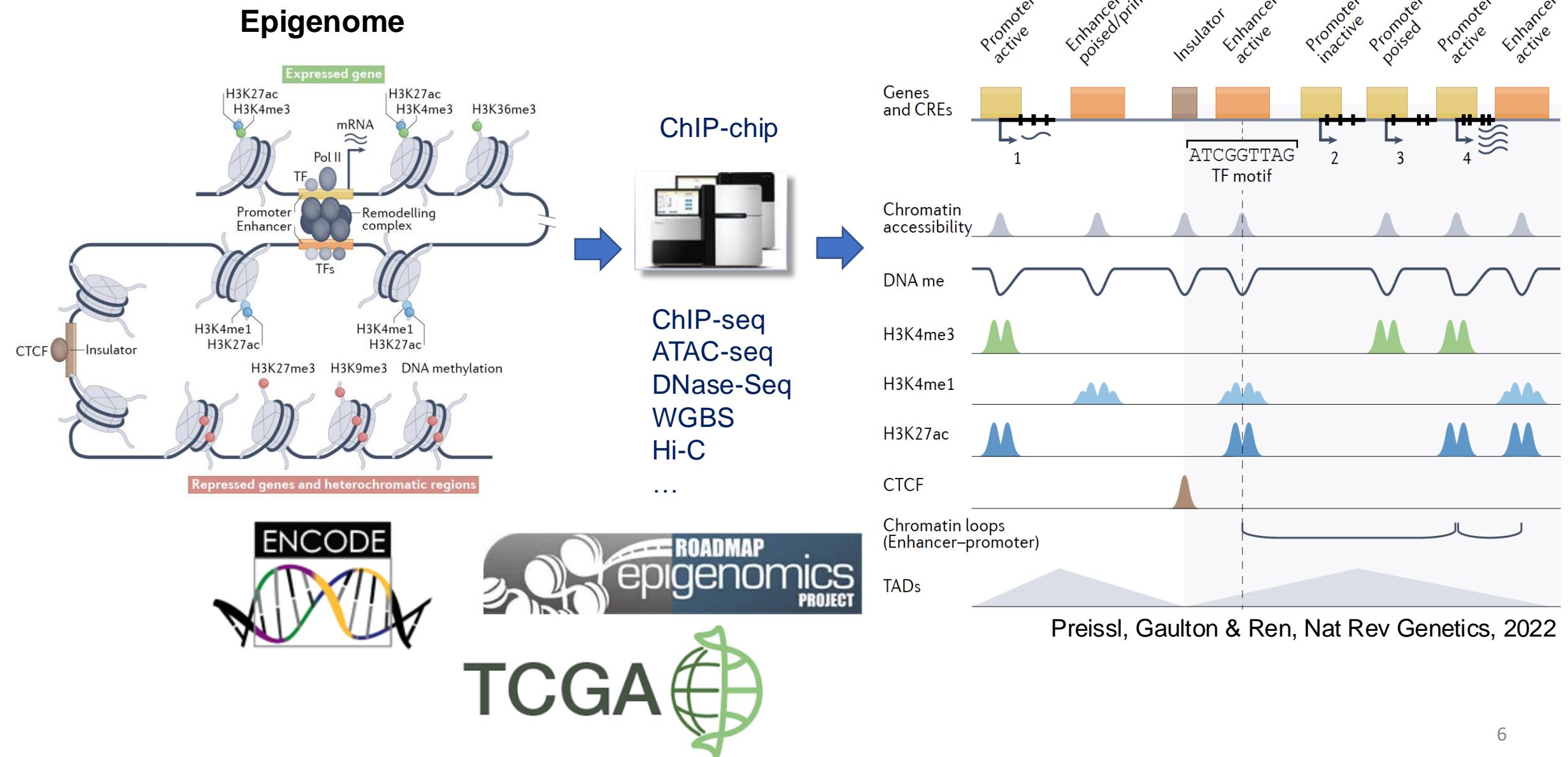
What target genes are controlled by the *cis* regulatory elements (CREs)?



What sequences act as *cis* regulatory elements (CREs), where and when do they act?

It has been estimated that >50% of risk variants act in this way (Nasser J, ... Engreitz, Nature, 2021)

# Epigenomic profiling reveals genome regulation



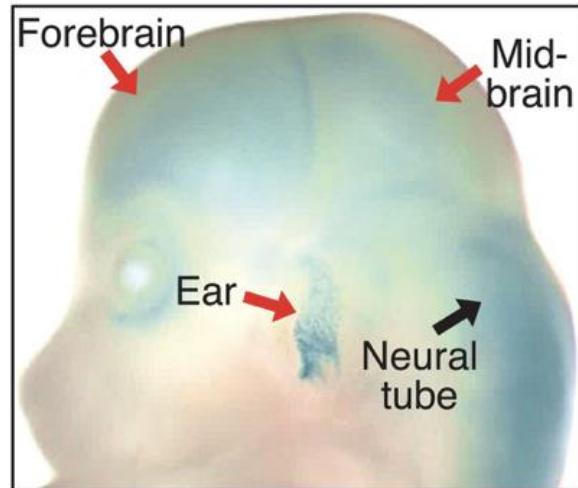


## A noncoding single-nucleotide polymorphism at 8q24 drives *IDH1*-mutant glioma formation

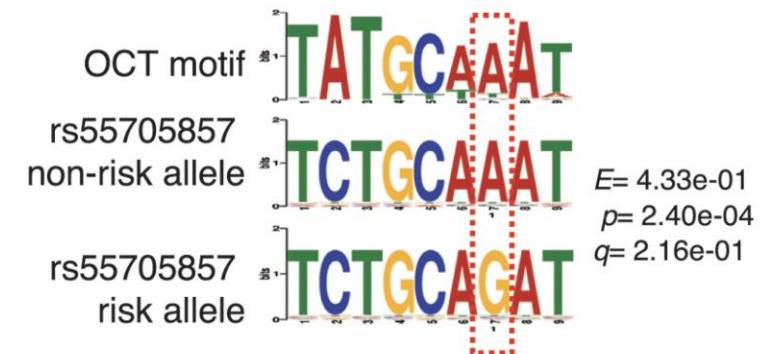
CONNOR YANCHUS , KRISTEN L. DRUCKER , THOMAS M. KOLLMAYER , RICKY TSAI , WARREN WINICK-NG , MINGGAO LIANG , AHMAD MALIK, JUDY PAWL-  
ING , SILVANA B. DE LORENZO, [...], AND DANIEL SCHRAMEK +49 authors [Authors Info & Affiliations](#)

SCIENCE • 6 Oct 2022 • Vol 378, Issue 6615 • pp. 68-78 • DOI: 10.1126/science.abj2890

rs55705857 resides within a brain-specific enhancer



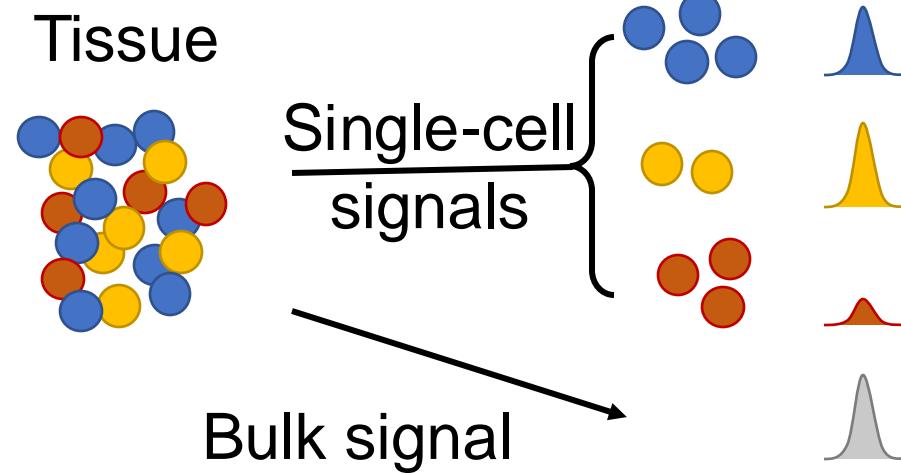
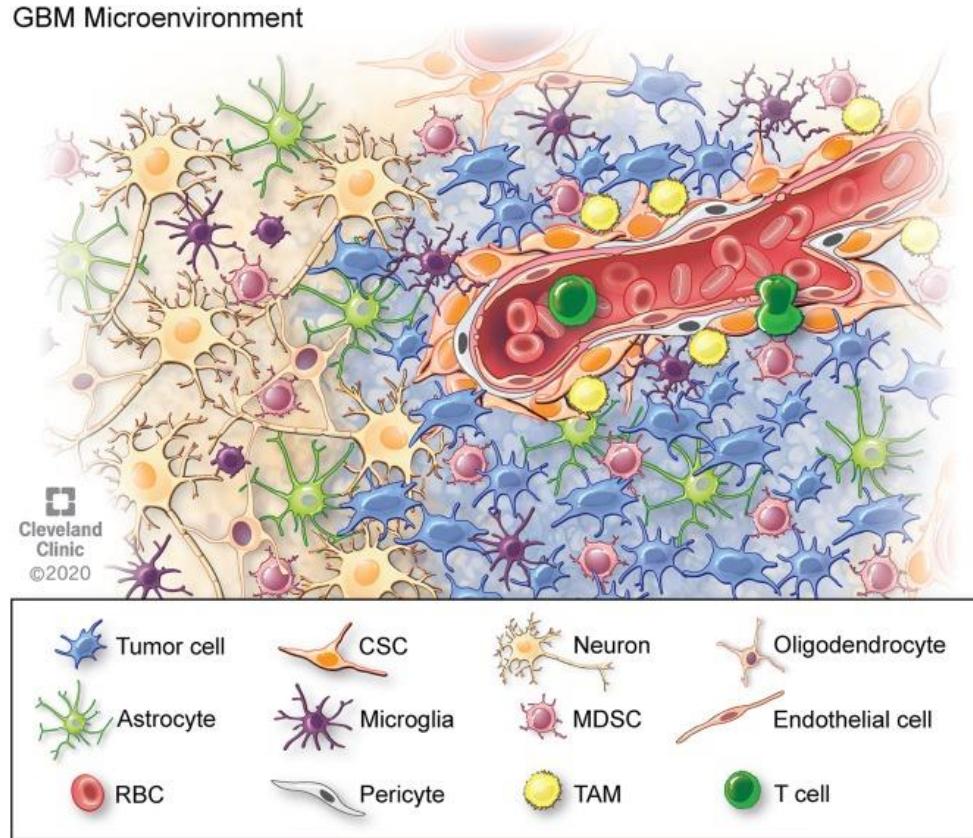
This risk variant disrupts OCT2/4 binding



..., allowing increased interaction with the Myc promoter and increased Myc expression in ~40% of *IDH*-mutant low-grade glioma (LGG) patients

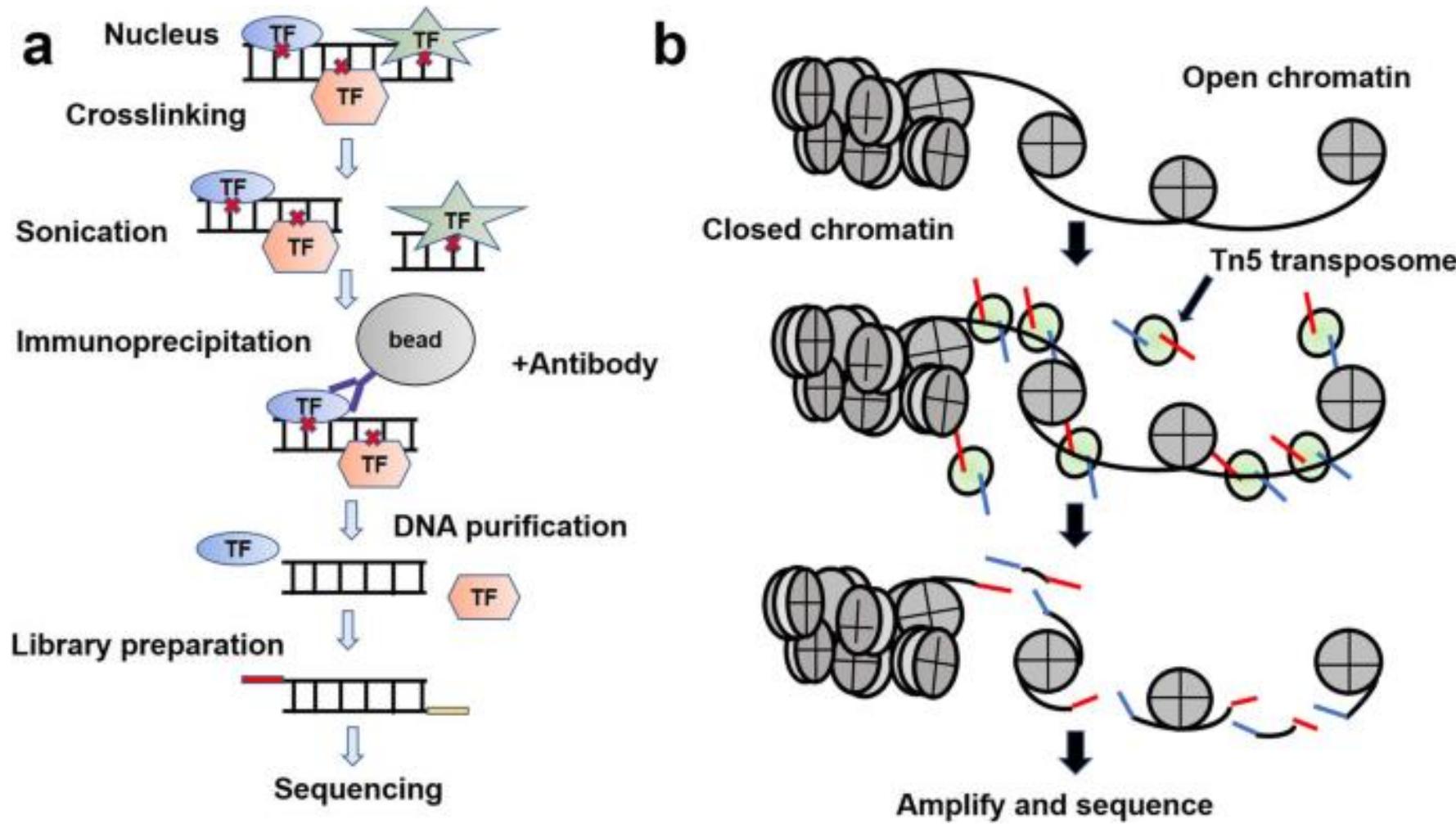
But, it's at BULK level!

# Single cell technologies solve cell heterogeneity



Lauko, et al. 2022

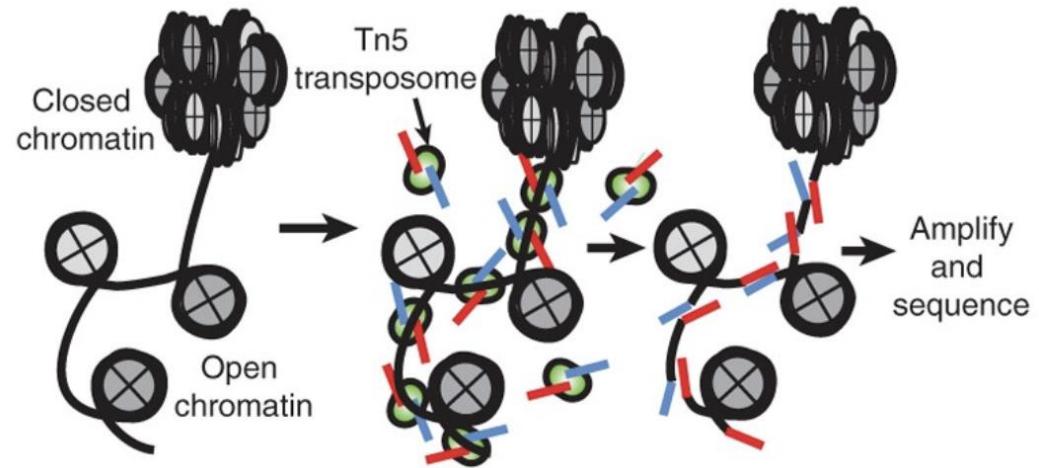
# The difference between ChIP-seq and ATAC-seq



**ATAC-seq** identifies accessible DNA regions by probing **open chromatin** with hyperactive mutant **Tn5 Transposase** that inserts sequencing adapters into open regions of the genome.

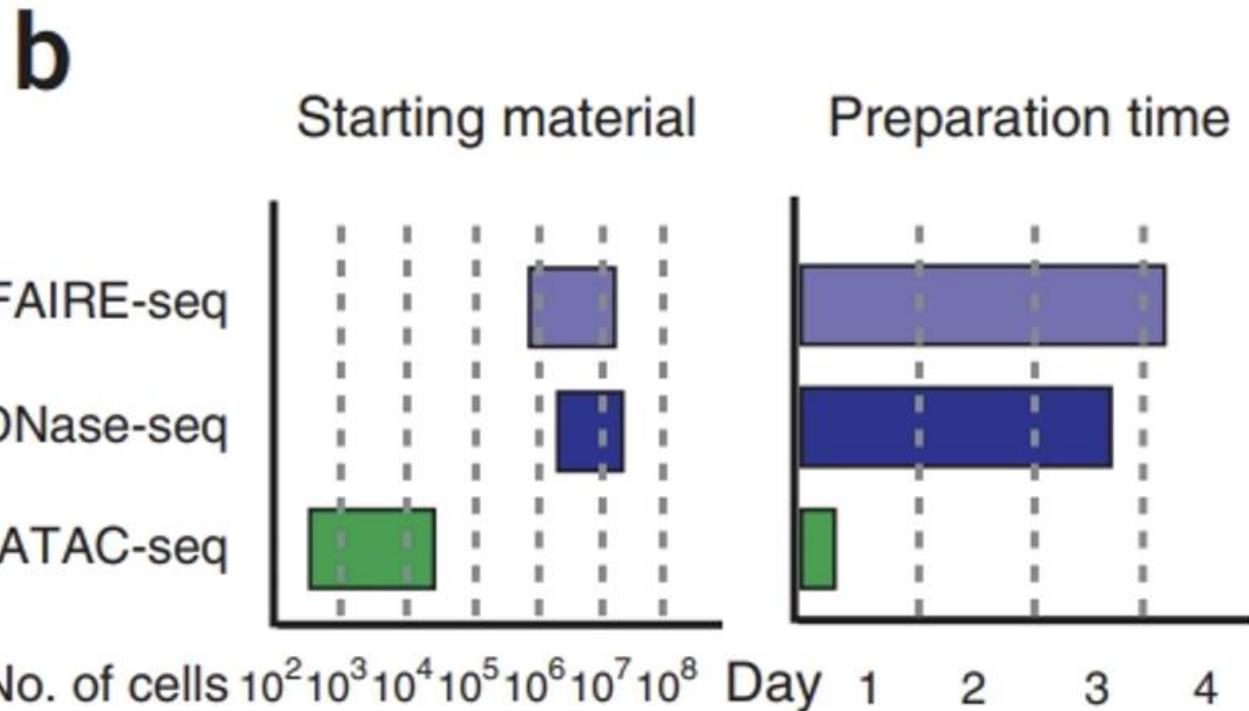
Buenrostro, et al. Nature Methods, 2013

# The advantage of ATAC-seq



**ATAC-seq** identifies accessible DNA regions by probing **open chromatin** with hyperactive mutant **Tn5 Transposase** that inserts sequencing adapters into open regions of the genome.

Buenrostro, et al. Nature Methods, 2013



# The data analysis of ATAC-seq

Yan et al. *Genome Biology* (2020) 21:22  
<https://doi.org/10.1186/s13059-020-1929-3>

Genome Biology

REVIEW

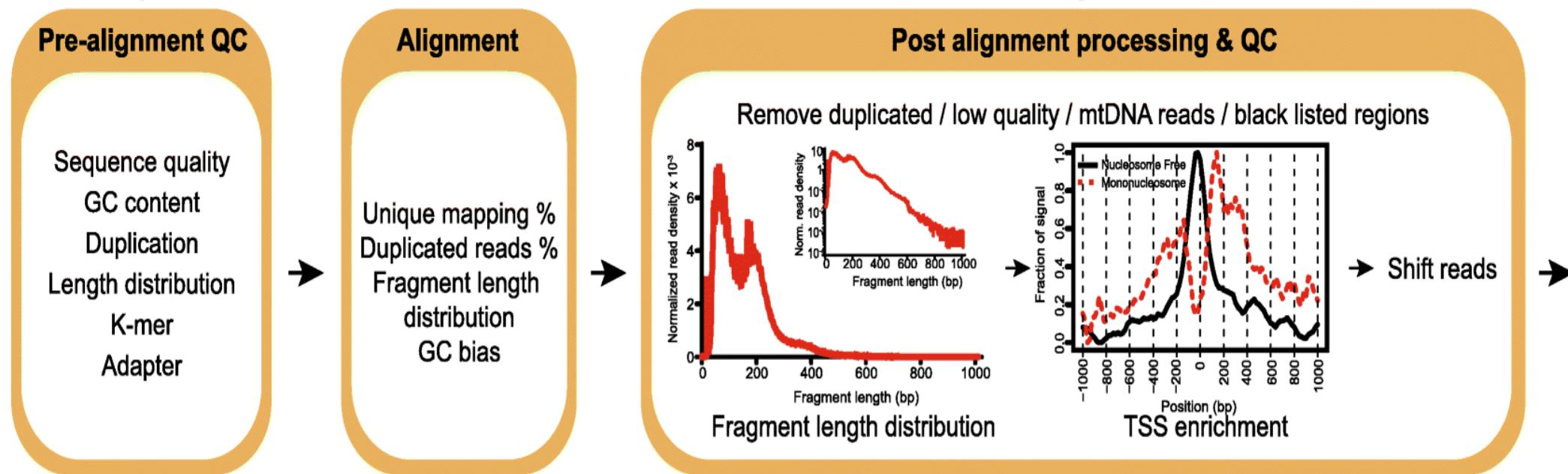
Open Access

## From reads to insight: a hitchhiker's guide to ATAC-seq data analysis

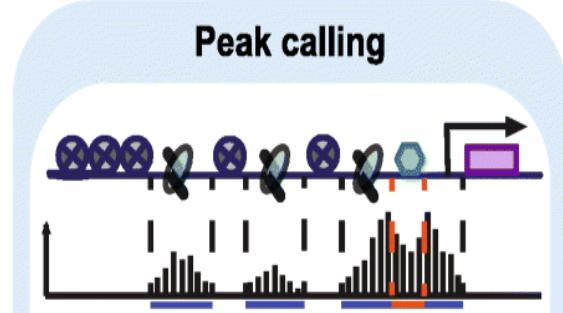


Feng Yan<sup>1</sup>, David R. Powell<sup>2</sup>, David J. Curtis<sup>1,3</sup> and Nicholas C. Wong<sup>1,2\*</sup> 

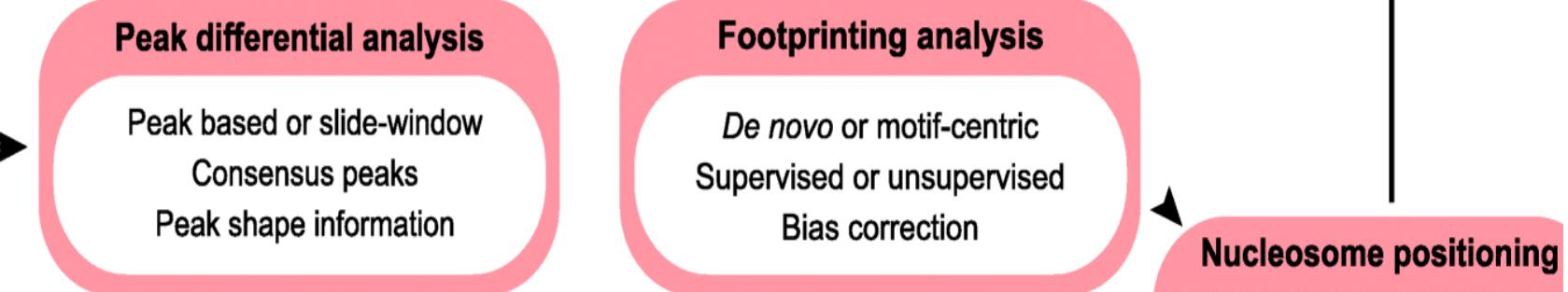
## Pre-analysis



## Core analysis

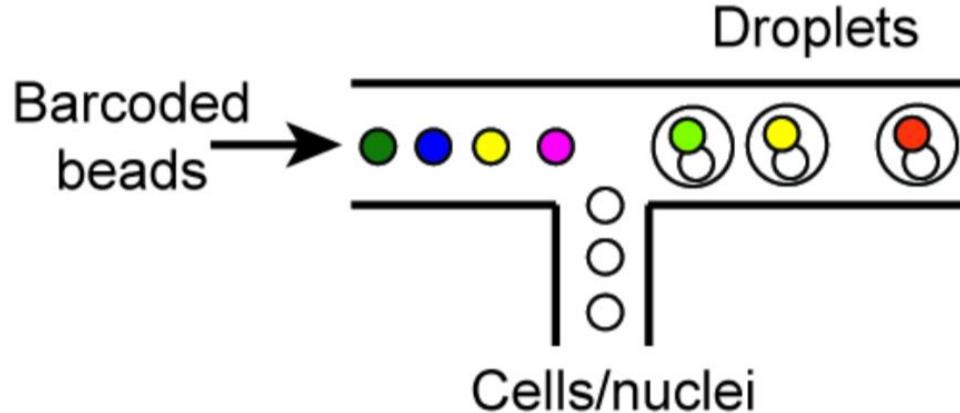


## Advanced analysis

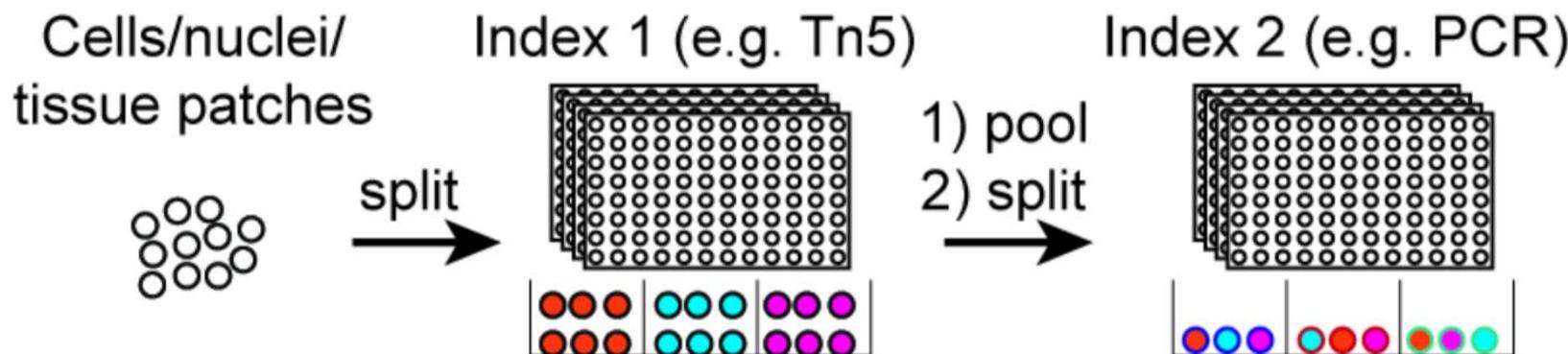


# Single-nucleus ATAC-seq – different platforms

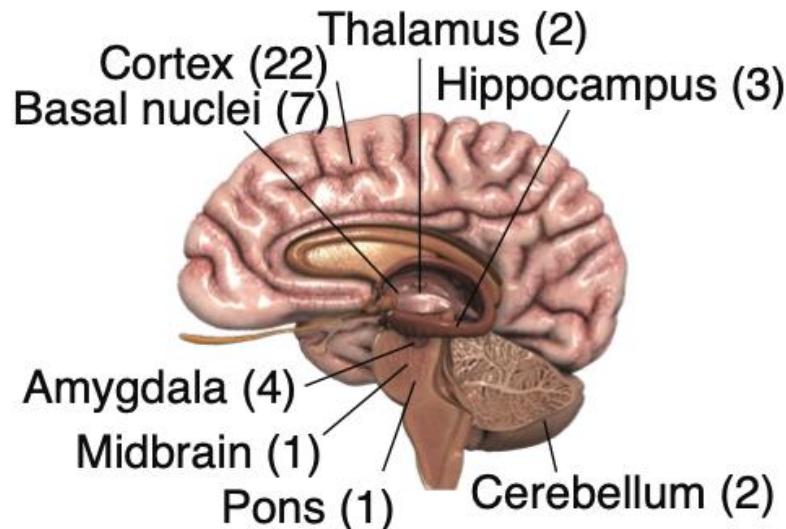
## □ Droplet-based (e.g. 10x Genomics)



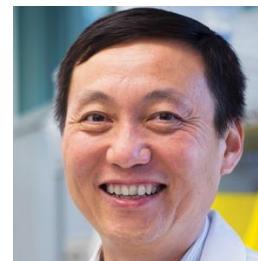
## □ Split-pool-barcoding/single cell combinatorial indexing (sci-)



# Single cell chromatin accessibility atlas of adult human brains



- 42 brain regions
- 3 neurotypical adult donors
- 1.1 million cells, 107 cell types



Bing Ren, Ph.D.  
(UCSD)



Marga Behrens,  
Ph.D. (Salk)



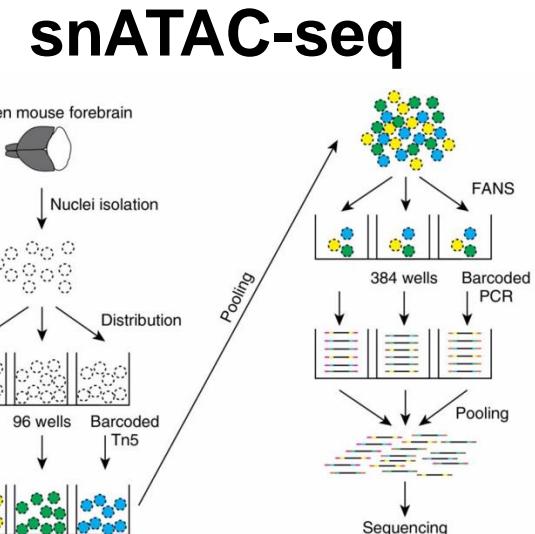
Ed Lein, Ph.D.  
(Allen Inst.)



Sebastian Preissl,  
Ph.D. (UCSD)

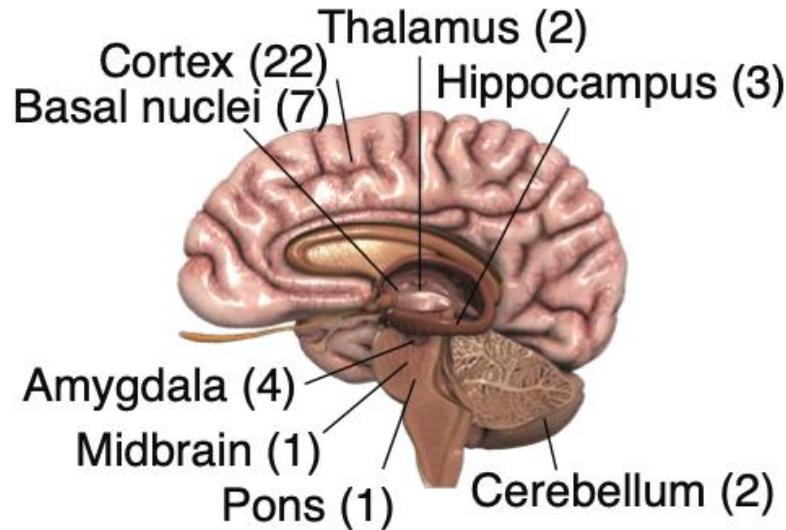


Joe Ecker, Ph.D.  
(Salk/HHMI)



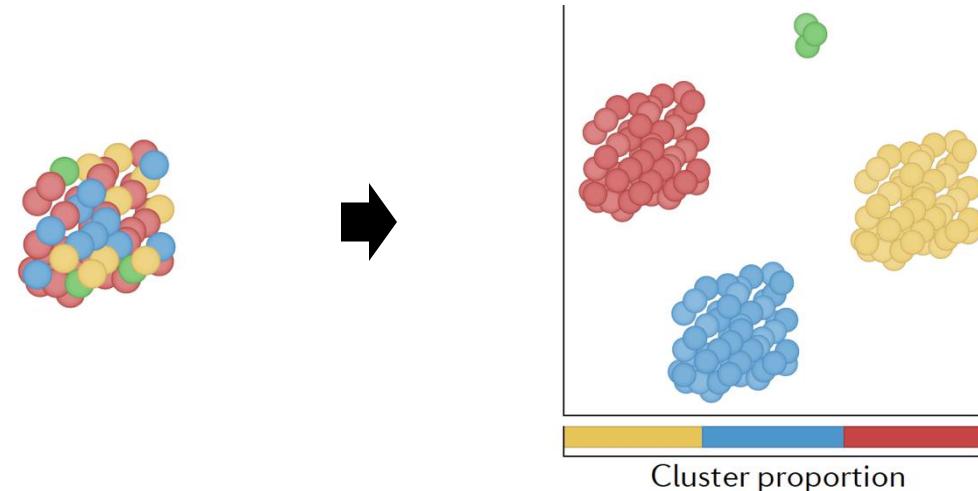
Preissl, et al. 2018

# Single cell chromatin accessibility atlas of adult human brains

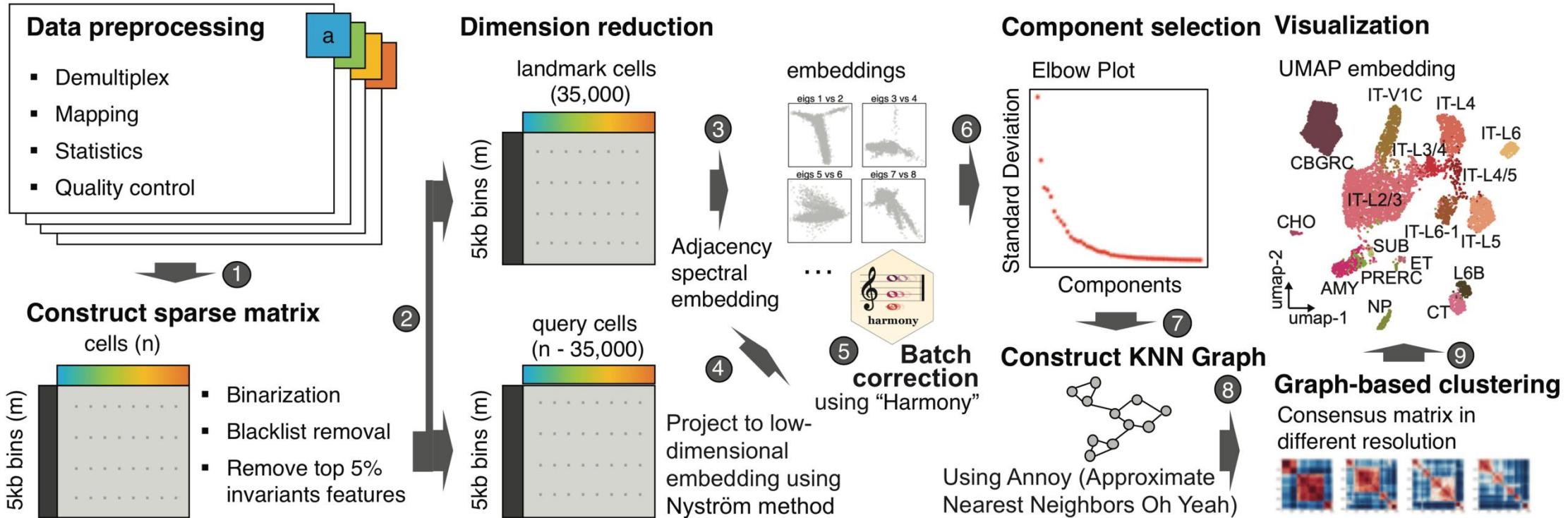


- 42 brain regions
- 3 neurotypical adult donors
- 1.1 million cells, 107 cell types

Single cell clustering



# Cell clustering pipeline of snATAC-seq



# Popular tools

nature methods

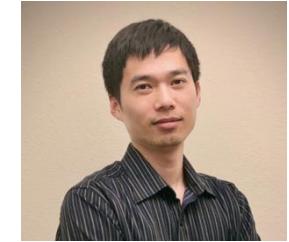
Article

## SnapATAC2



<https://doi.org/10.1038/s41592-023-02139-9>

**A fast, scalable and versatile tool for analysis of single-cell omics data**



Dr. Kai Zhang



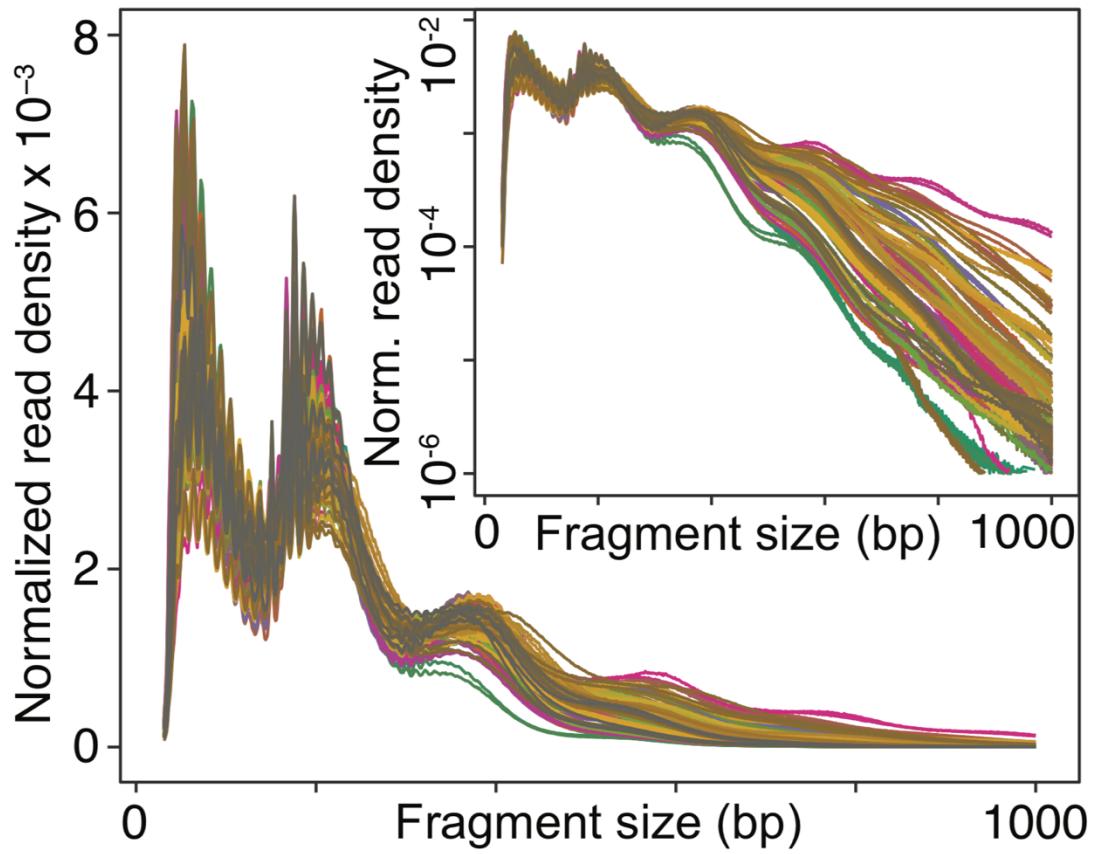
pycisTopic



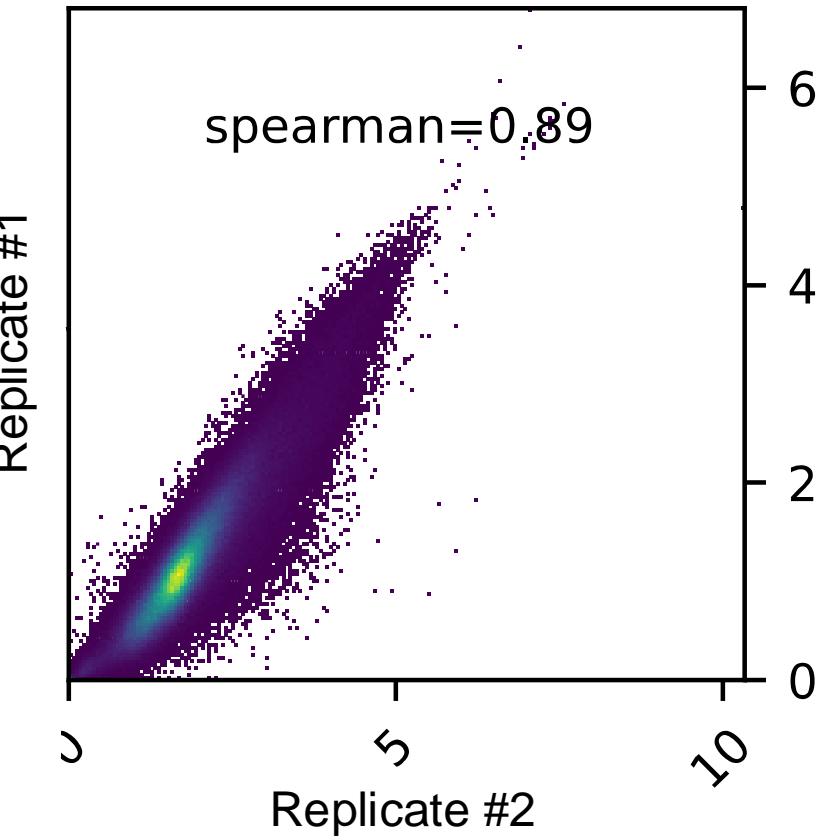
chromVAR

# Quality control of snATAC-seq – Bulk level

The distribution of fragment size

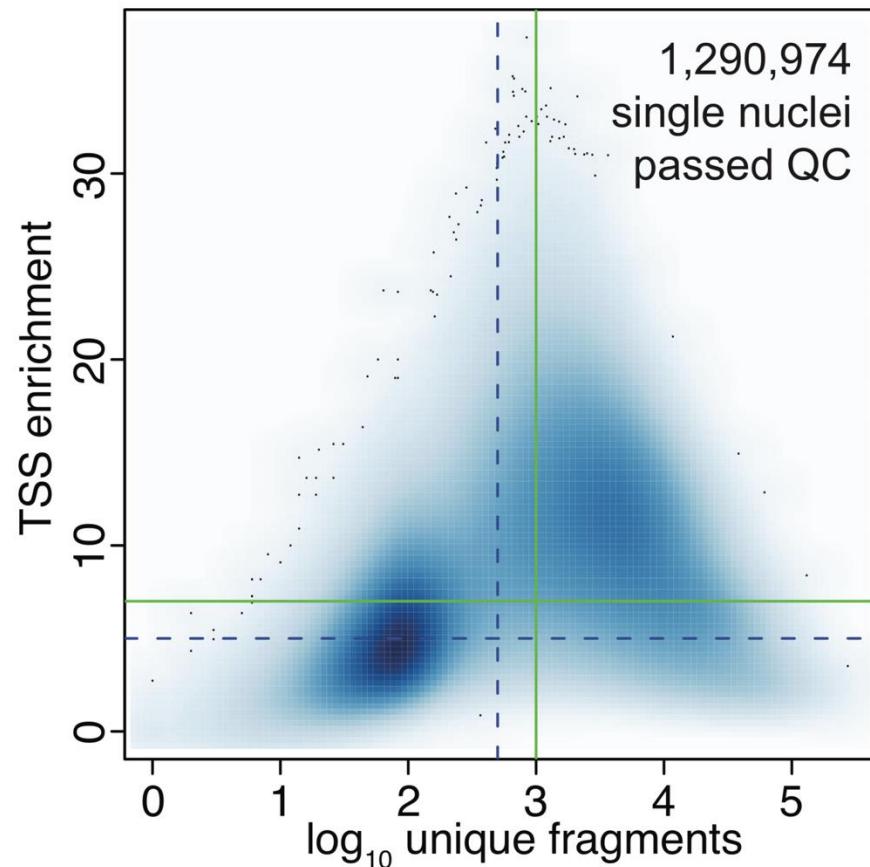


Correlation between replicates

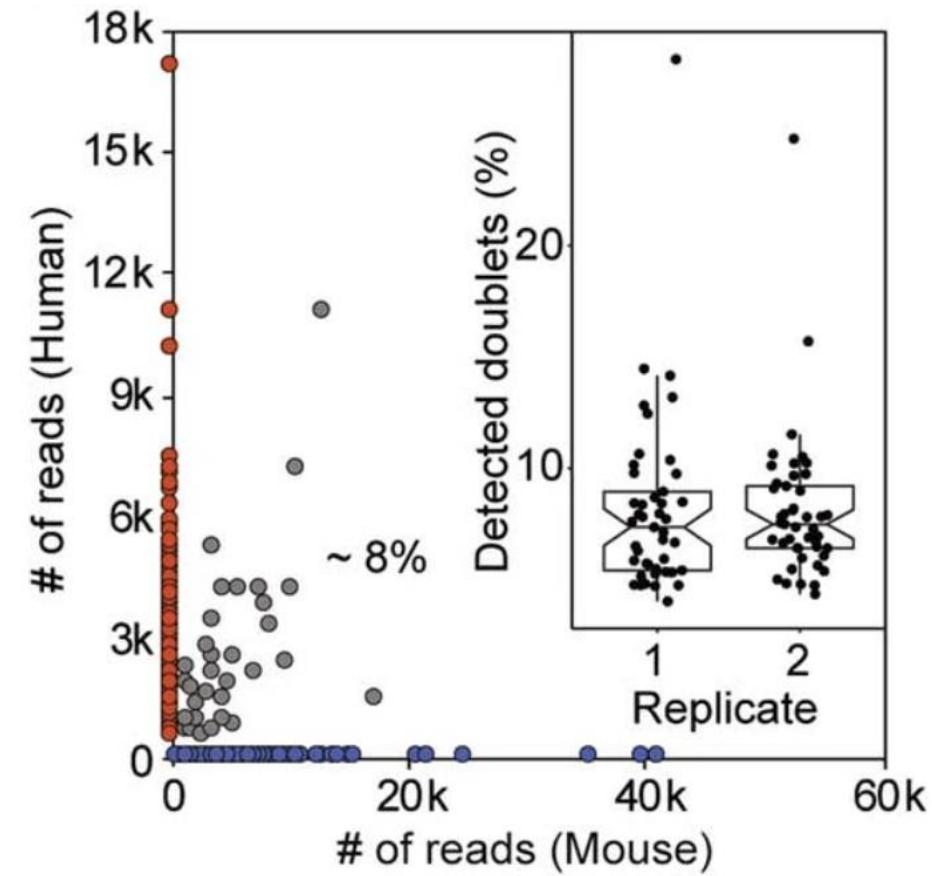


# Quality control of snATAC-seq – Single-cell level

TSS enrichment & unique fragments

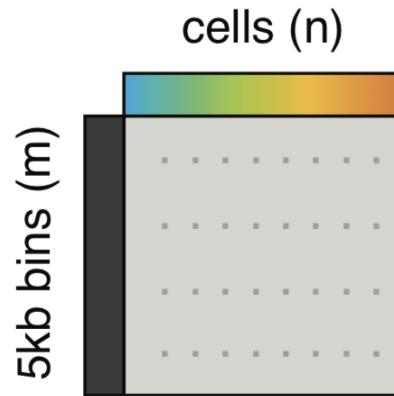


Doublet removal



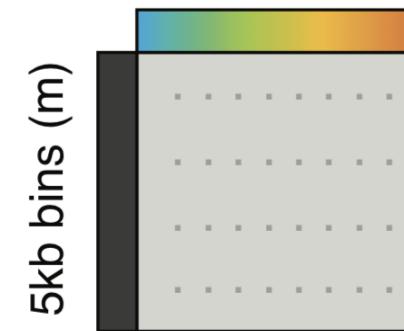
# Challenges in clustering extremely large sparse matrix

- Million cells
- Signals on whole genome
- Very sparse

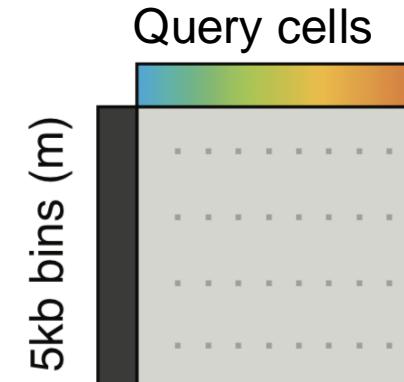
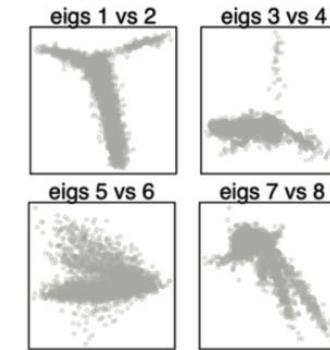


- Memory issues
- Time consuming

Landmark cells



Dimension reduction



Nyström algorithm  
an efficient technique  
to generate low-rank  
matrix approximations

# Challenges in identifying robust cell clusters

Resolution parameters in community detection algorithm (Leiden)

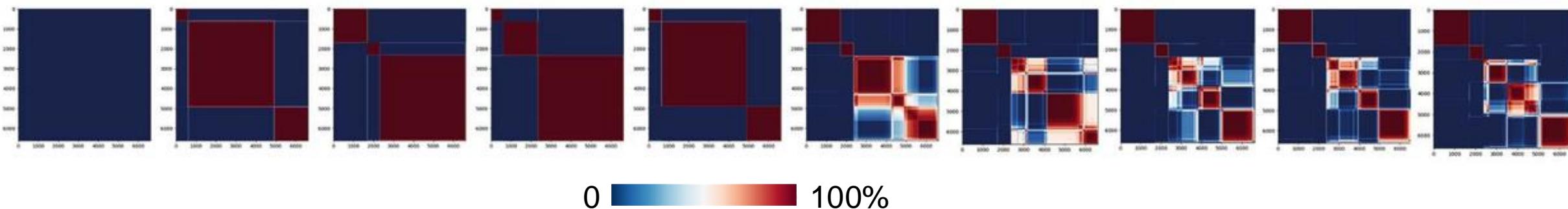
0.1

1.0

UMAP embedding by cluster

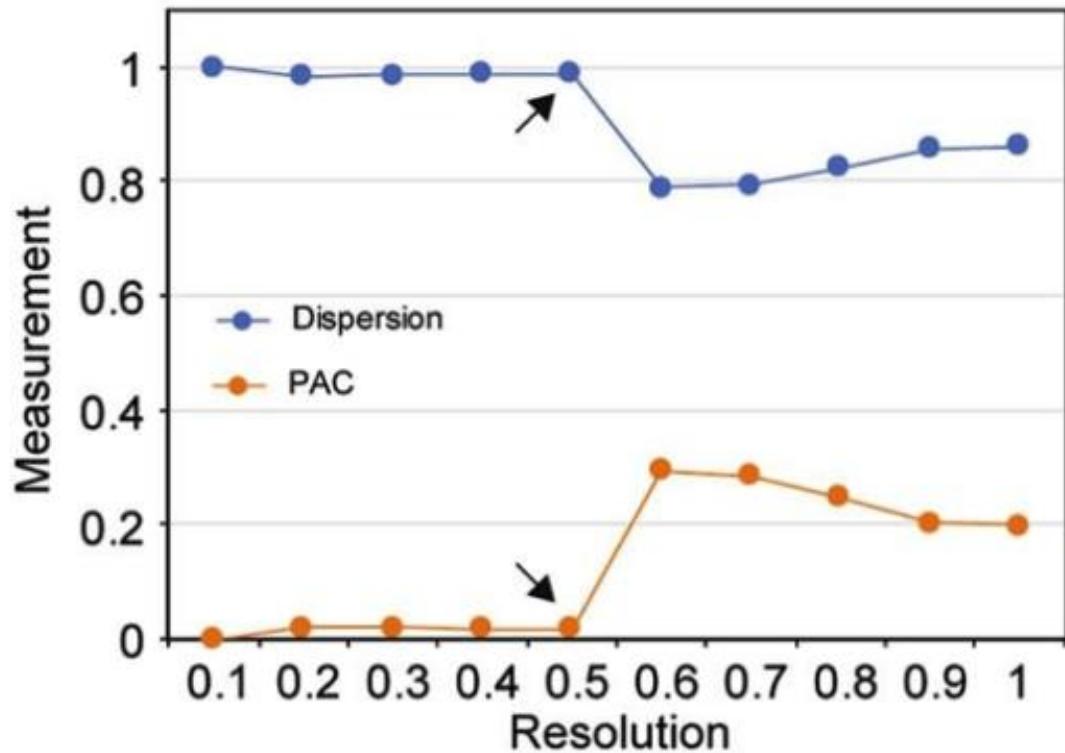


Generate consensus matrix from iterative runs



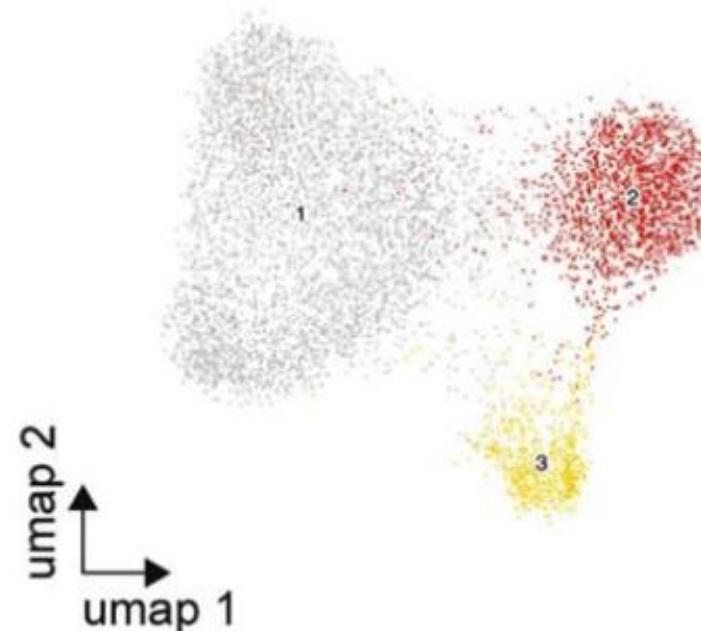
# Challenges in identifying robust cell clusters

## Selecting the most robust resolution

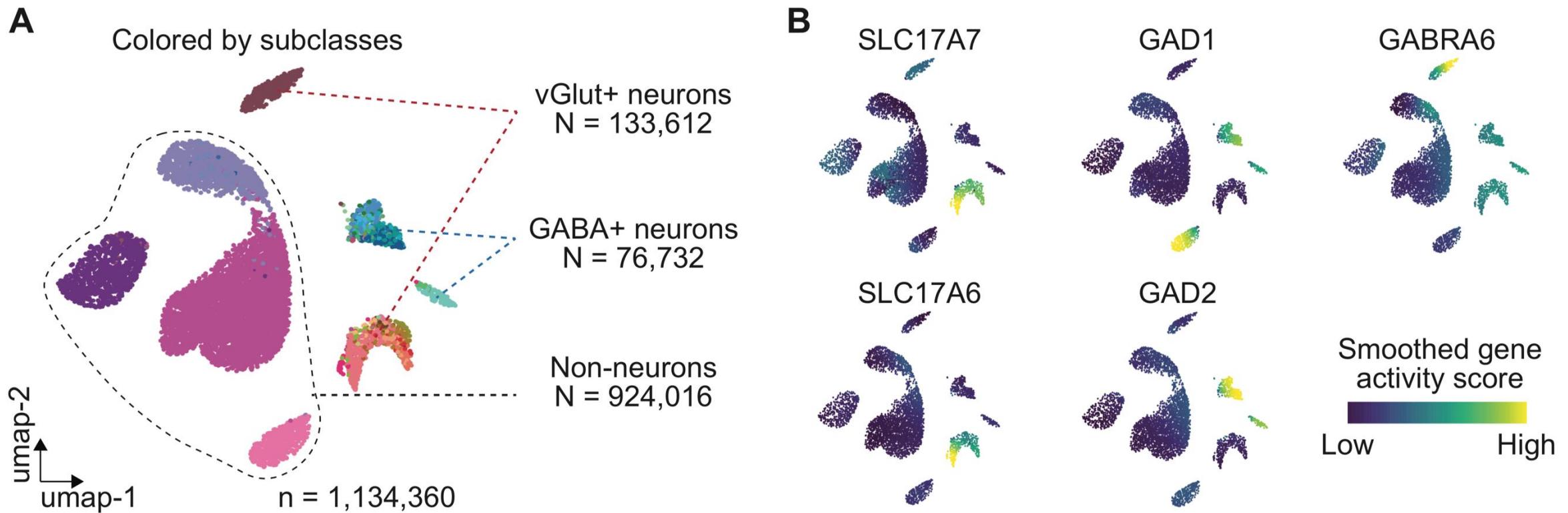


**Dispersion:** dispersion coefficient

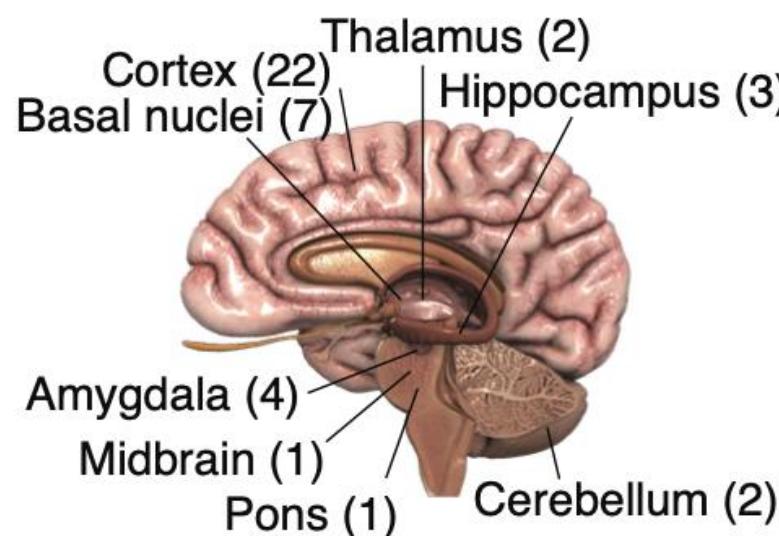
**PAC:** proportion of ambiguous clustering score



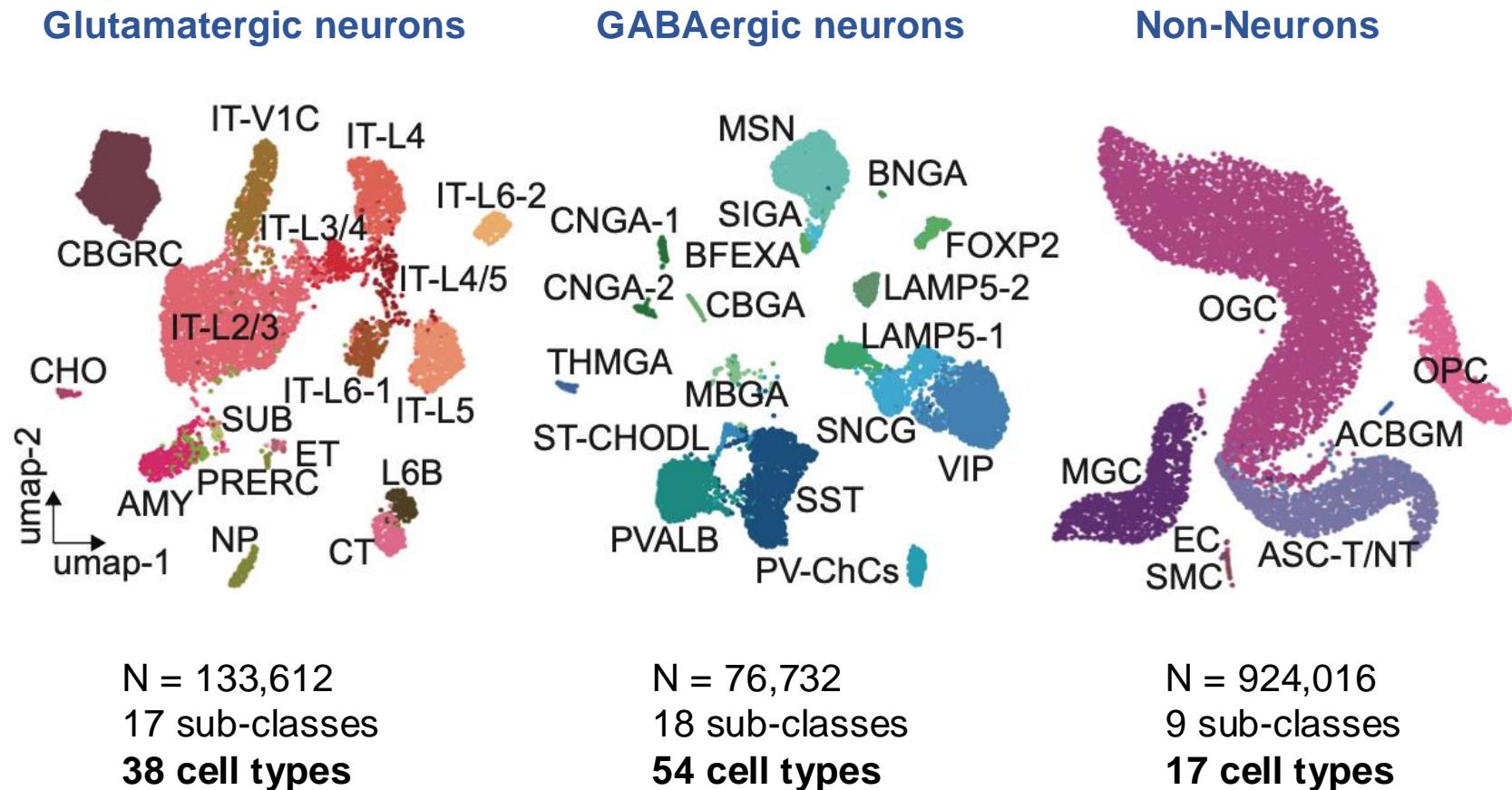
# Cell cluster annotation based on TSS/promoter + gene body



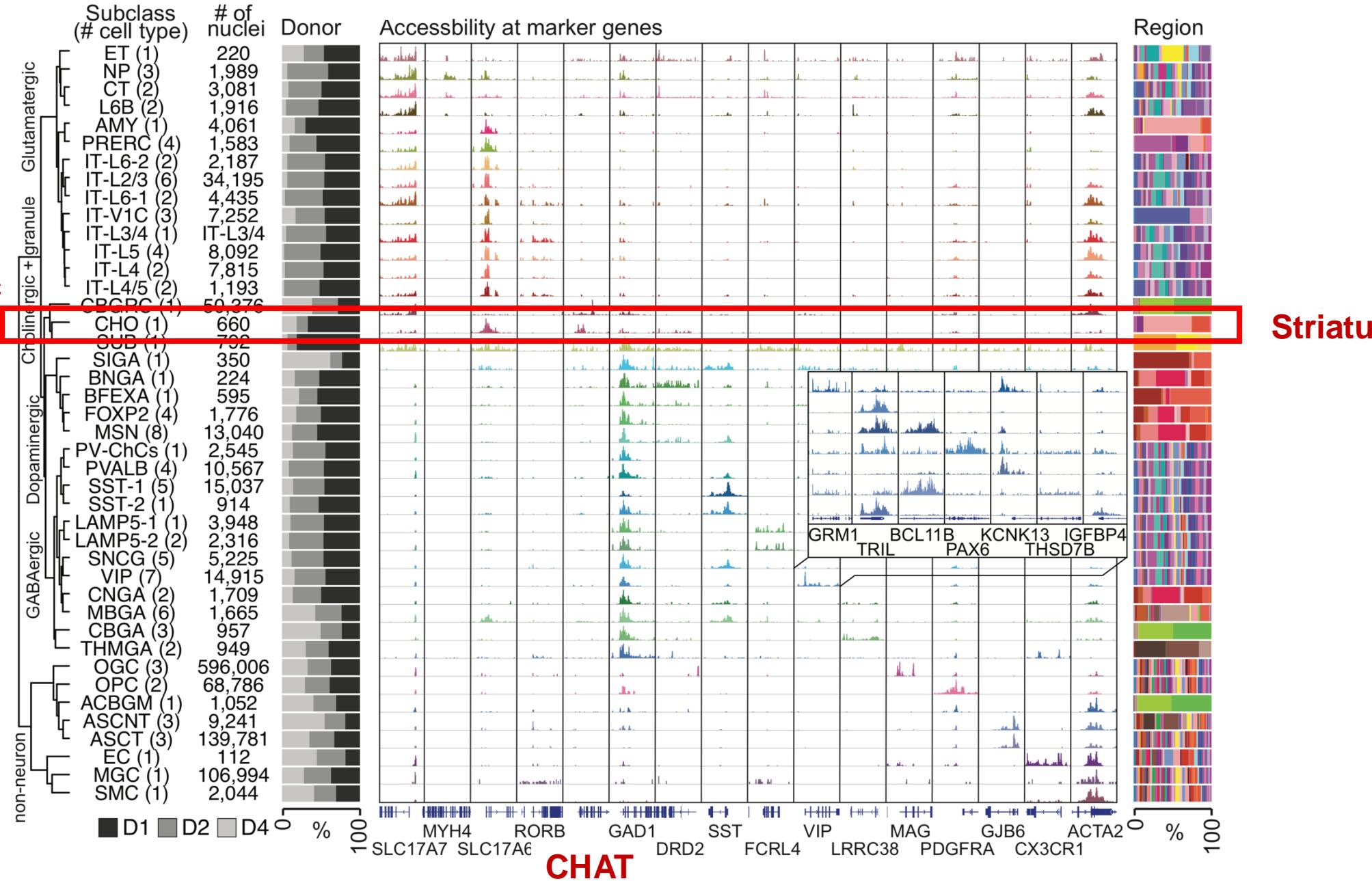
# 42 cell subclasses and 107 cell types identified from >1.1 million single nuclei in adult human brain



- 42 brain regions
- 3 adult donors
- 1.1 million cells, 107 cell types

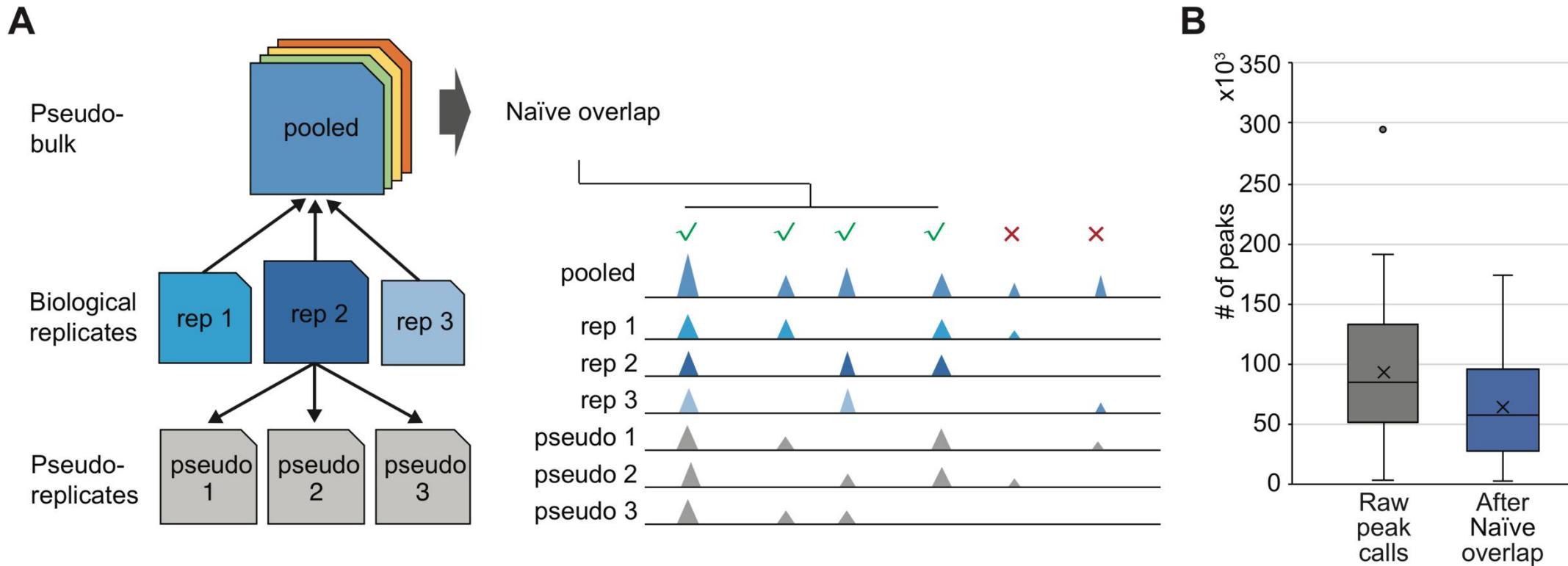


## Cholinergic neurons (CHO)



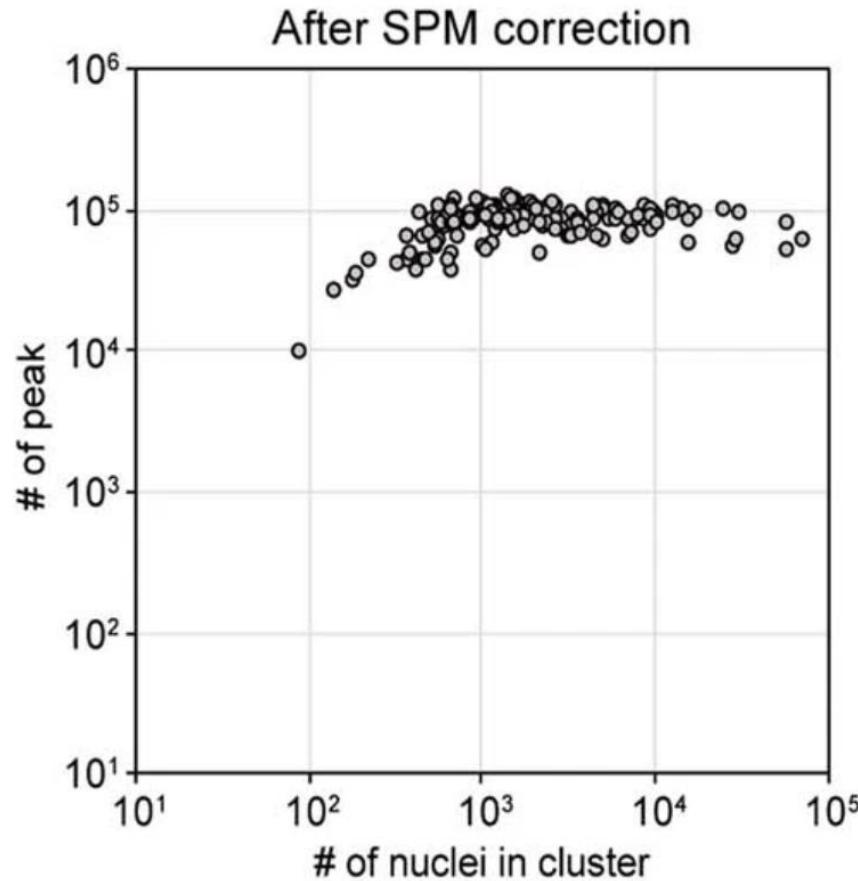
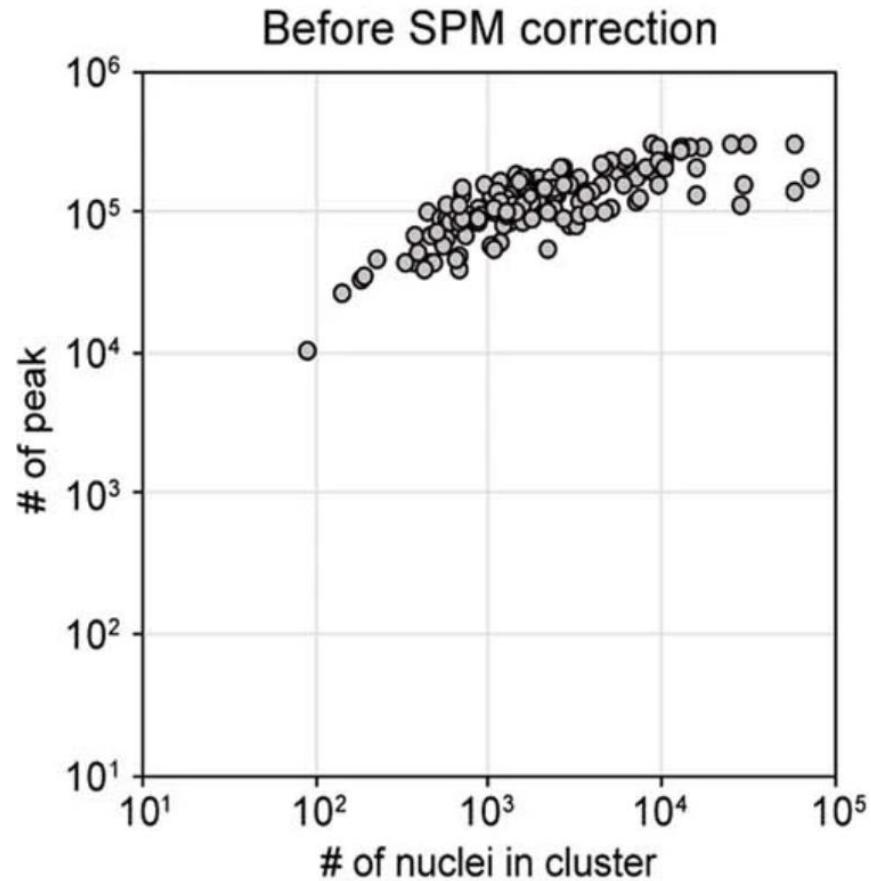
# Peak calling pipeline

## Identify reproducible peaks



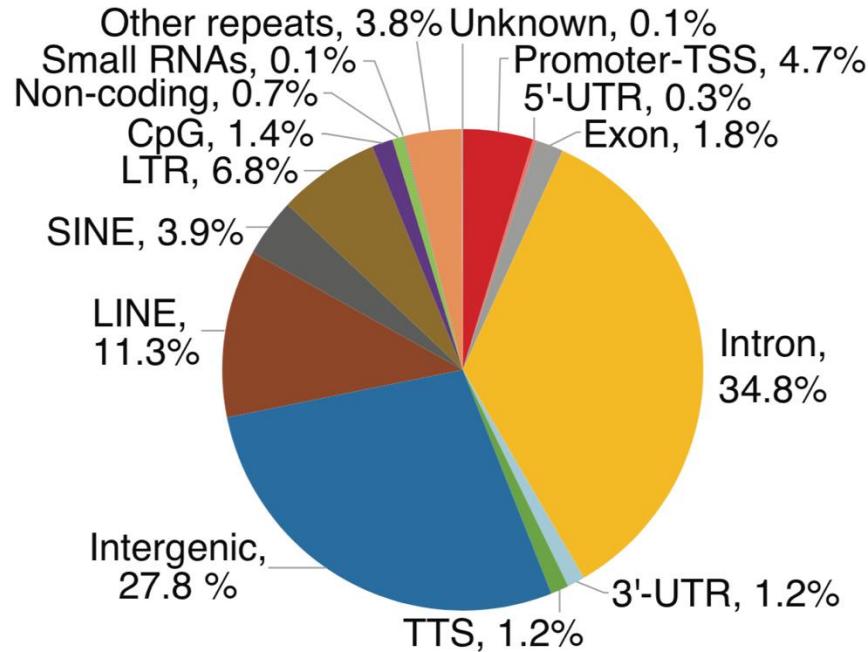
# Peak calling pipeline

Correction for potential bias from sequencing depth (# of cells)

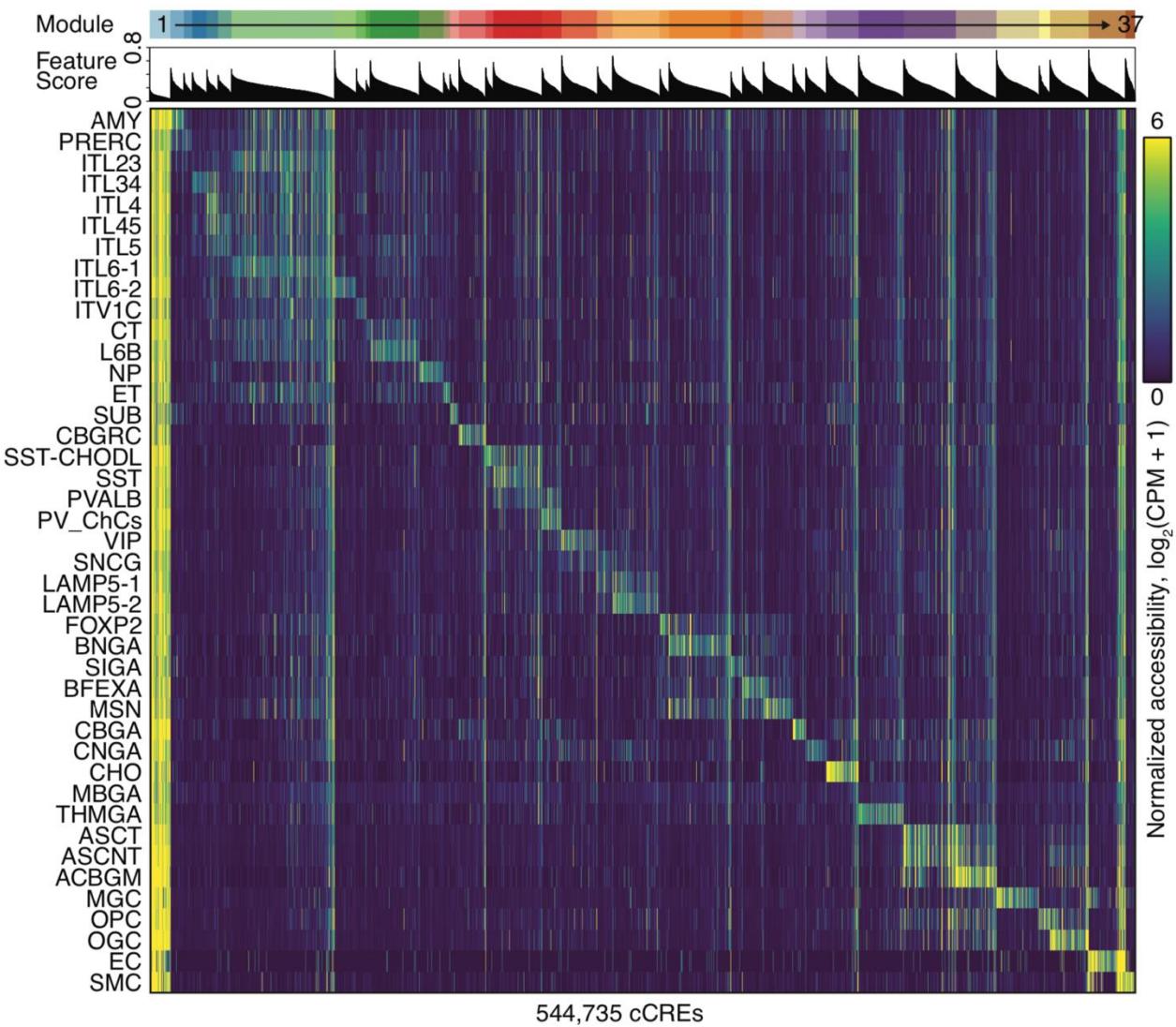


# Brain cCREs display highly variable chromatin accessibility

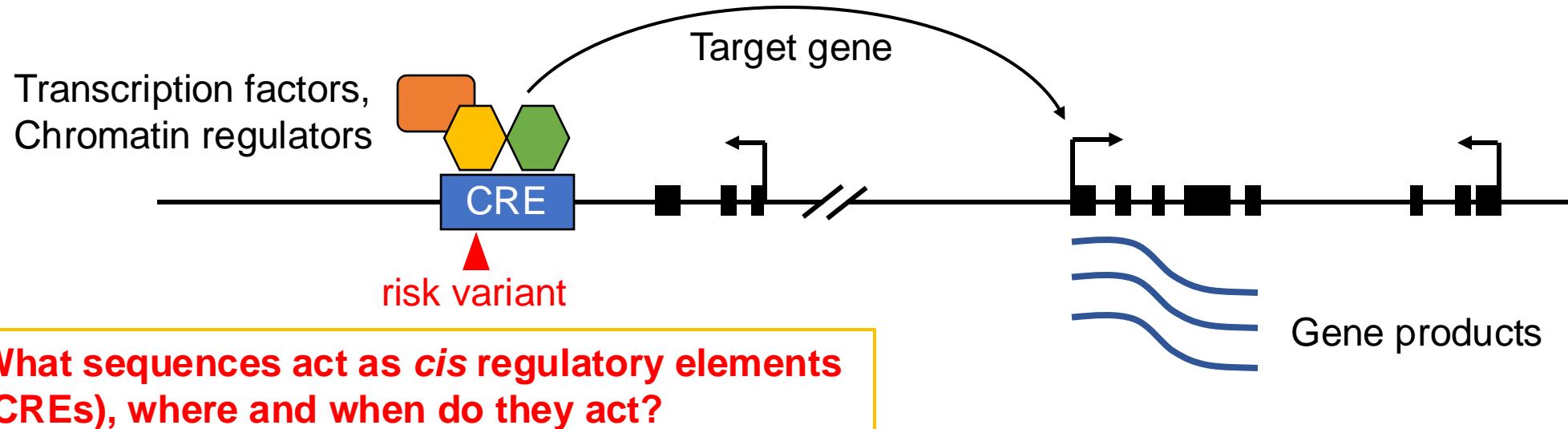
**544,735 candidate cis-regulatory elements (cCREs)**



**~ 46.9% are new cCREs comparing to ENCODE**

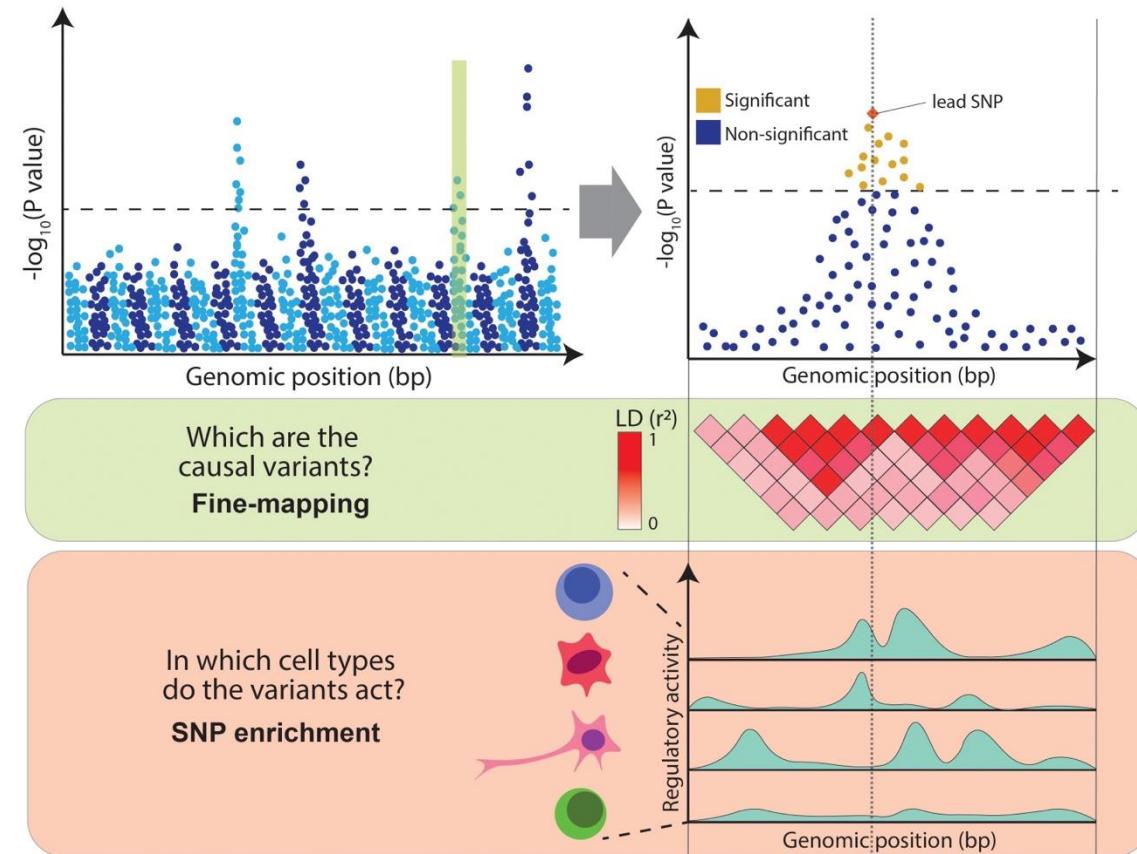


# A paradigm for understanding the noncoding disease risk variants, and the key barriers



It has been estimated that >50% of risk variants act in this way (Nasser J, ... Engreitz, Nature, 2021)

# Predicting Disease Relevant Cell Type(s)



Transcription factors,  
Chromatin regulators

CRE  
risk variant

Target gene

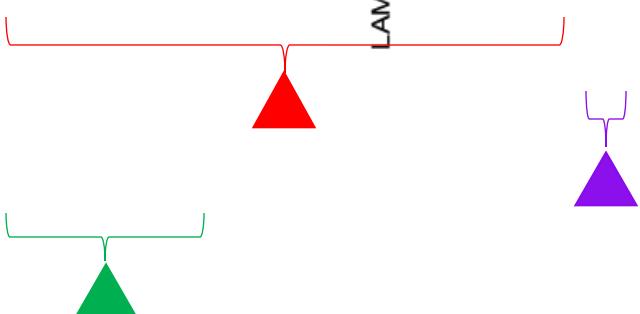
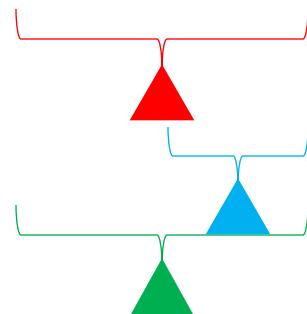
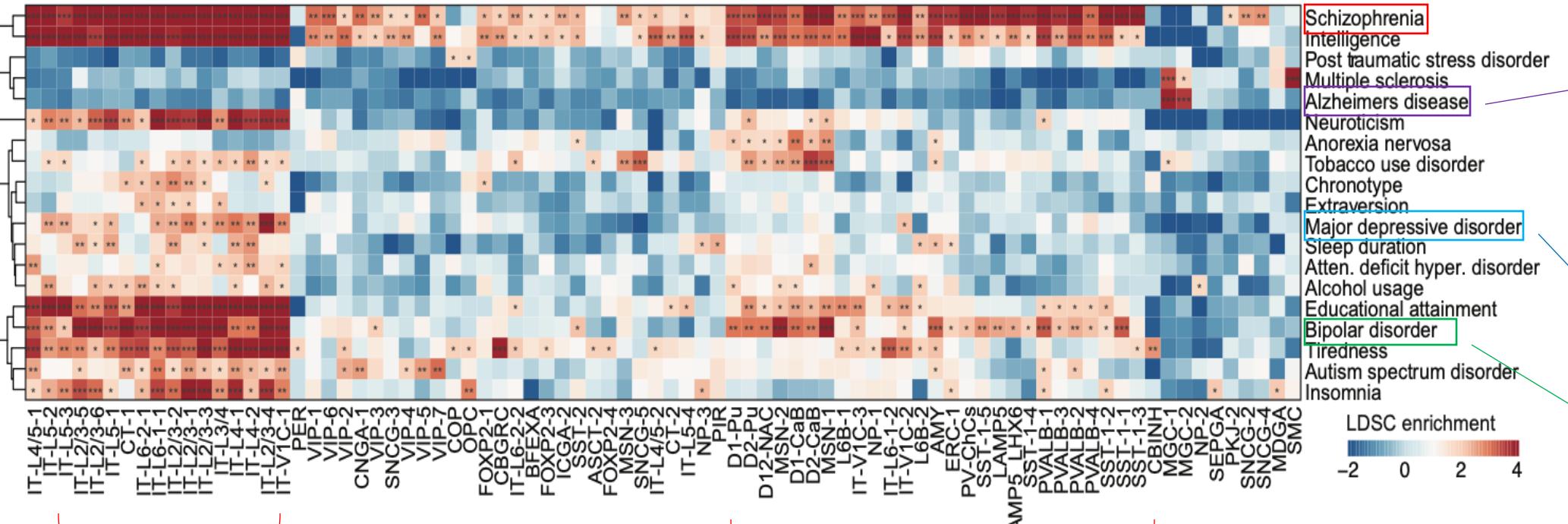
Gene products

Disease risk variants are enriched in  
enhancer of disease relevant cell type(s)?

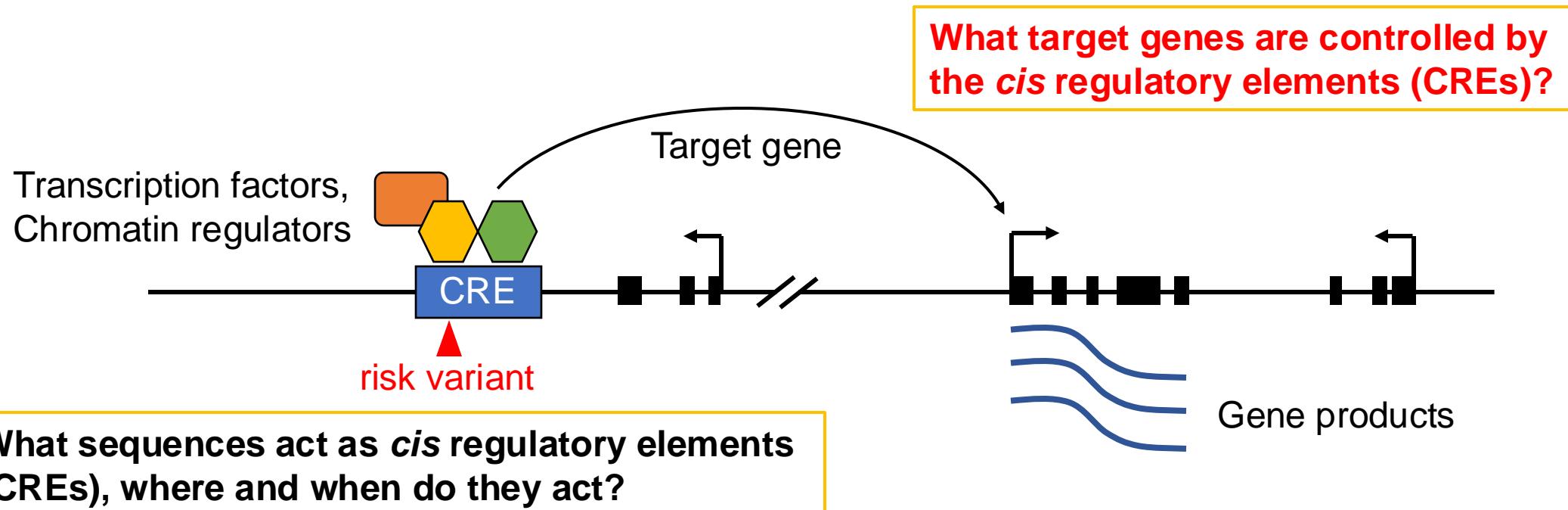
# Cell-type resolved chromatin landscapes help interpret non-coding variants in mental diseases



Linkage disequilibrium score regression (LDSC) analysis

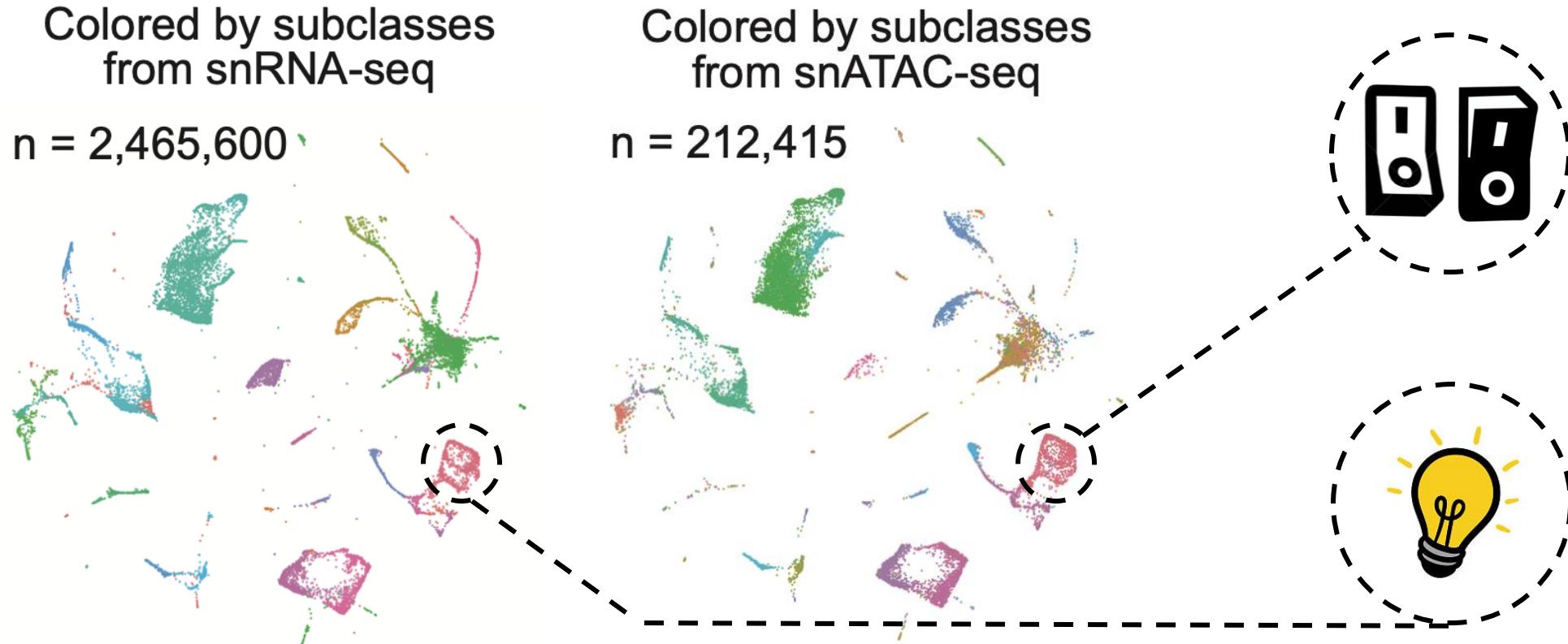


# A paradigm for understanding the noncoding disease risk variants, and the key barriers



It has been estimated that >50% of risk variants act in this way (Nasser J, ... Engreitz, Nature, 2021)

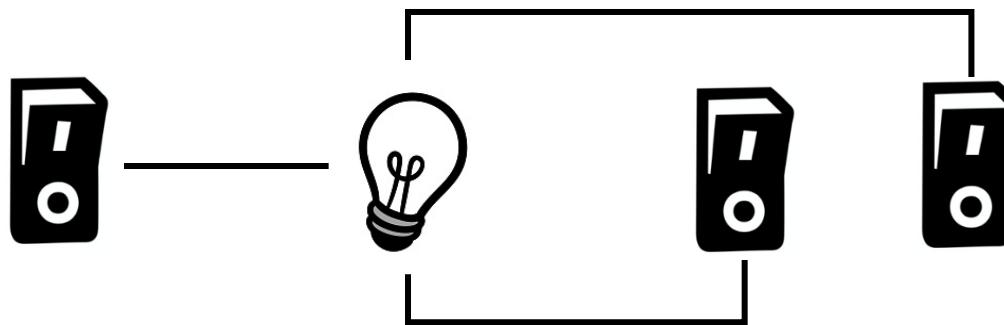
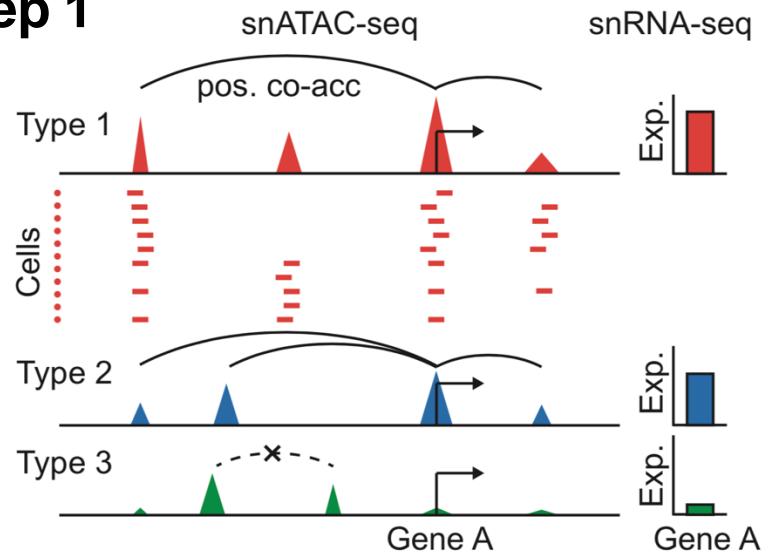
# Integration with snRNA-seq



scRNA-seq data: Linnarsson lab

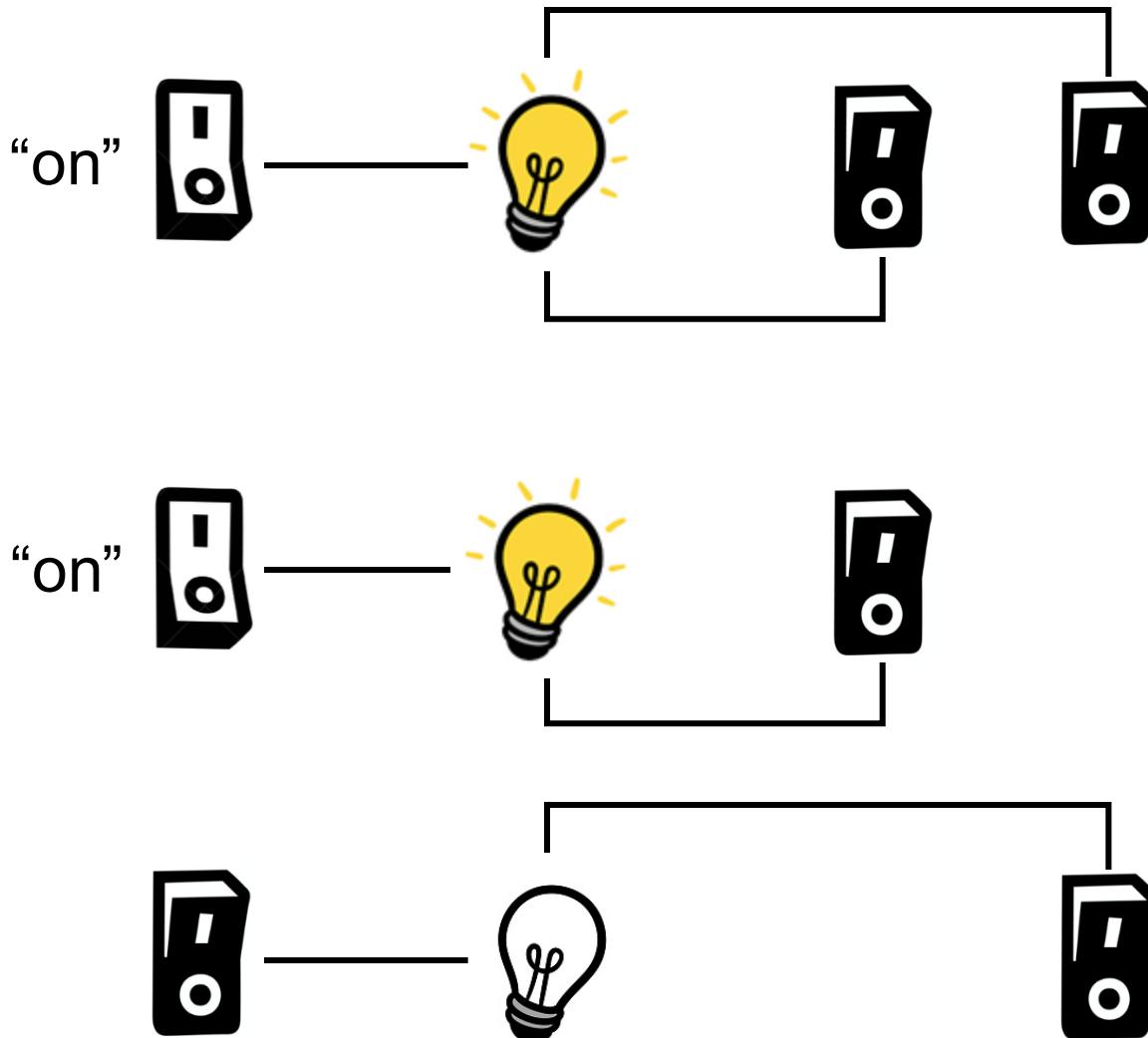
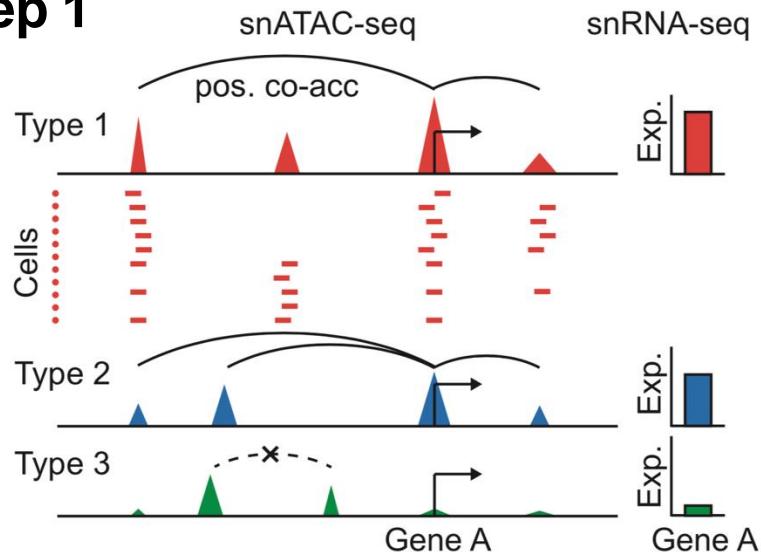
# Linking distal cCREs to target genes

## Step 1

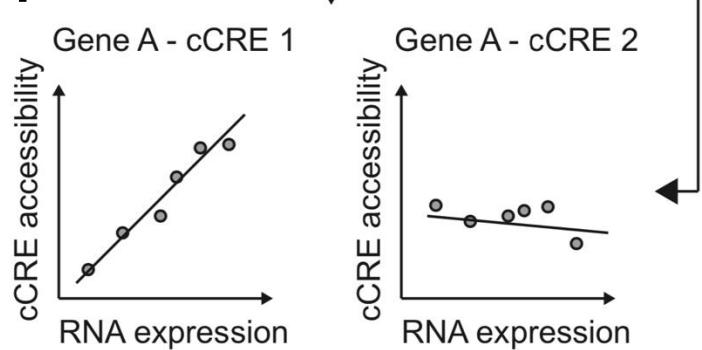


# Linking distal cCREs to target genes

## Step 1



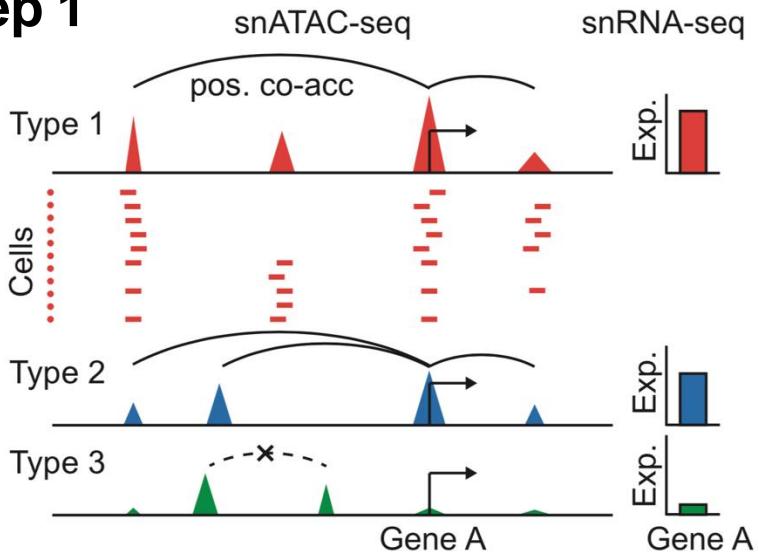
## Step 2



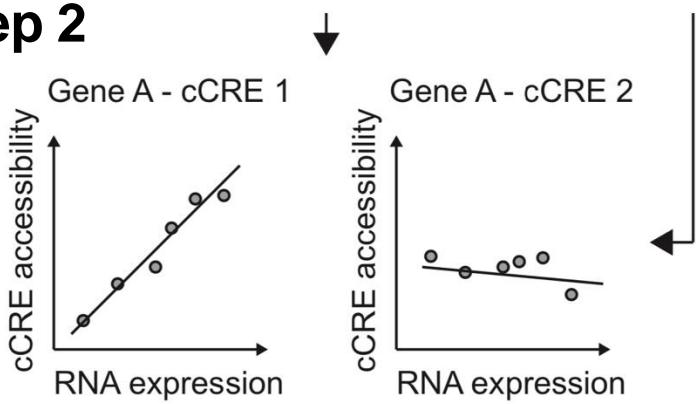
114,877 putative enhancers  
13,094 potential target genes

# Linking distal cCREs to target genes

## Step 1

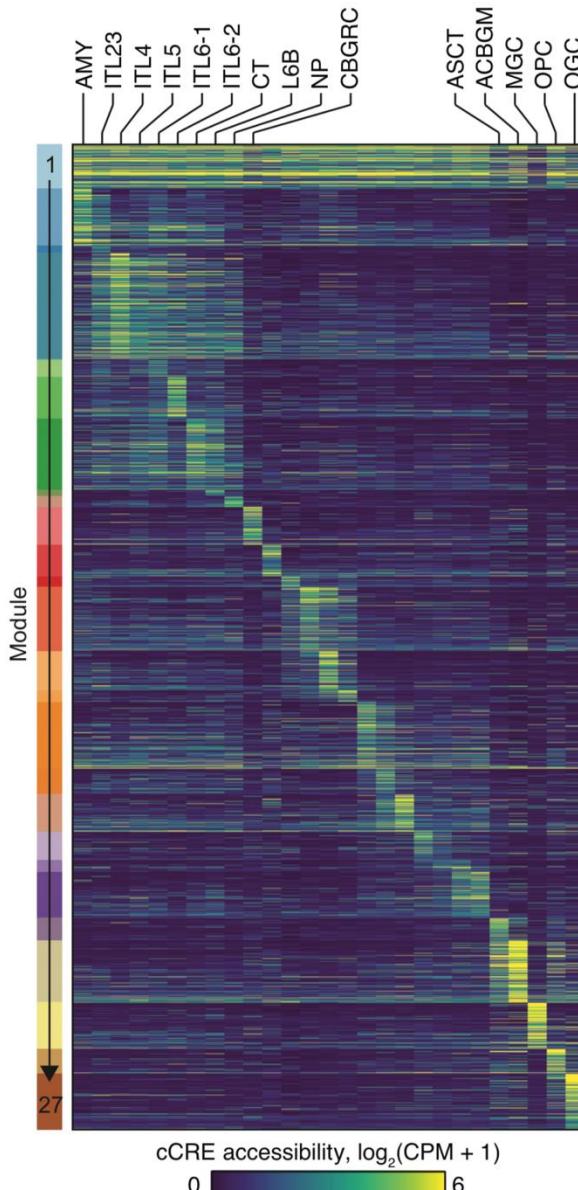


## Step 2

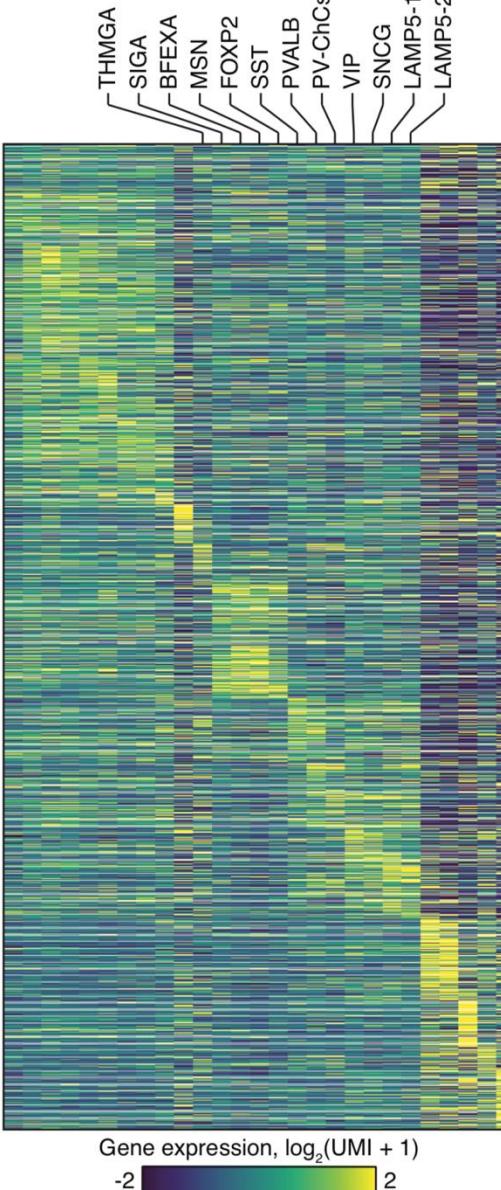


**114,877 putative enhancers**  
**13,094 potential target genes**

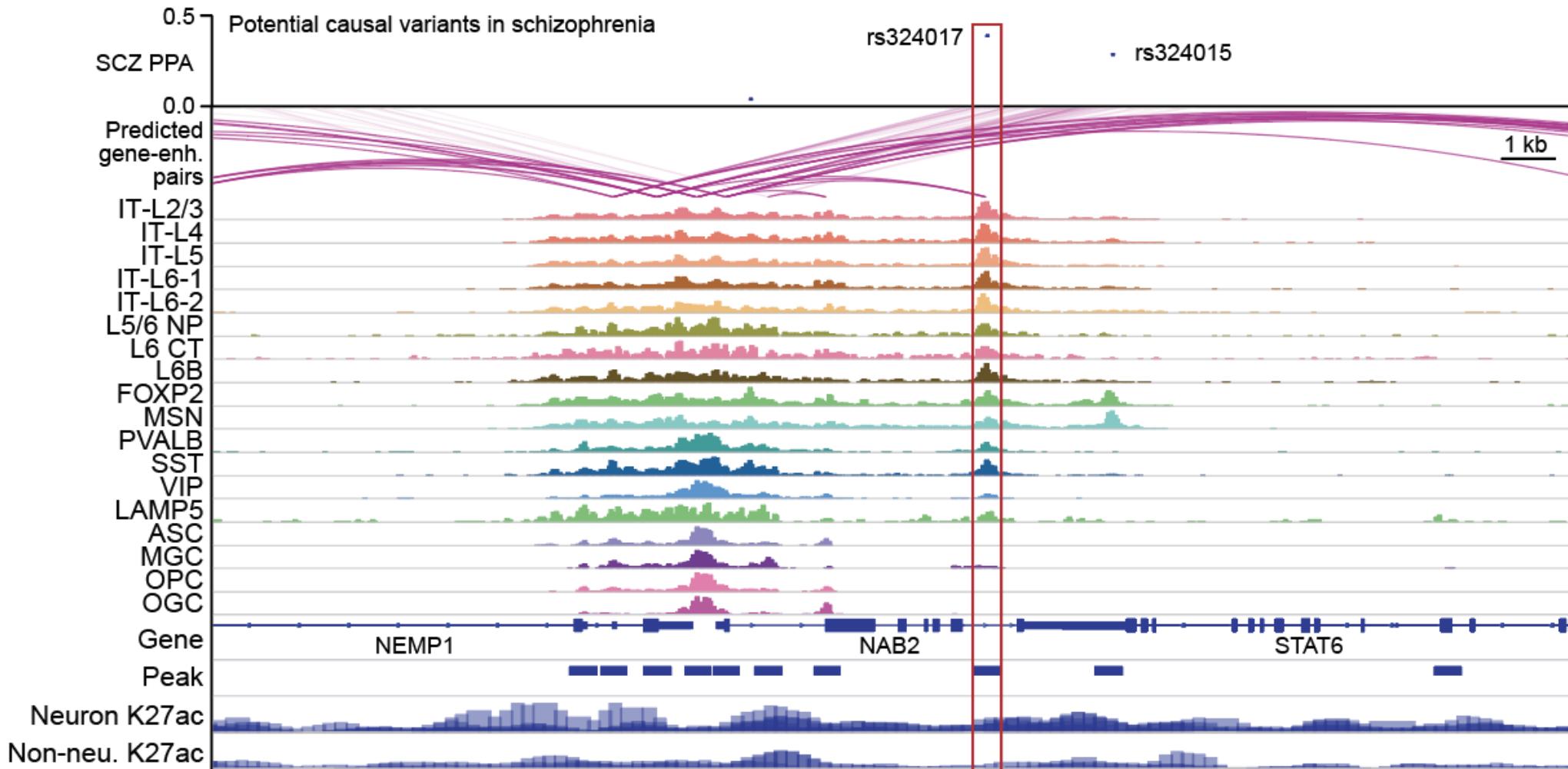
## Enhancer accessibility



## Expression of putative target genes



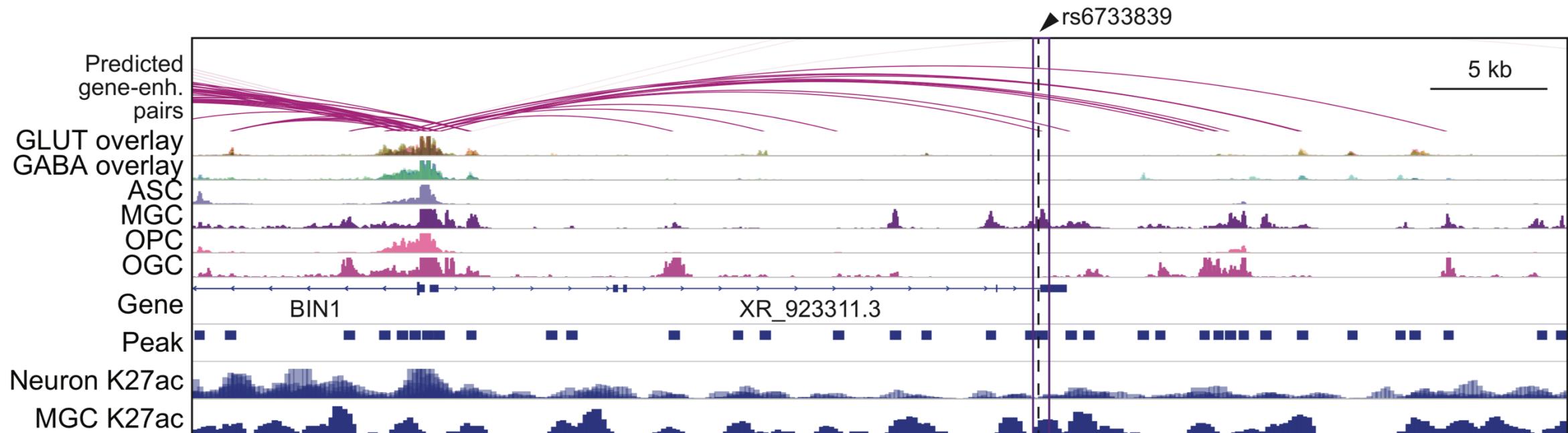
# Representative risk variants in schizophrenia



**NAB2:** Immediate early genes anchor a biological pathway of proteins required for memory formation, long-term depression and risk for schizophrenia

# Representative risk variants in Alzheimer's disease

## A microglia-specific enhancer containing AD risk variant drives microglia-specific expression of the BIN1 gene

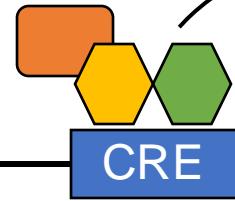


**BIN1**: This gene encodes several isoforms of a nucleocytoplasmic adaptor protein.  
This gene is also a risk gene in Alzheimer's disease.

# A paradigm for understanding the noncoding disease risk variants, and the key barriers

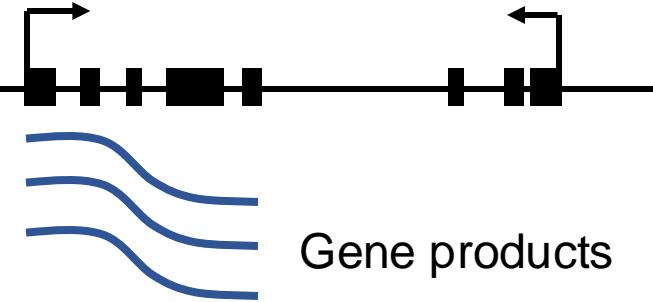
How do risk variants affect transcription factor binding, and which ones?

Transcription factors,  
Chromatin regulators



Target gene

What target genes are controlled by the *cis* regulatory elements (CREs)?



What sequences act as *cis* regulatory elements (CREs), where and when do they act?

It has been estimated that >50% of risk variants act in this way (Nasser J, ... Engreitz, Nature, 2021)

# Predicting chromatin accessibility from DNA sequences

Article | Open Access | Published: 04 October 2021

## Effective gene expression prediction from sequence by integrating long-range interactions

Žiga Avsec  , Vikram Agarwal, Daniel Visentin, Joseph R. Ledsam, Agnieszka Grabska-Barwinska, Kyle R. Taylor, Yannis Assael, John Jumper, Pushmeet Kohli  & David R. Kelley 

*Nature Methods* 18, 1196–1203 (2021) | Cite this article

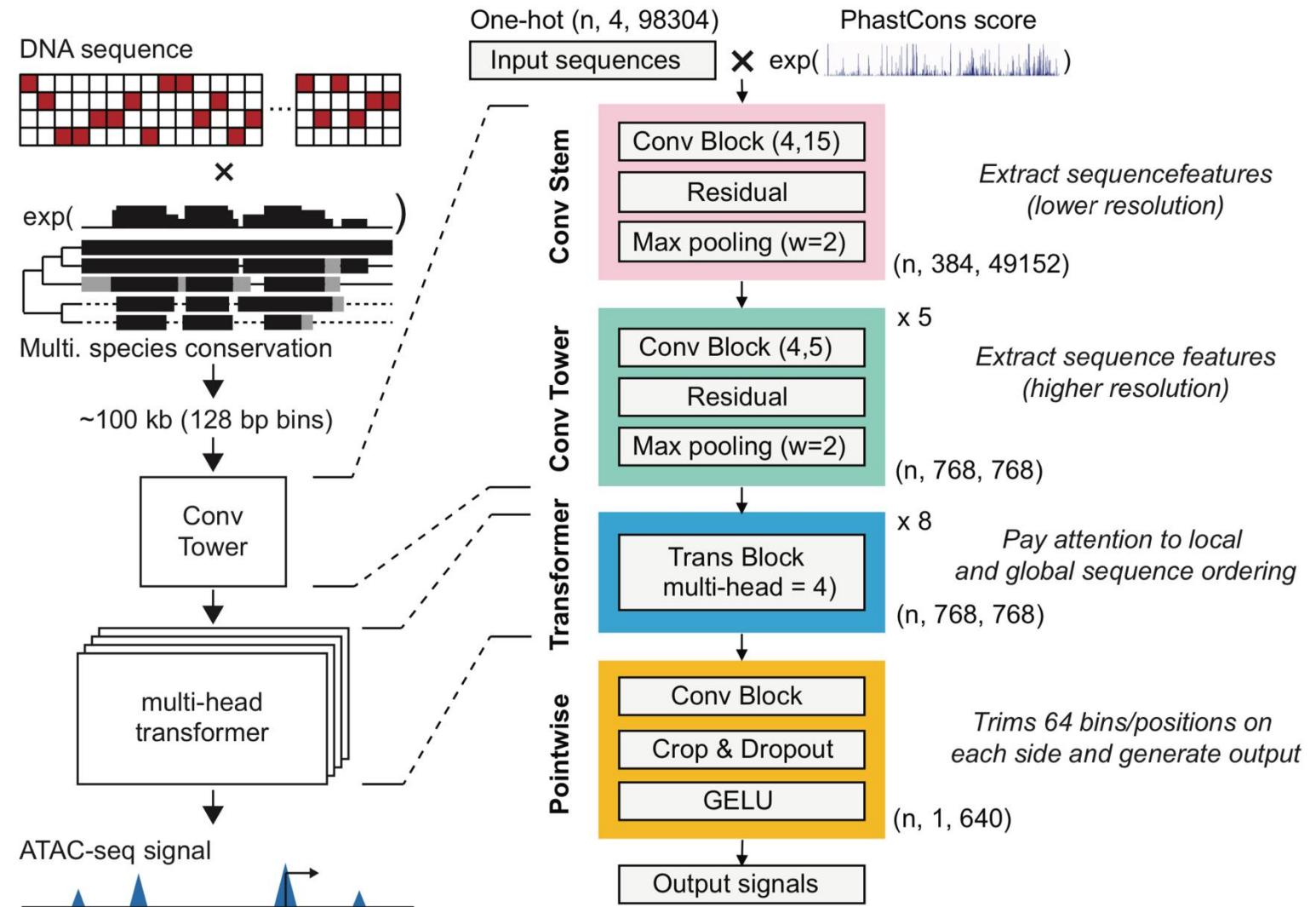
88k Accesses | 52 Citations | 396 Altmetric | Metrics



Zihan Wang  
CSE, UCSD



Dr. Jingbo Shang  
CSE, UCSD



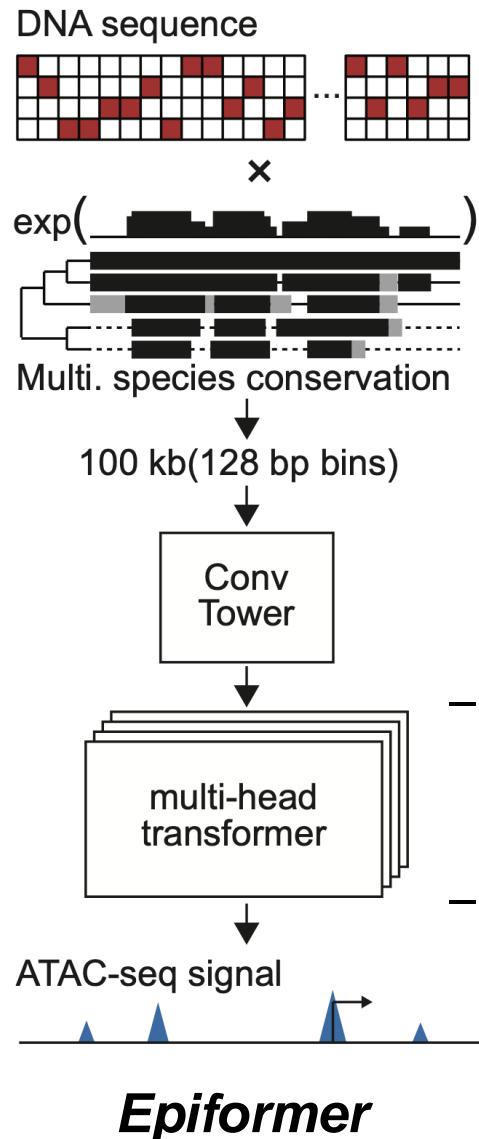
*Epiformer*

# Predicting chromatin accessibility from DNA sequences

DNA sequence



Potential function  
(genomic signals)



Natural language processing

The agreement on the European Economic Area was signed in August 1992.

Translator

L'accord sur la zone économique européenne a été signé en août 1992.

# Predicting chromatin accessibility from DNA sequences

Astrocytes  
(ASC)

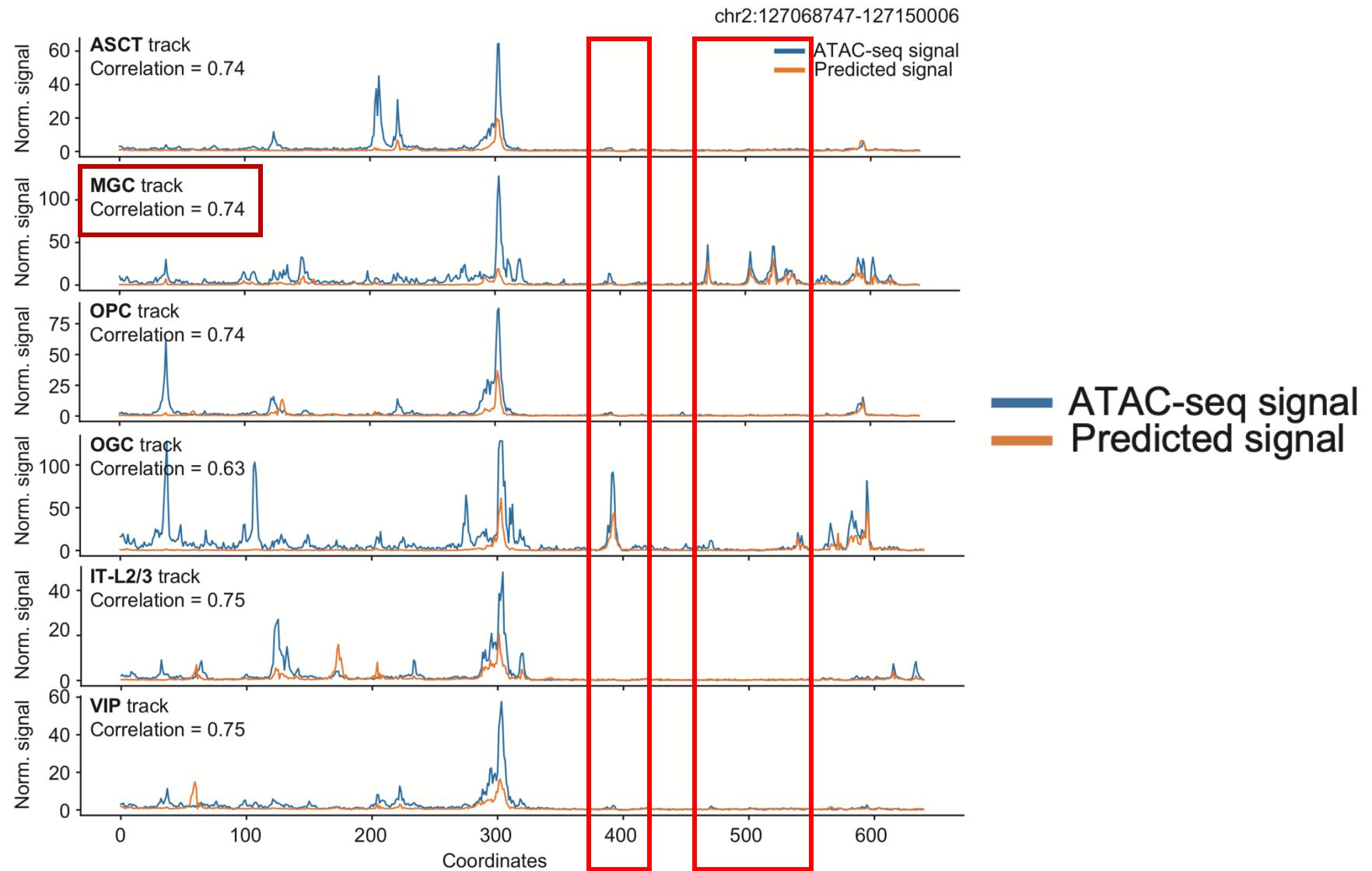
**Microglia (MGC)**

Oligodendrocytes  
precursor cell (OPC)

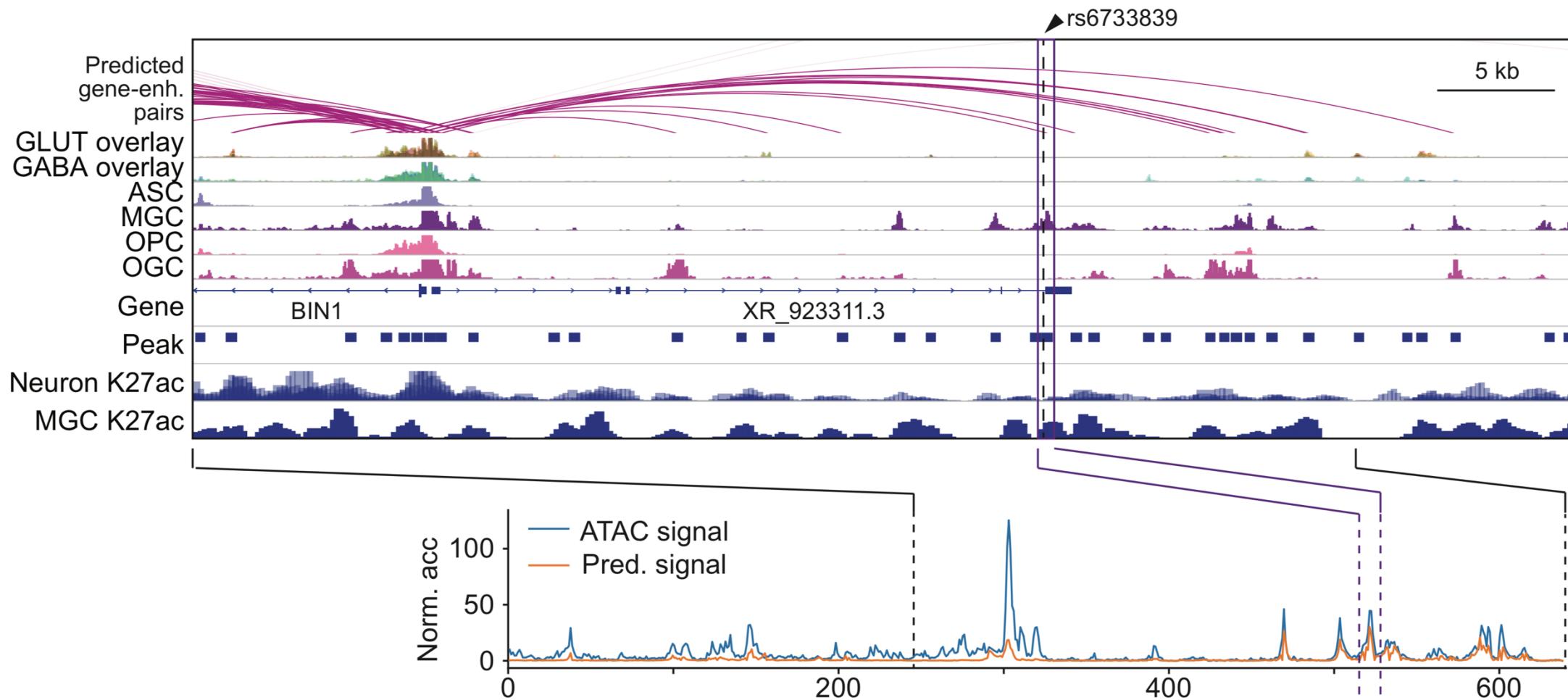
Oligodendrocytes  
(OGC)

Excitatory neurons  
at cortical layer 2/3

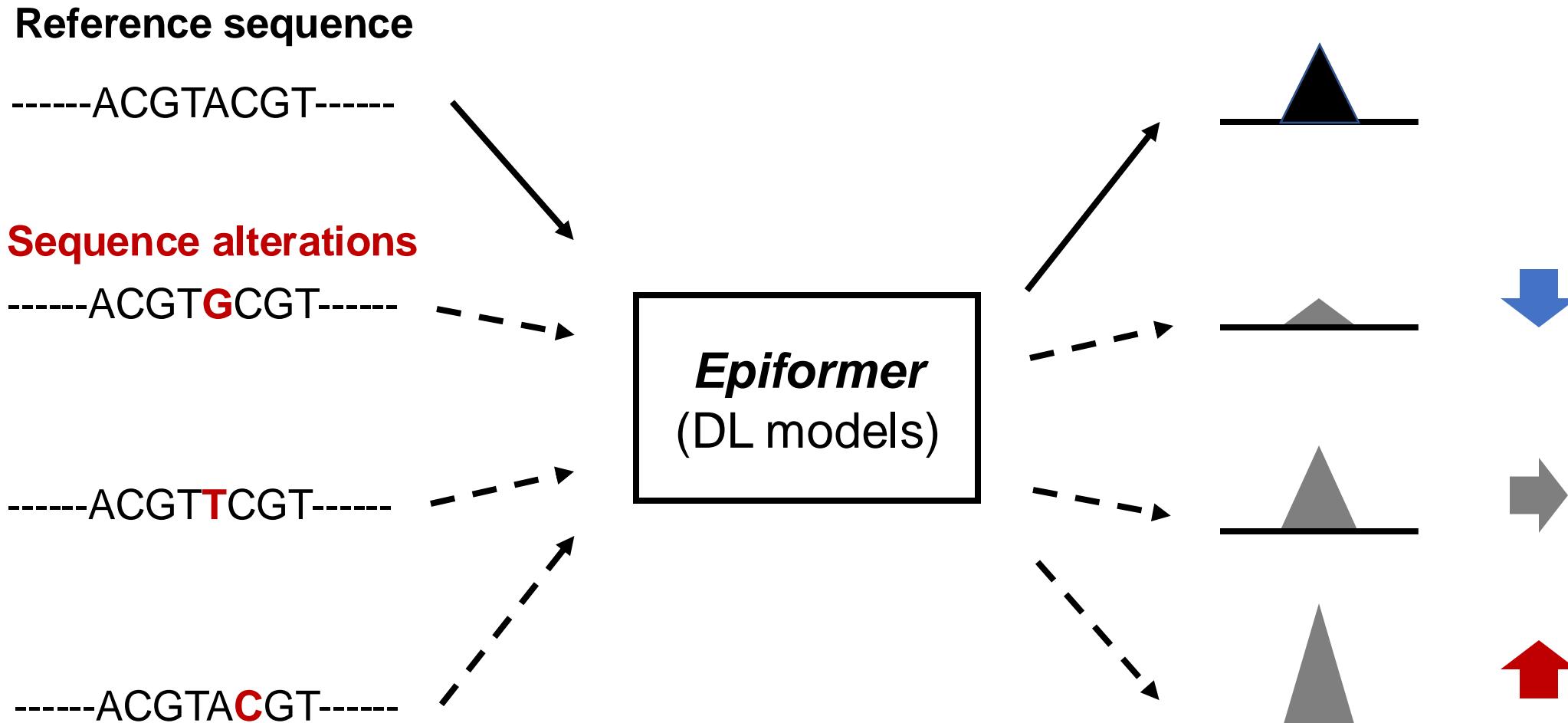
VIP+ Inhibitory  
neurons



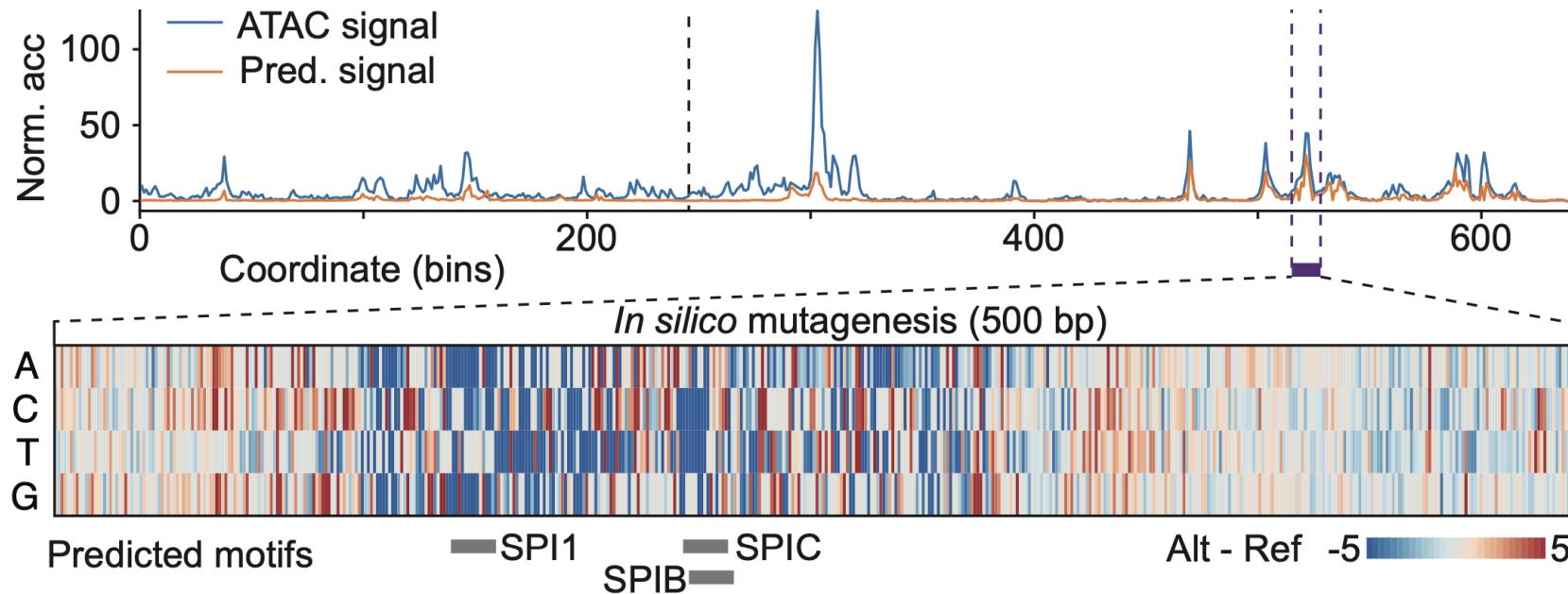
# Predicting chromatin accessibility from DNA sequences



# *In silico* mutagenesis accelerates interpretation

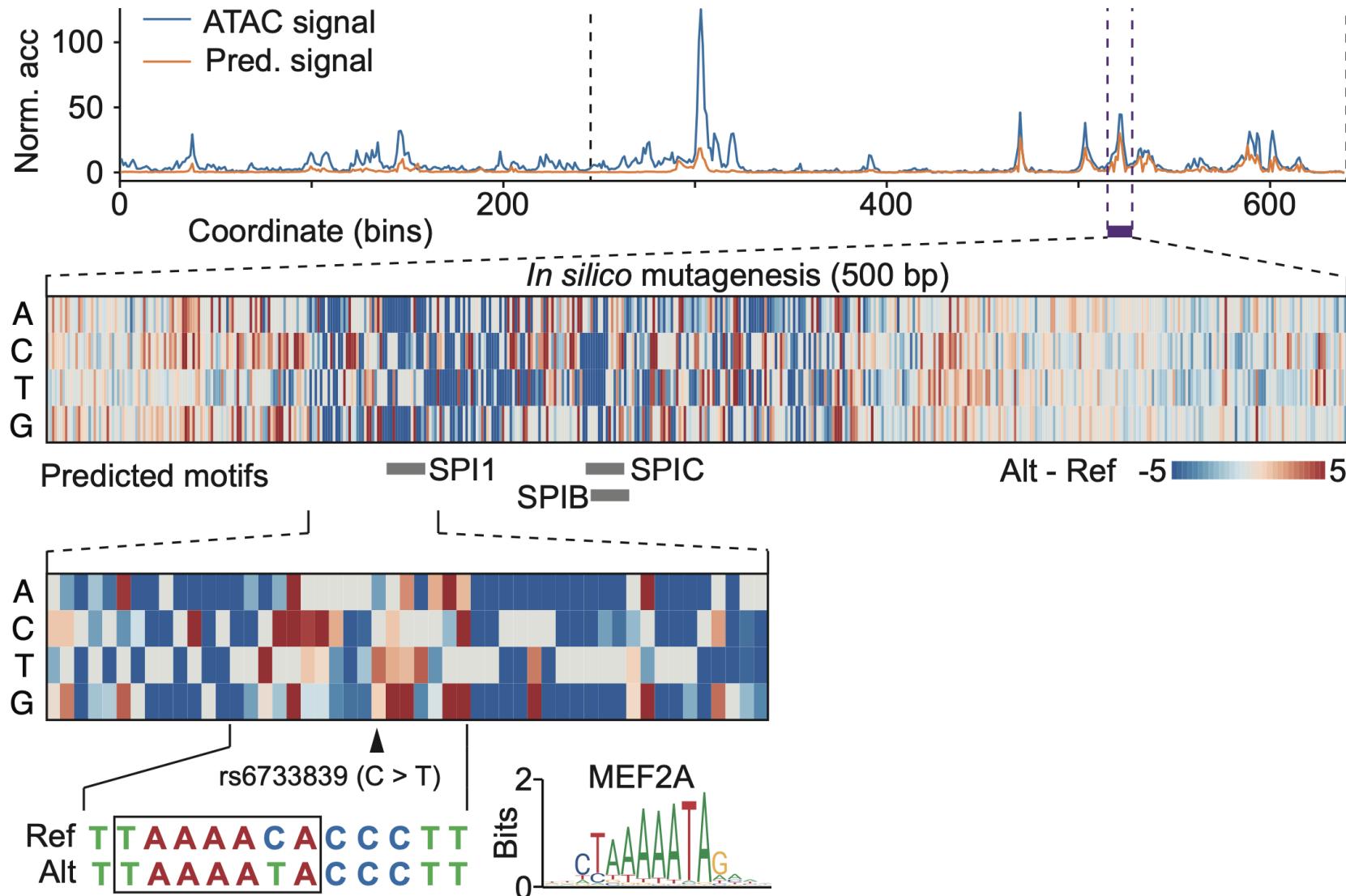


# Interpret risk variants through in silico mutagenesis

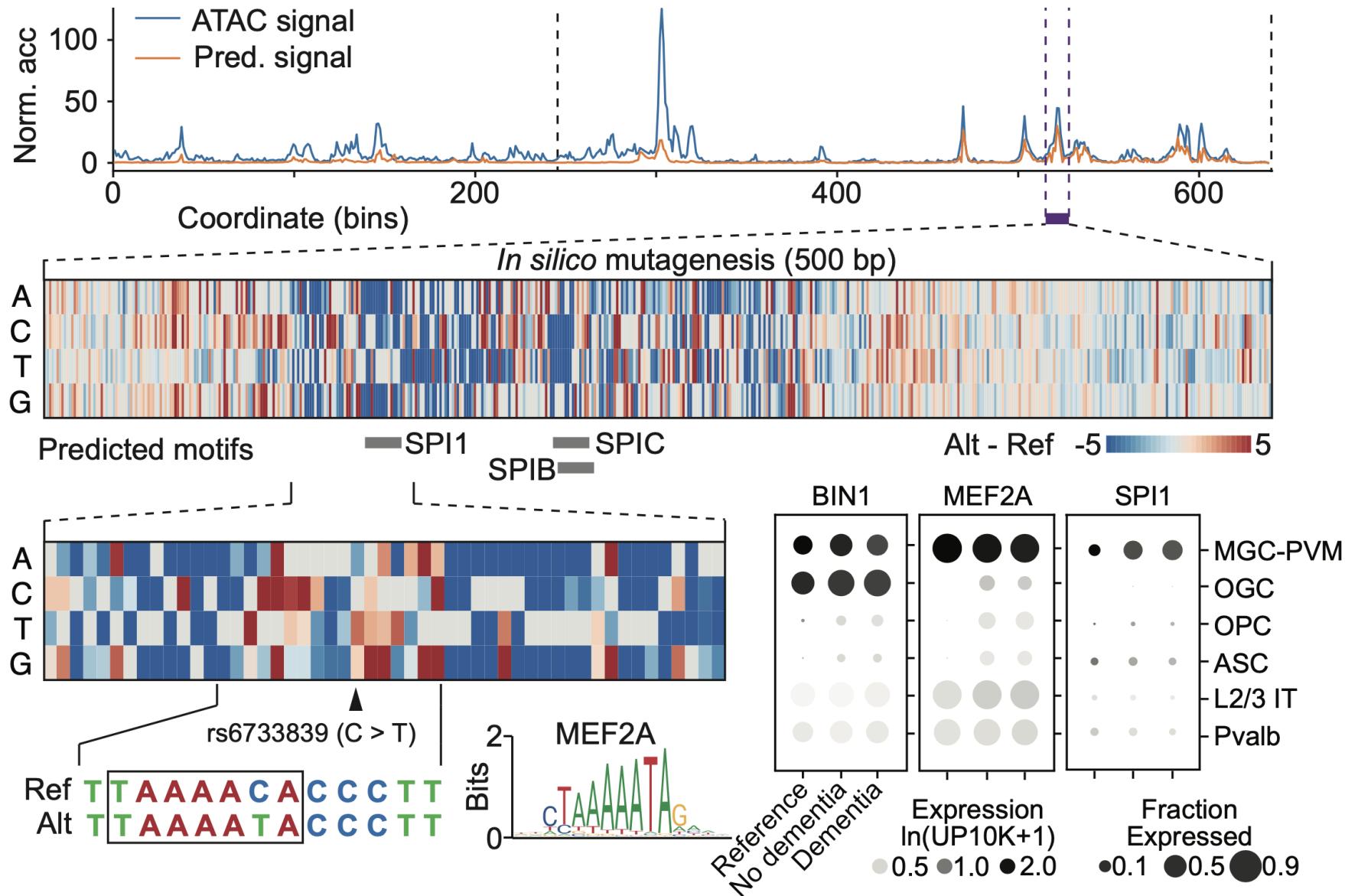


The transcription factor **PU.1 (SPI1)** regulates the normal ***development of microglial cells and their function*** in the brain (Smith et al., 2013).

# Interpret risk variants through in silico mutagenesis



# Interpret risk variants through in silico mutagenesis



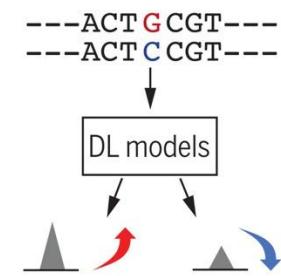
Seattle Alzheimer's Disease  
Brain Cell Atlas (SEA-AD)

# Summary

## Cell clustering



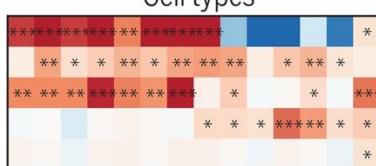
## Interpret risk variants



## Associations with neuropsychiatric disorders

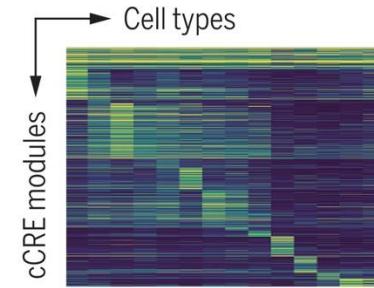
### LD score regression

19 GWAS traits and disorders

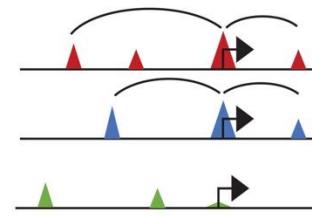


## Characterization of 544k cCREs

Novel cCREs

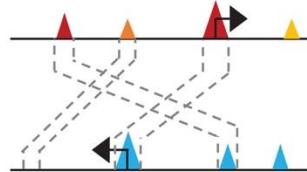


## Predict putative target genes

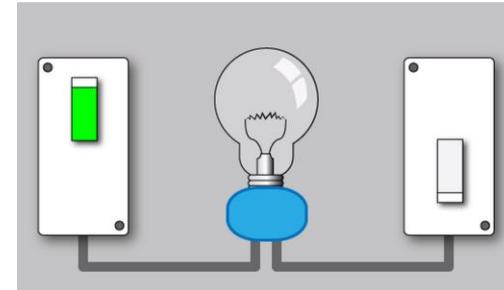


## Cross-species conservation

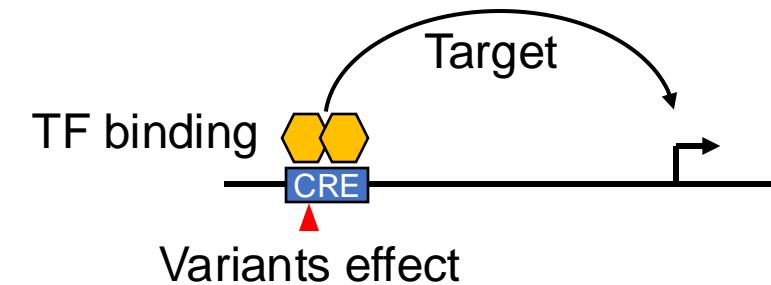
Human genome



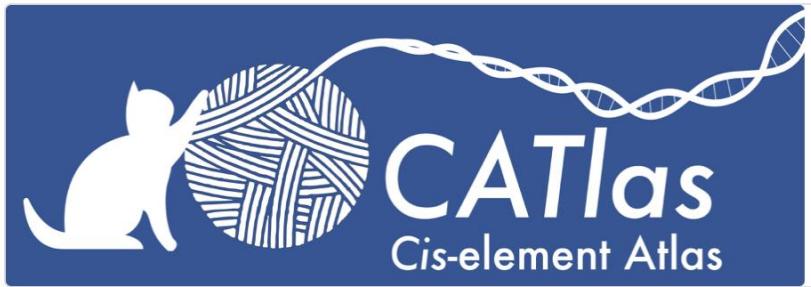
~ 1/3 of cCREs show conservation and accessibility in mouse



Light is “regulated” by switches.



# Cis-element atlas

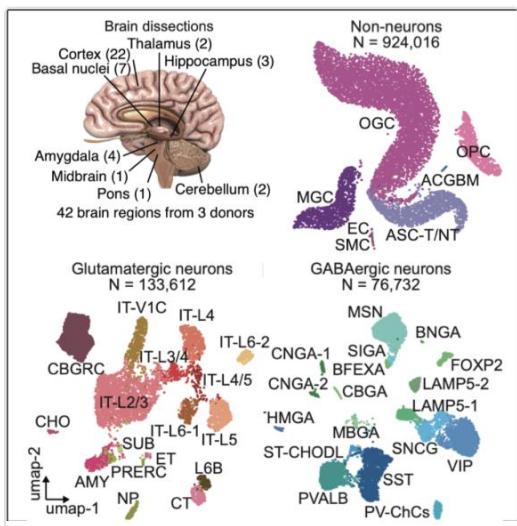


## Human Brain

Sample: 42 brain regions

Method: snATAC-seq + multi-modal

Nuclei count: > 1.1 million



## Cell Clusters

Explore cell clusters in different datasets

Last updated Nov,2022

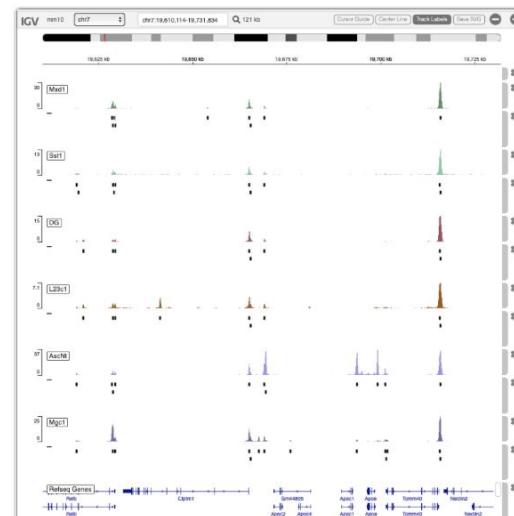
Explore

Class	L3 Cluster	Description
● GABAergic	PVGA1	MGE-derived neurogliaform cells, Pvalb
	Average TSSE:	20.89
	Average LogUMI:	3.63
	More Info	
● GABAergic	SSTGA5	MGE-derived neurogliaform cells, Sst
● GABAergic	LAMGA2	CGE-derived neurogliaform cells, Lamp5
● GABAergic	PVGA6	MGE-derived neurogliaform cells, Pvalb
● GABAergic	VIPGA3	CGE-derived neurogliaform cells, Vip
● GABAergic	LAMGA4	CGE-derived neurogliaform cells, Lamp5
● GABAergic	SSTGA1	MGE-derived neurogliaform cells, Sst
● GABAergic	VIPGA1	Cholinergic neurons, Cortex
● GABAergic	SSTGA4	MGE-derived neurogliaform cells, Sst
● GABAergic	SSTGA3	MGE-derived neurogliaform cells, Sst

## Cell Types

Explore detailed information for every cell type

Explore



## Explore Tracks

Load signal tracks and compare between cell types/regions

Explore

- Explore cell cluster
- Cell type annotations
- Compare signal tracks



Data now available via our portal:  
<http://catlas.org>

> 55,599 visits

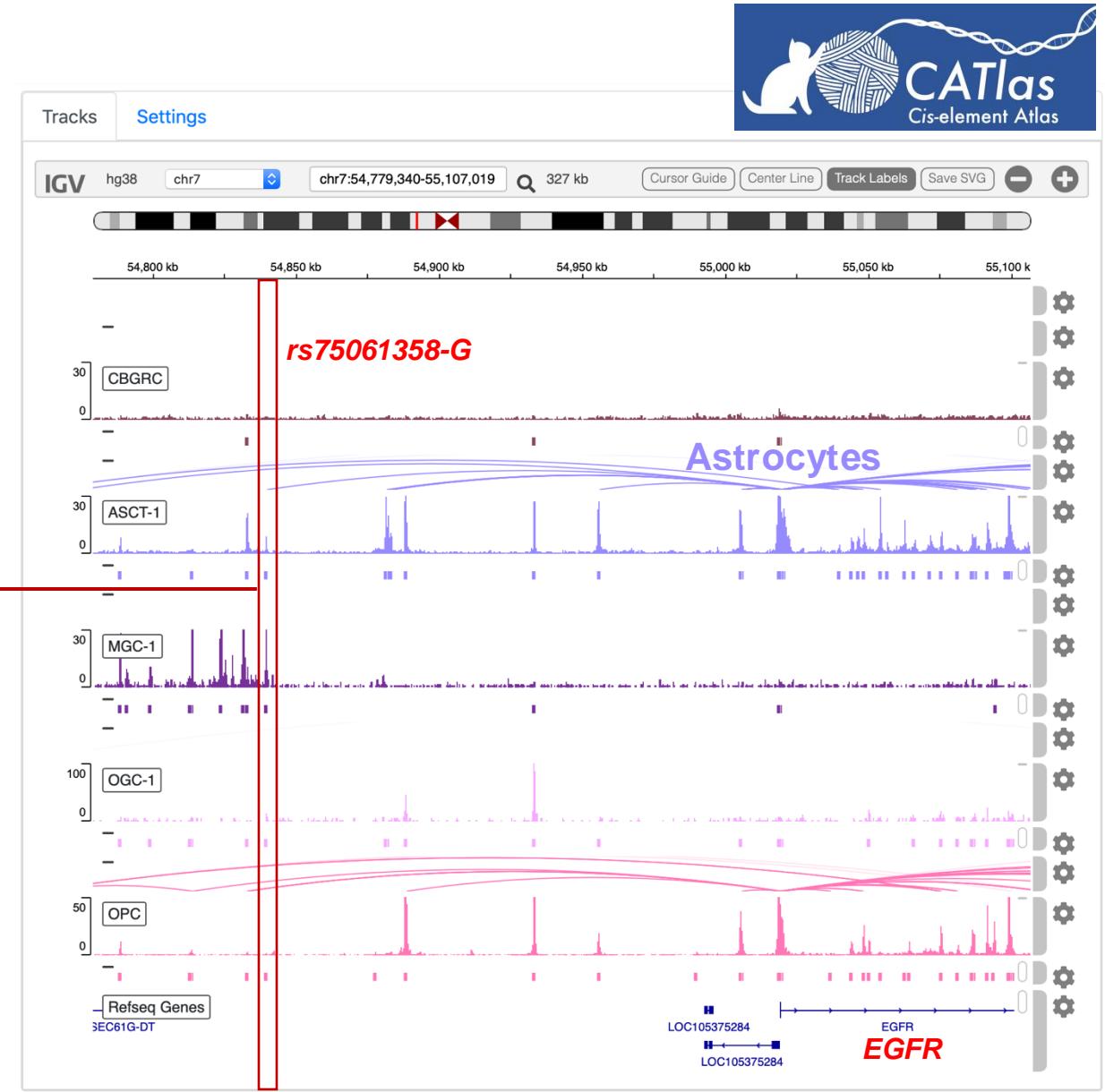
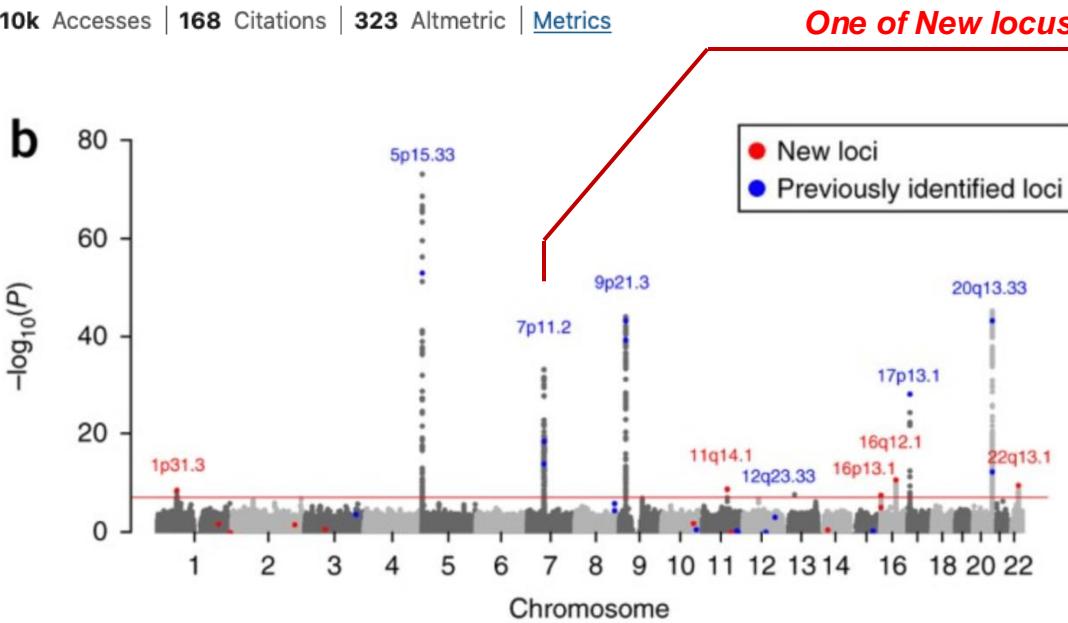
Published: 27 March 2017

# Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors

Beatrice S Melin , Jill S Barnholtz-Sloan, Margaret R Wrensch, Christoffer Johansen, Dora Il'yasova, Ben Kinnersley, Quinn T Ostrom, Karim Labreche, Yanwen Chen, Georgina Armstrong, Yanhong Liu, Jeanette E Eckel-Passow, Paul A Decker, Marianne Labussière, Ahmed Idbaih, Khe Hoang-Xuan, Anna-Luisa Di Stefano, Karima Mokhtari, Jean-Yves Delattre, Peter Broderick, Pilar Galan, Konstantinos Gousias, Johannes Schramm, Minouk J Schoemaker, GliomaScan Consortium, ... Melissa L Bondy   
+ Show authors

Nature Genetics 49, 789–794 (2017) | Cite this article

10k Accesses | 168 Citations | 323 Altmetric | Metrics



# CATlas (version 2)

The screenshot shows the homepage of CATlas version 2. At the top, there is a navigation bar with links for Home, Browse, Search, Modeling, Predictions, Upload, Help, Contact, and a 'SUBSCRIBE' button. The main content area includes:

- Overview of CATlas**: A section with a brief description of the Cis-element Atlas (CATlas) as a crowdsourcing resource for regulatory elements and functional annotations.
- Logo**: A logo featuring a black cat sitting next to a circular pattern representing DNA, with the text "CATlas Cis-element Atlas".
- Search, Browse, Modeling, Upload**: Buttons for navigating to these specific sections.
- Quick Search**: A search bar with examples like VIP, VIP+ GABAergic neurons and VIP\_1, VIP+ GABAergic neurons - type 1.
- Data Statistics**: A section displaying various metrics:
  - 5+ Species: Homo Sapiens, Mus musculus, Macaca mulatta...
  - 1700+ Cell types: Brain, Heart, Liver...
  - 11+ million Cells: Excitatory neurons, Hepatocyte, Cardiomyocyte...
  - 6+ Sequencing: snATAC-seq, paired-Tag...
- Number of cells in CATlas**: A pie chart showing the distribution of cells across different species and sequencing methods. The data is summarized in the following table:

Species/Method	#Cells
Homo Sapiens	6,655,229
Mus musculus	8,827,529
Macaca mulatta	128,954
Callithrix jacchus	668,492
snATAC-Seq	900,000
Paired-Tag	64,849
sci-ATAC-seq	10x multiome

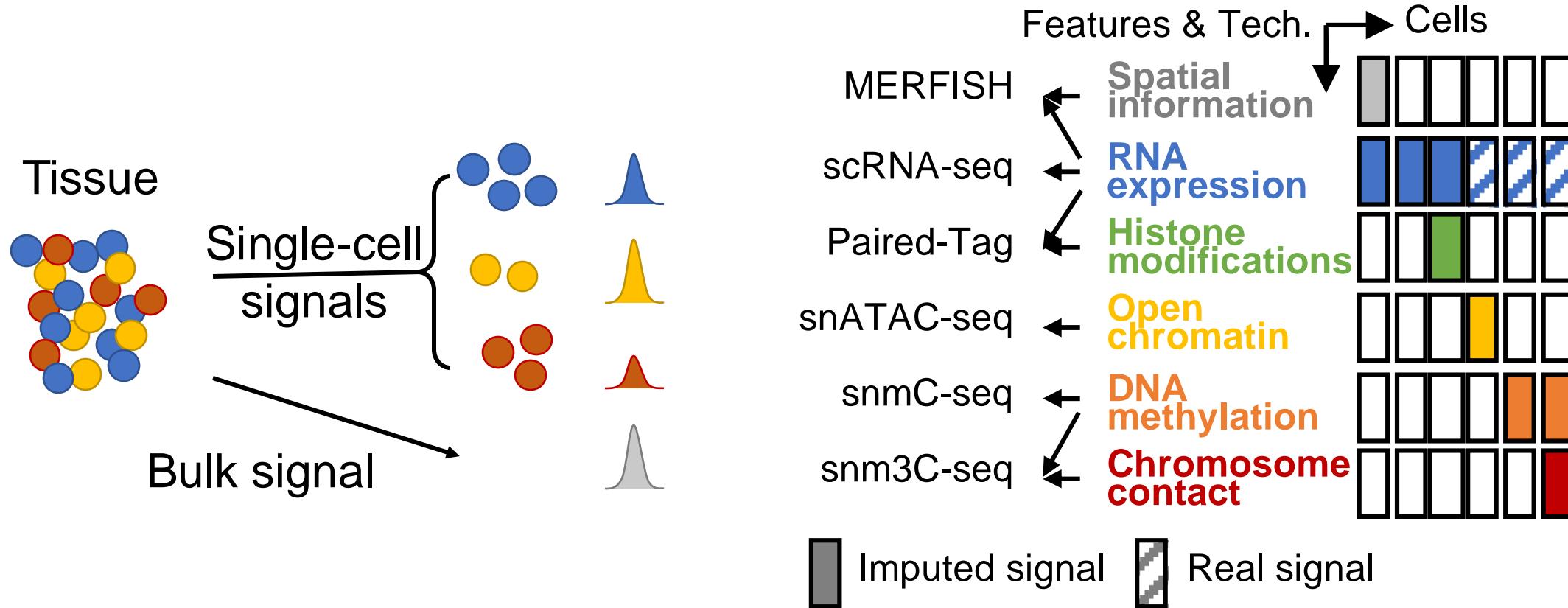
- Explore cell cluster
- Cell type annotations
- Compare signal tracks

**More data now available via our new portal:**



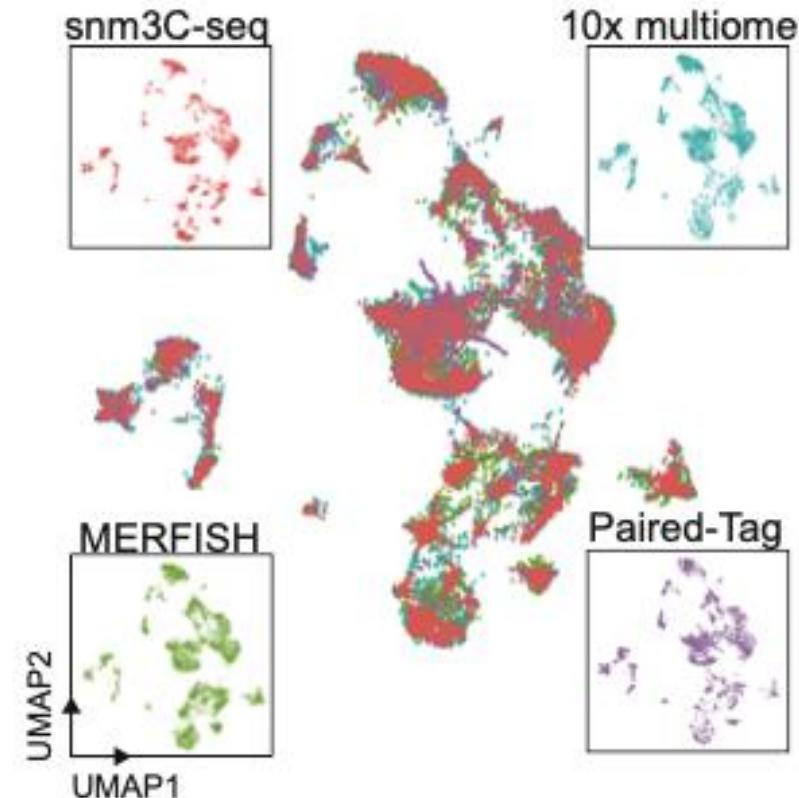
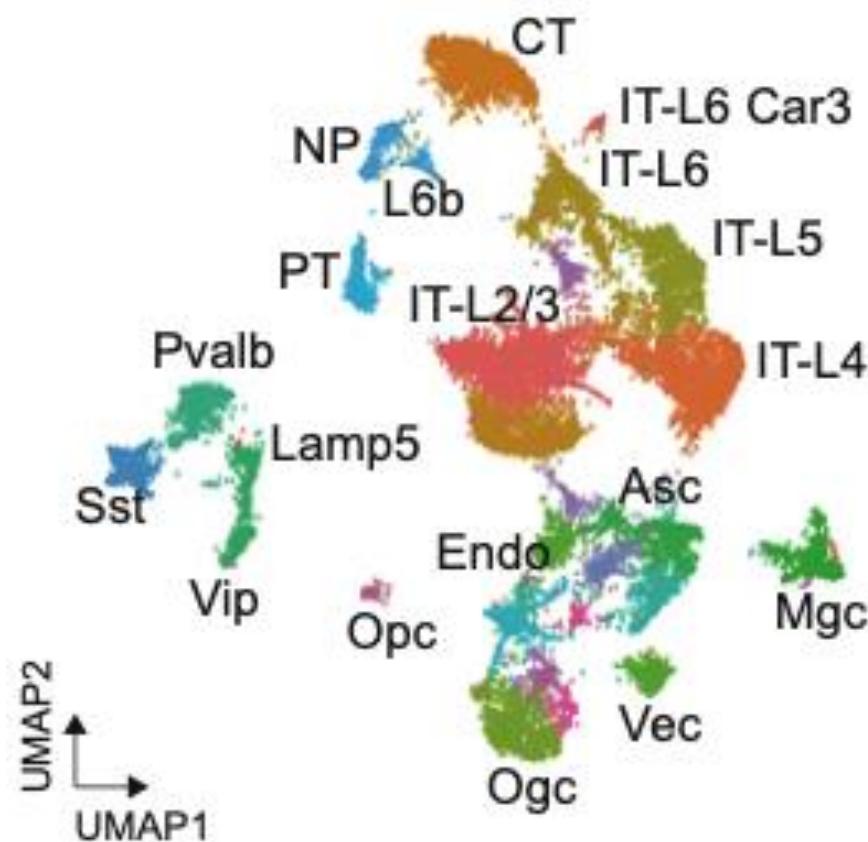
# Integration of single-cell multimodal omics

# Integration of single-cell multimodal omics

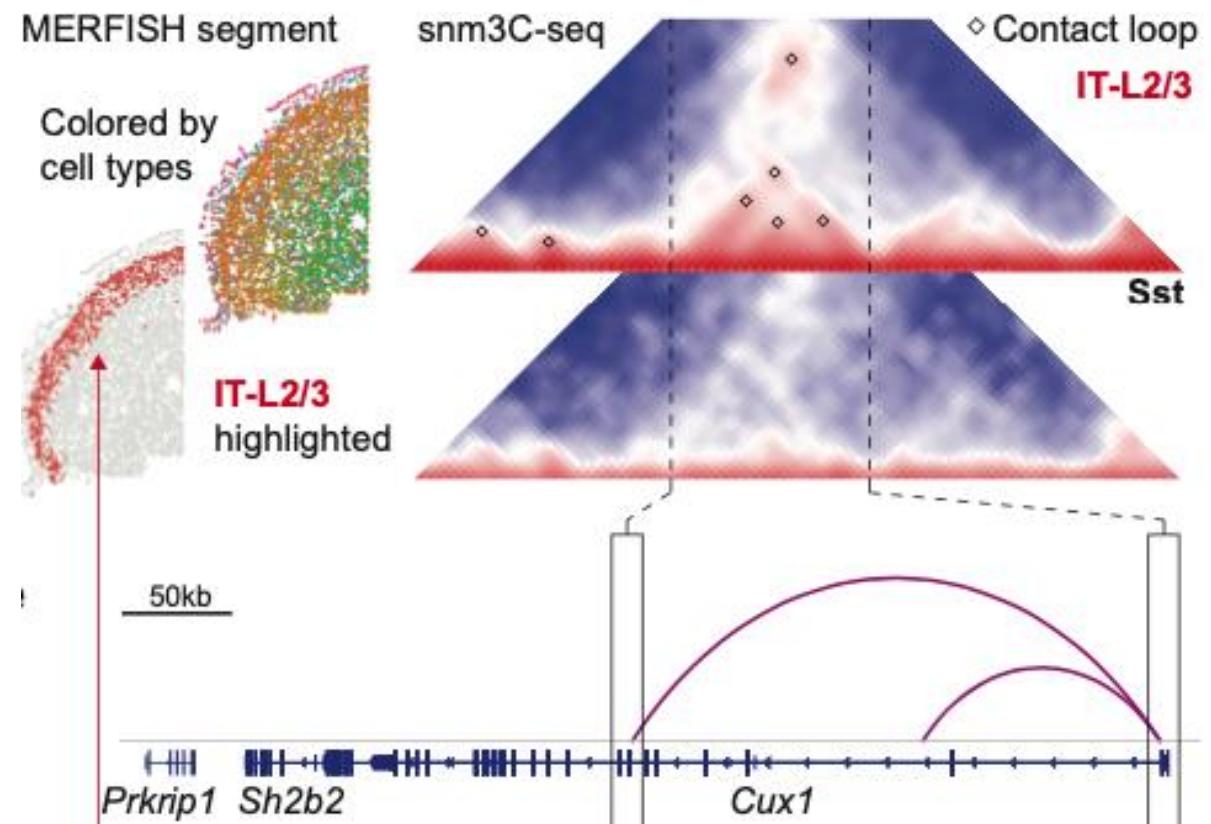
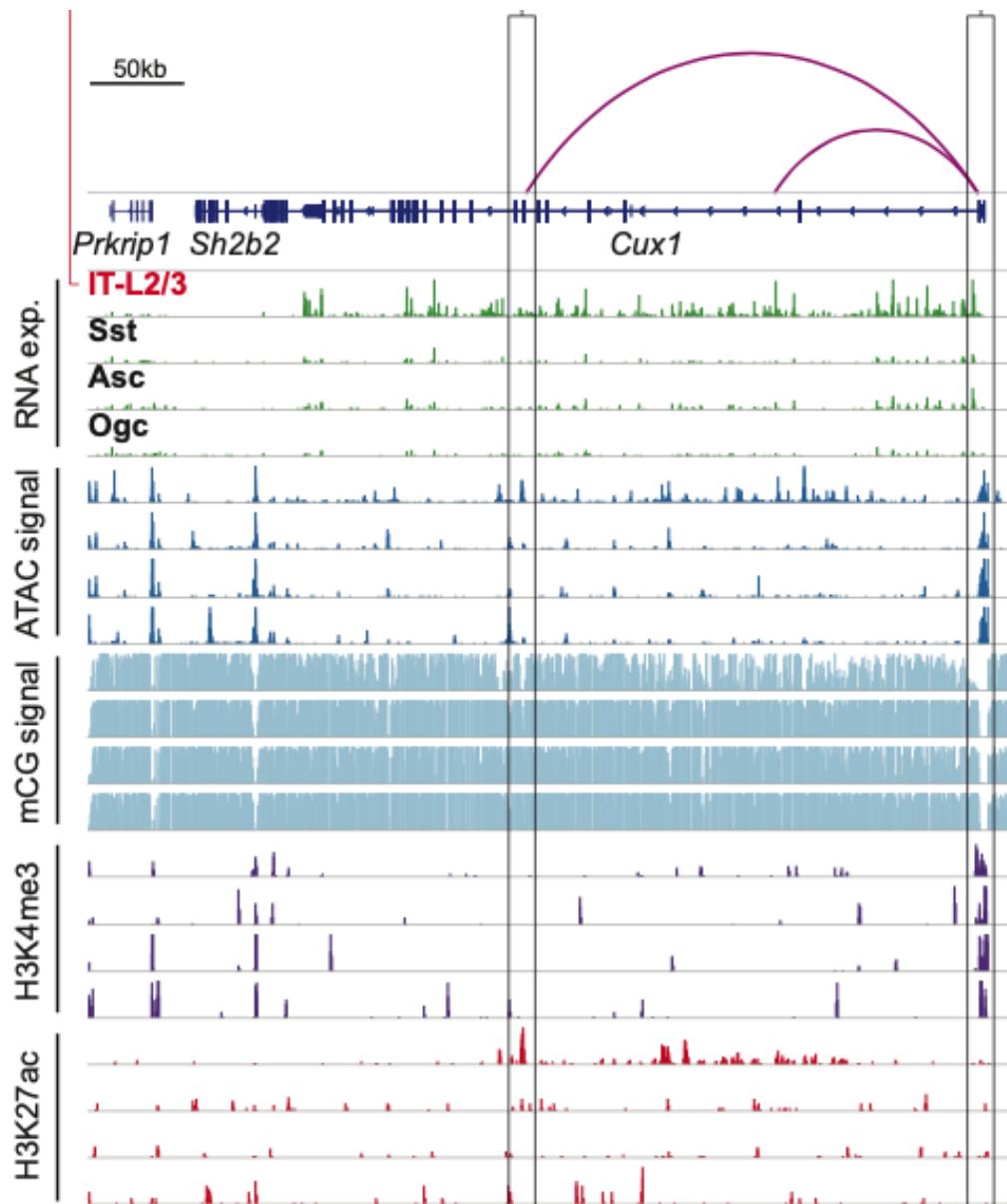


Technologies have been developed to map DNA methylation (**snmC-seq**), chromatin accessibility (**snATAC-seq**), histone modifications (**Paired-Tag**), 3D genome conformation (**snm3C-seq**), either alone or in combination with RNA-seq, in single cells.

# Integration of various single cell technologies

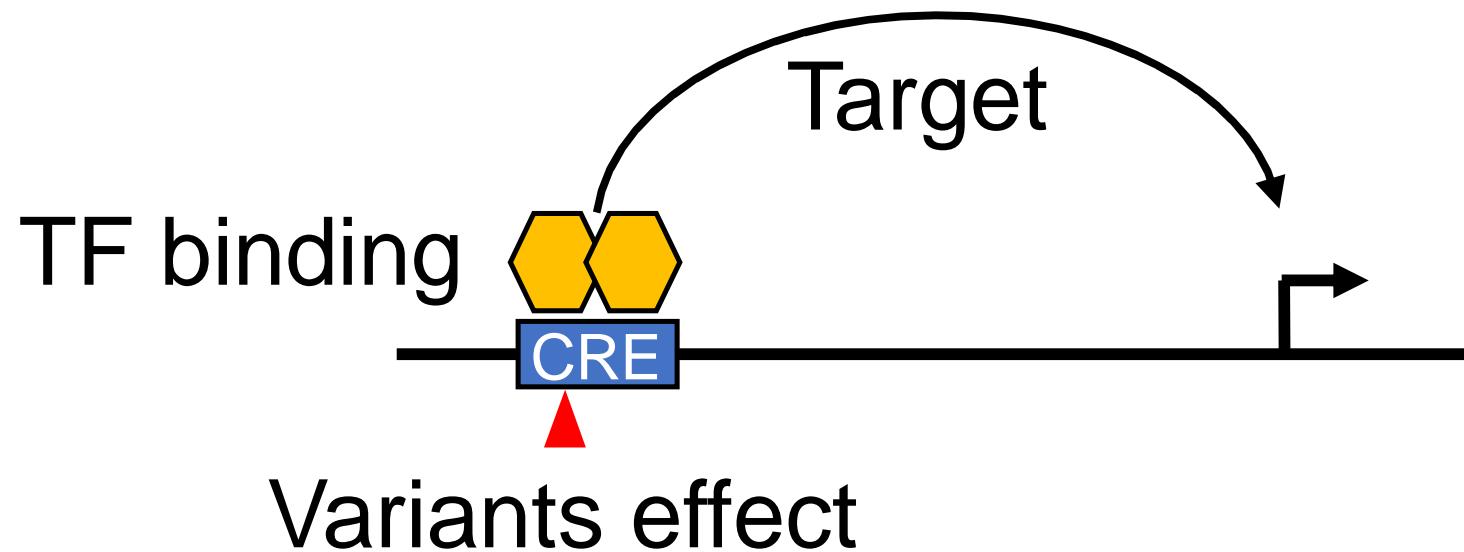


# Multi-view of genomic and spatial dataset

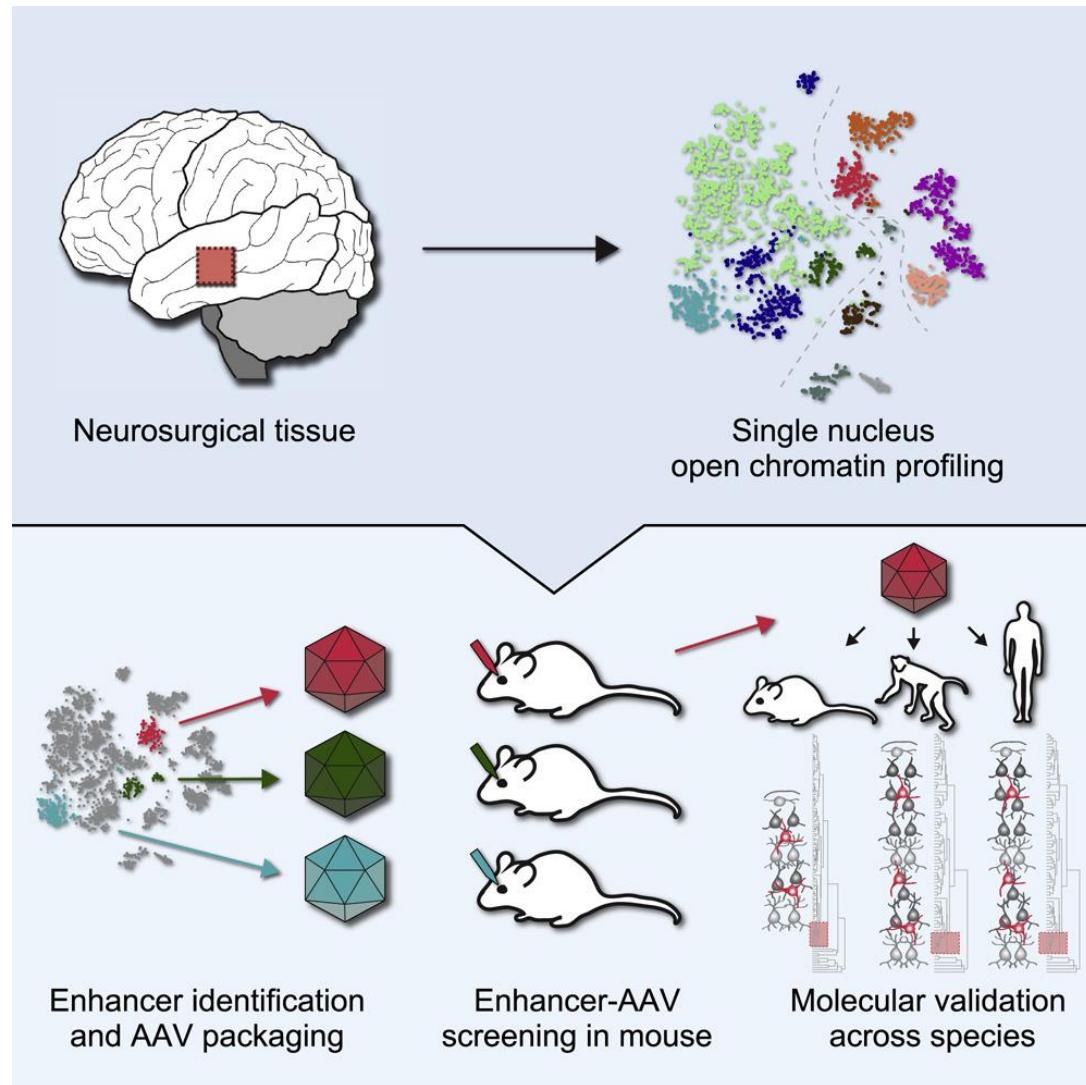
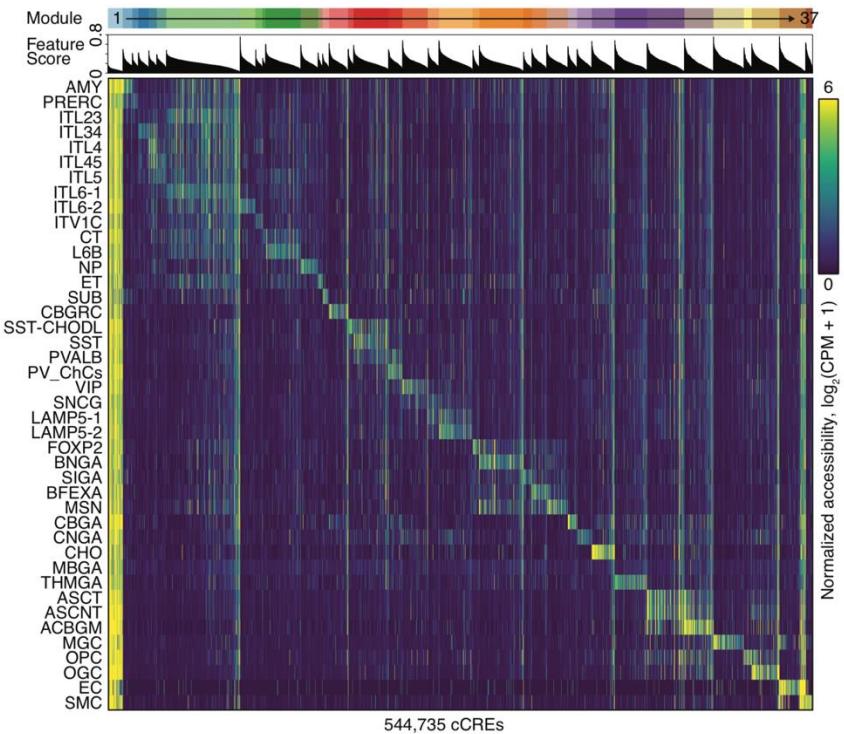


# Why do we care about CREs?

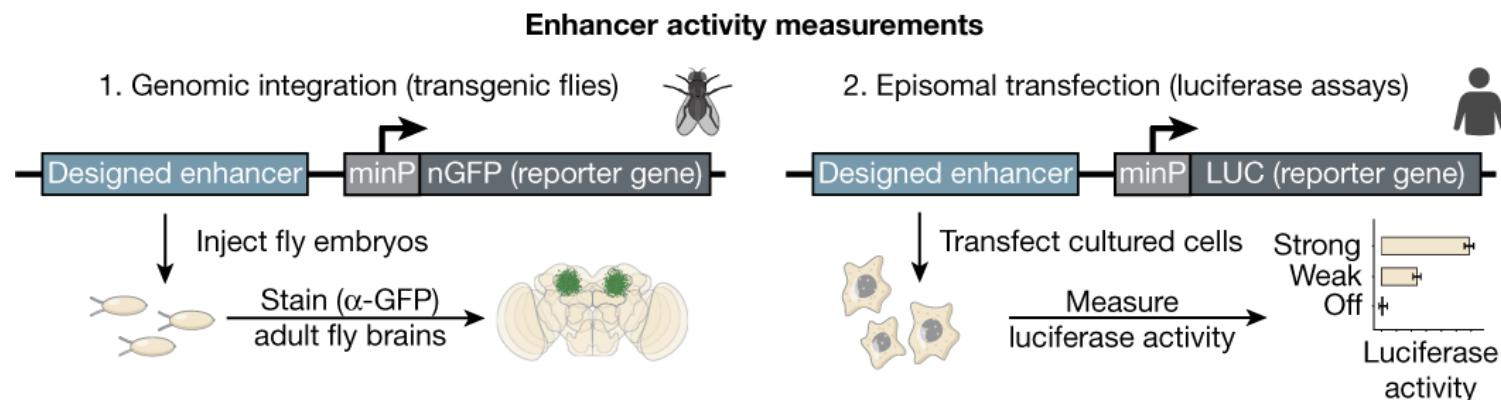
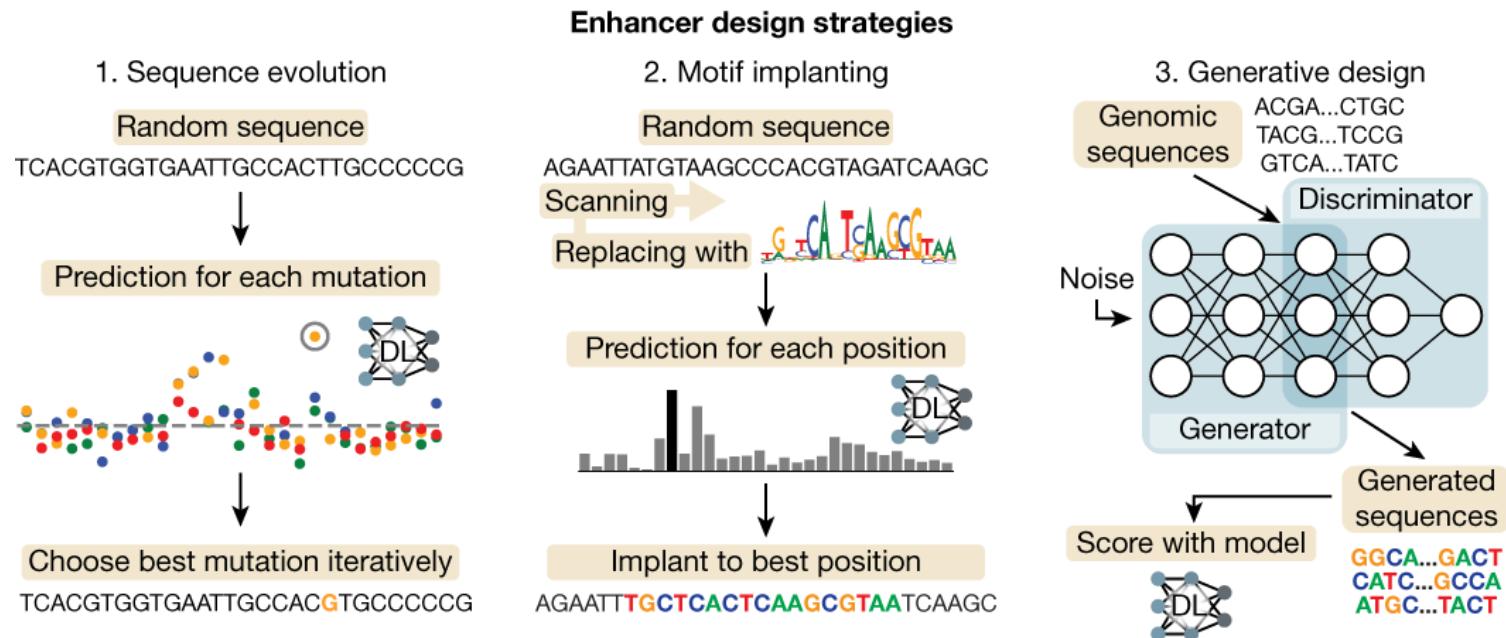
# Disease association & gene regulation



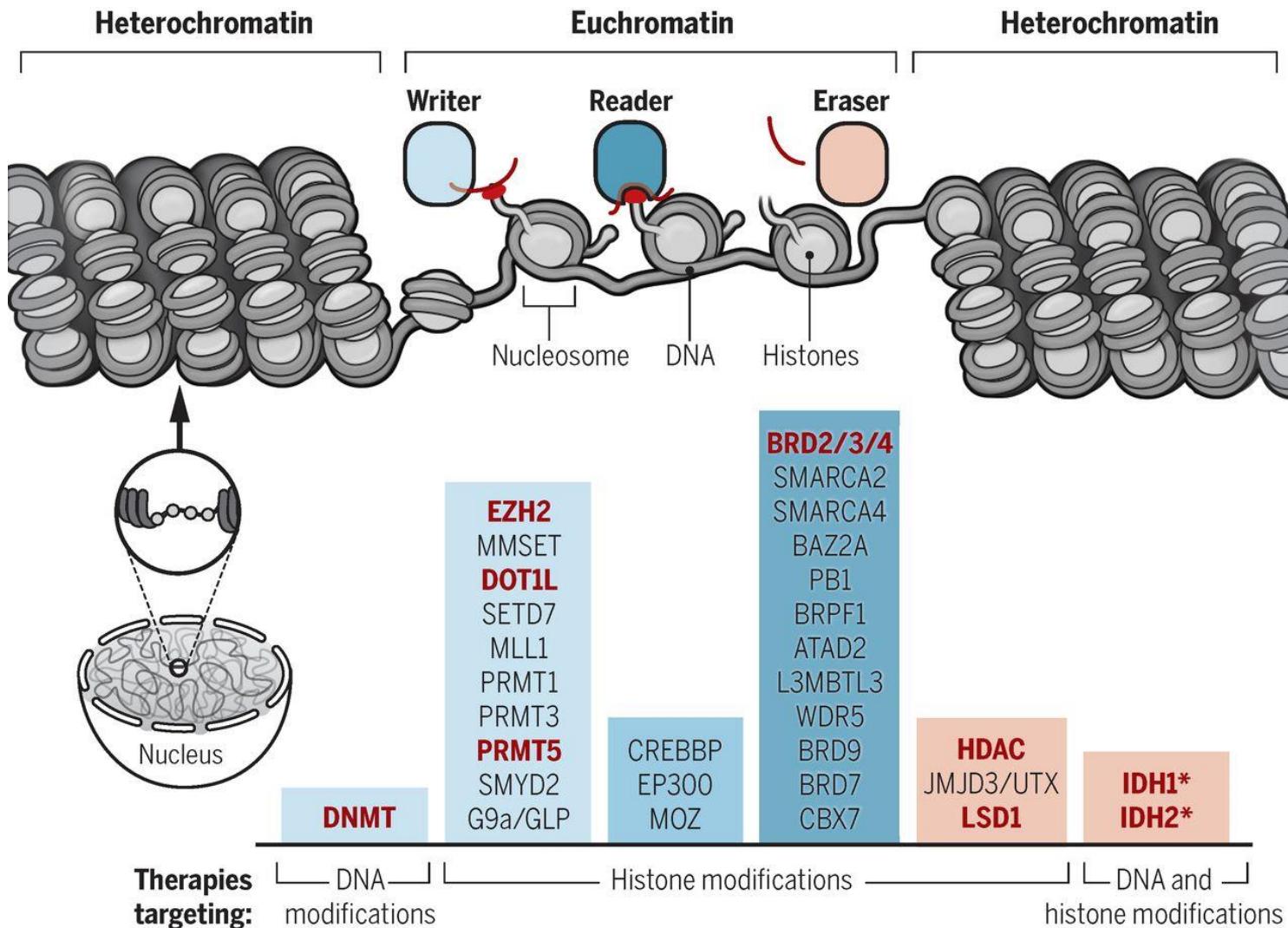
# "enhancer AAV (Adeno-associated virus)" labels a specific cell type



# Cell-type-directed design of synthetic enhancers

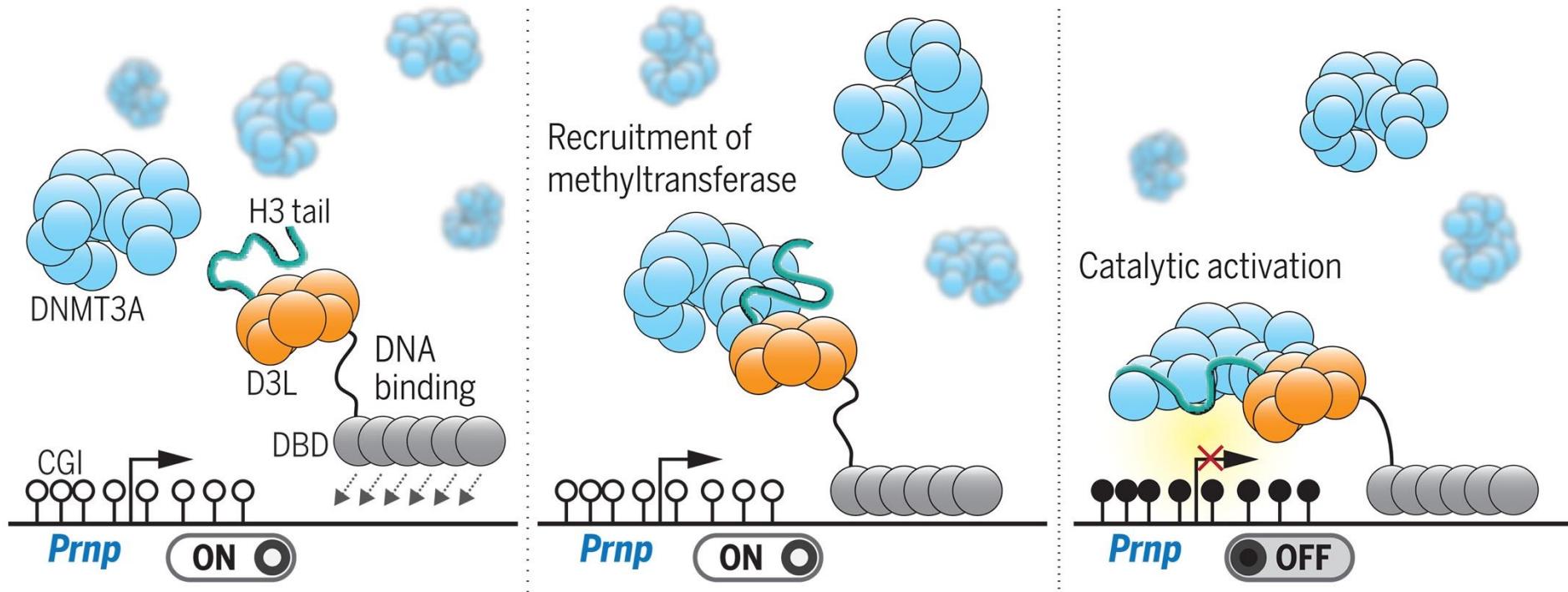


# Epigenetic therapy



Mark A. Dawson. Science. 2017

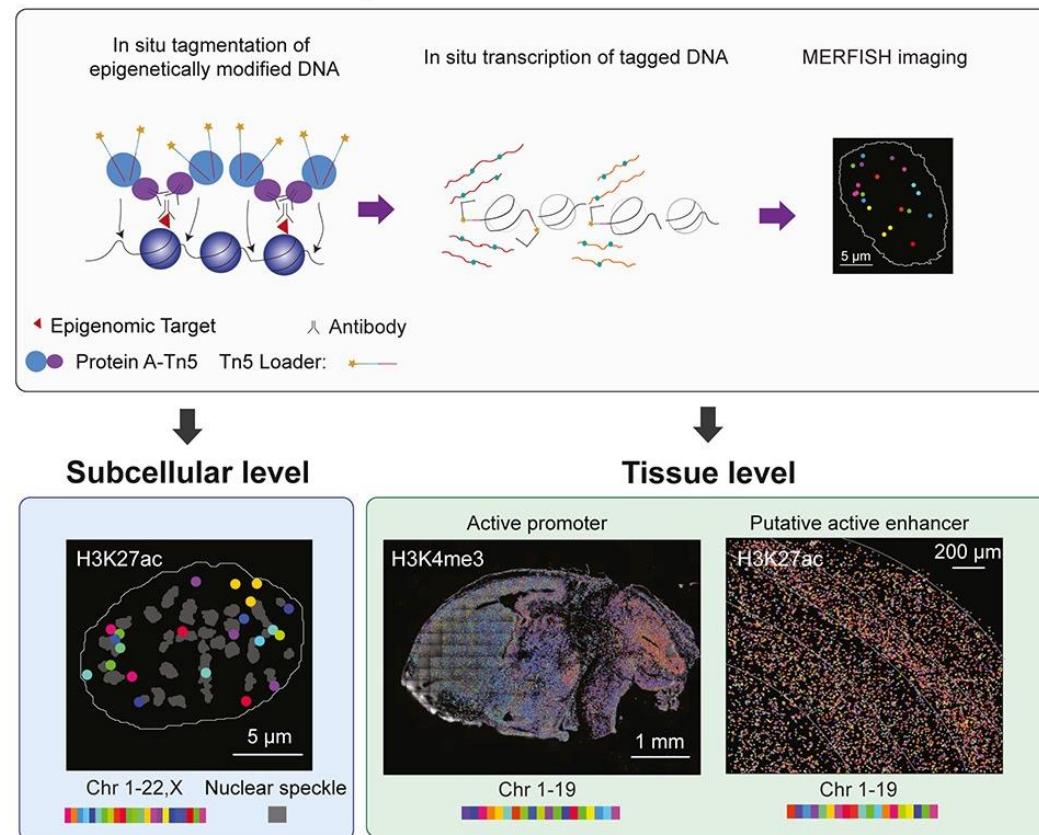
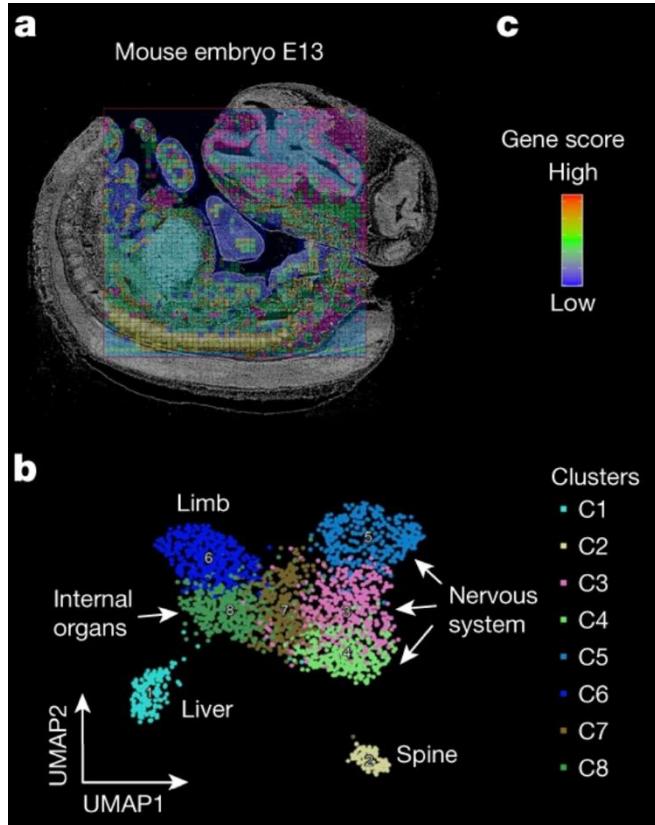
# Epigenetic therapy



Neumann, et al. Science. 2024

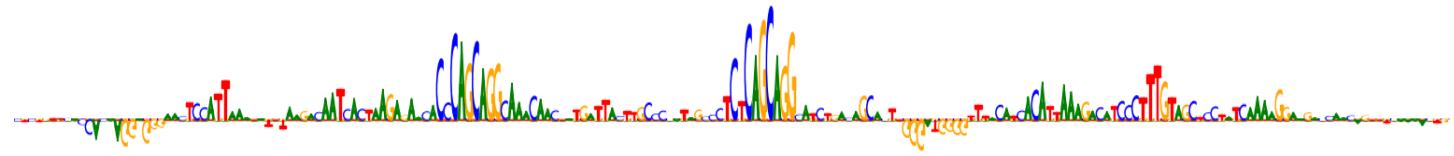
# The emerging of spatial epigenome

## Epigenomic MERFISH



Deng, et al. Nature. 2022

Lu, et al. Cell. 2022



**Thanks for listening**

