



UC San Diego

Comprehensive Analysis of Single Cell Epigenomic Data

Yang (Eric) Li

CSHL Sequencing Technologies and Bioinformatics Analysis

Nov 19, 2021

salk
Where cures begin.

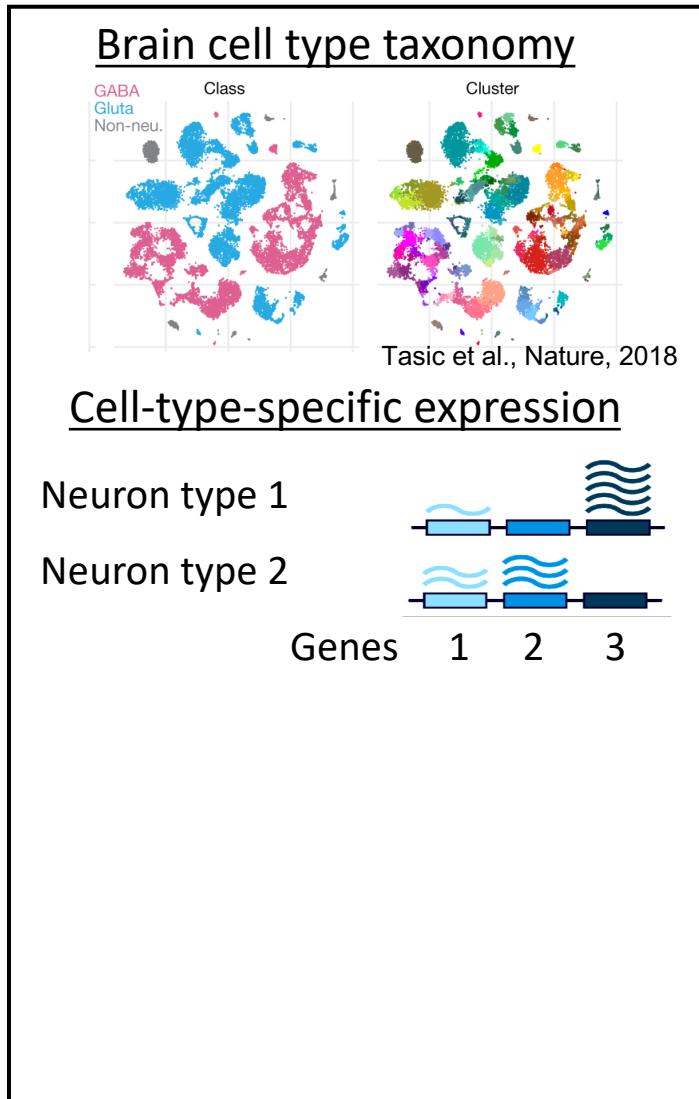

CENTER for
EPIGENOMICS
at UC San Diego


BRAIN
INITIATIVE

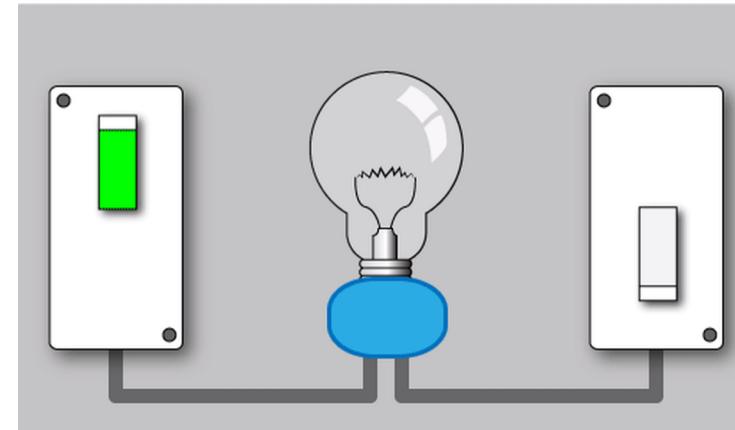
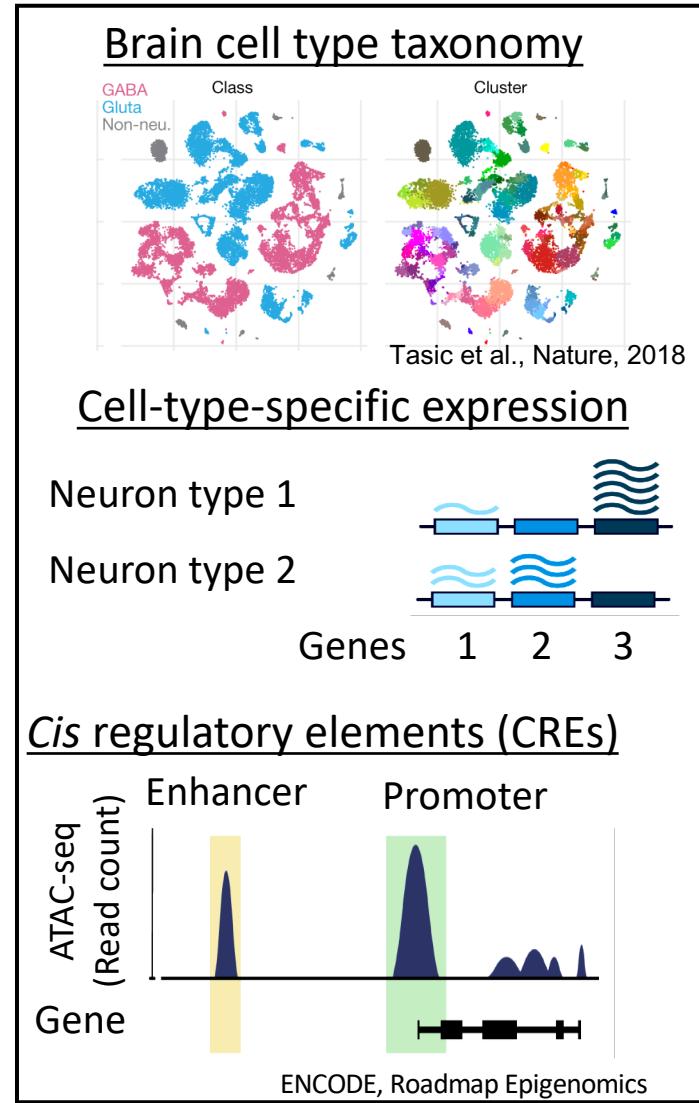
 BICCN

Bing Ren Lab
Department of Cellular and Molecular Medicine,
University of California, San Diego, School of Medicine
CMM-East 2071, San Diego, CA

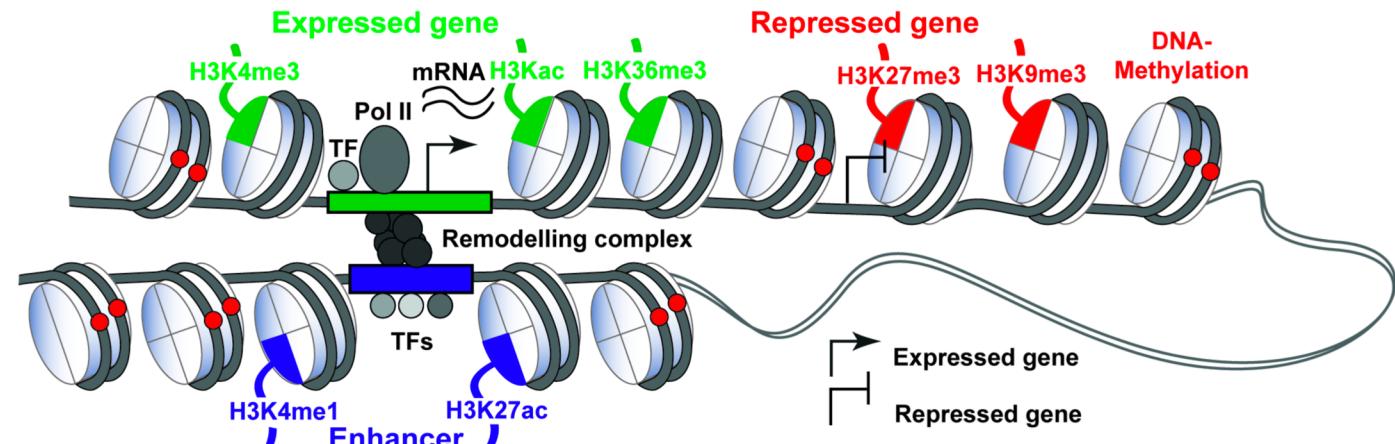
What gene regulatory programs are underlying the unique identity and function of cell types?



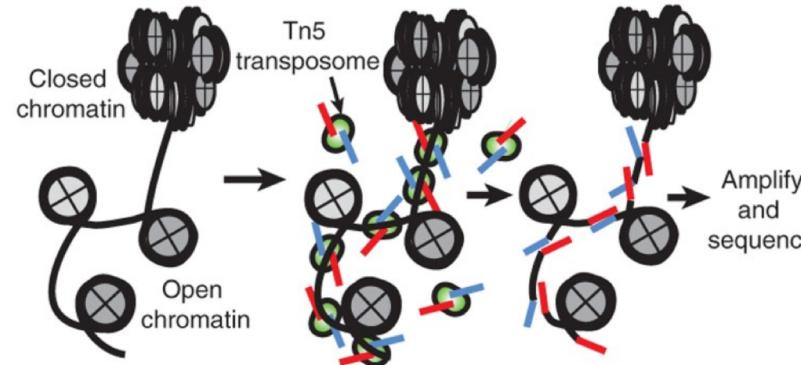
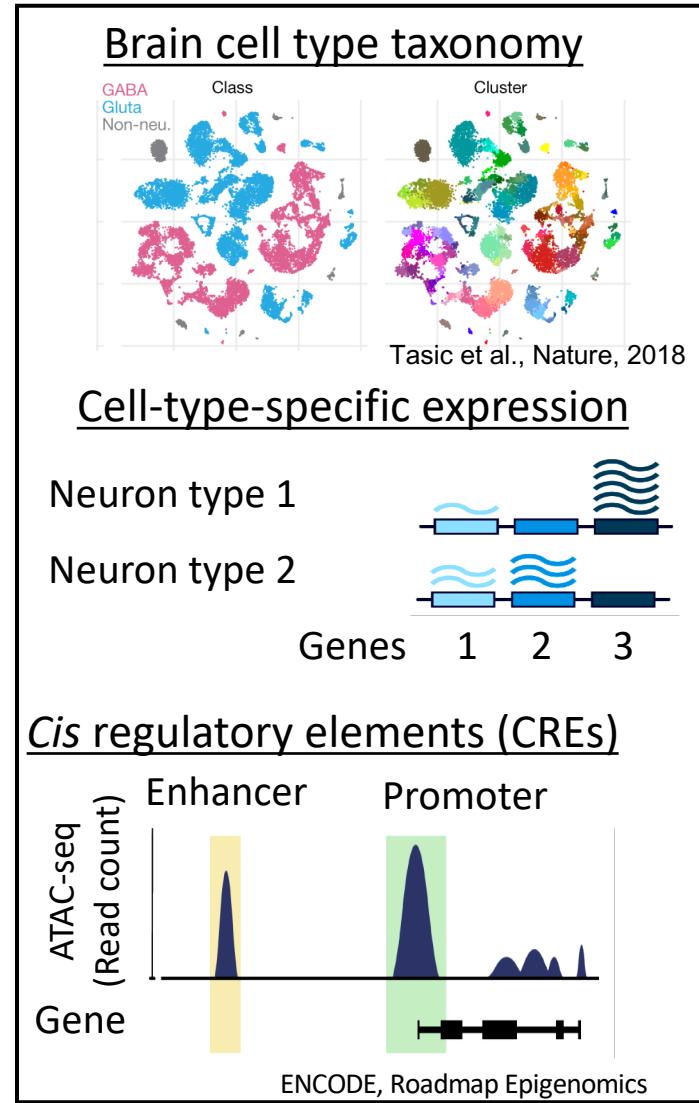
What gene regulatory programs are underlying the unique identity and function of cell types?



Light is “regulated” by switches.



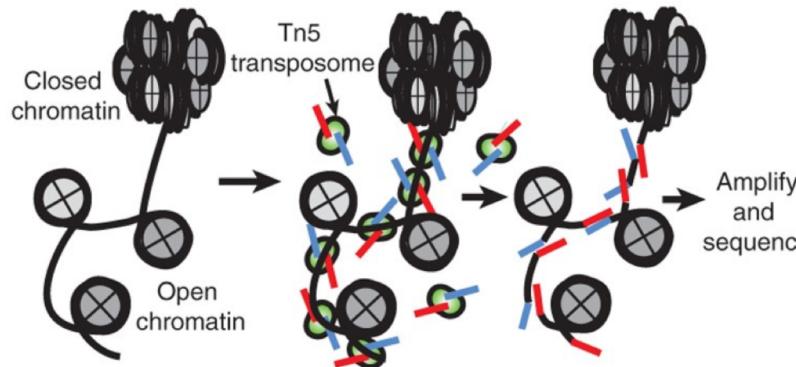
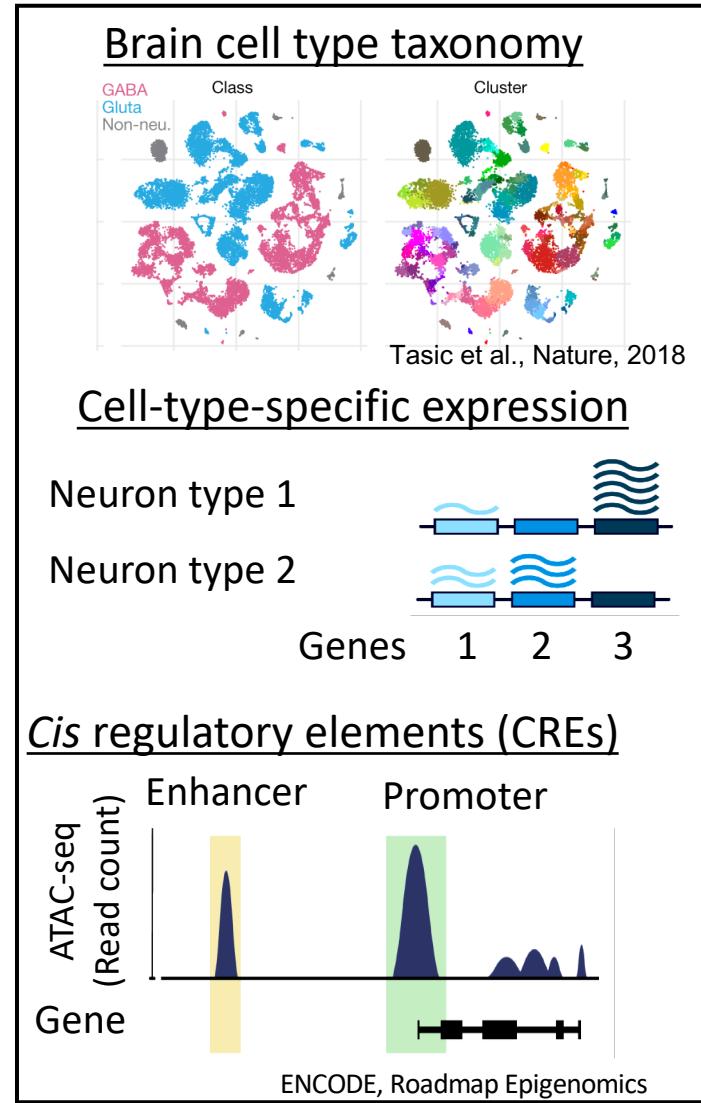
What gene regulatory programs are underlying the unique identity and function of cell types?



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

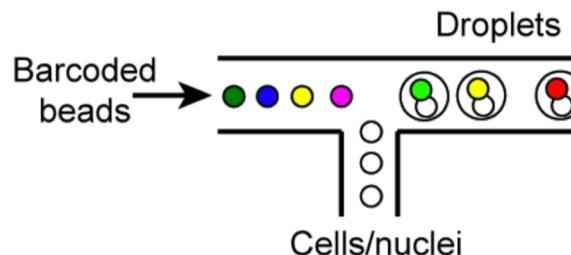
What gene regulatory programs are underlying the unique identity and function of cell types?



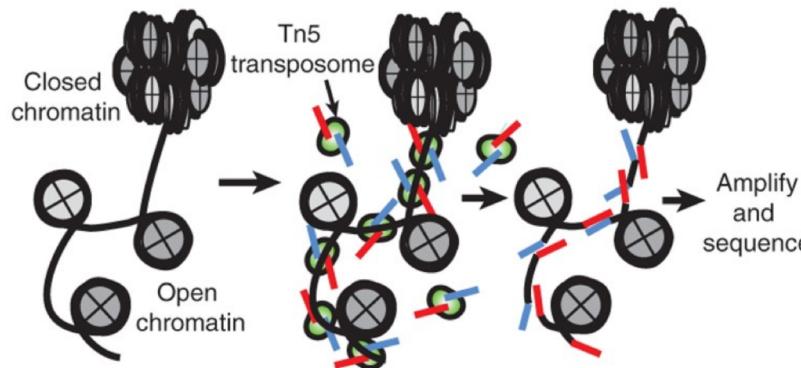
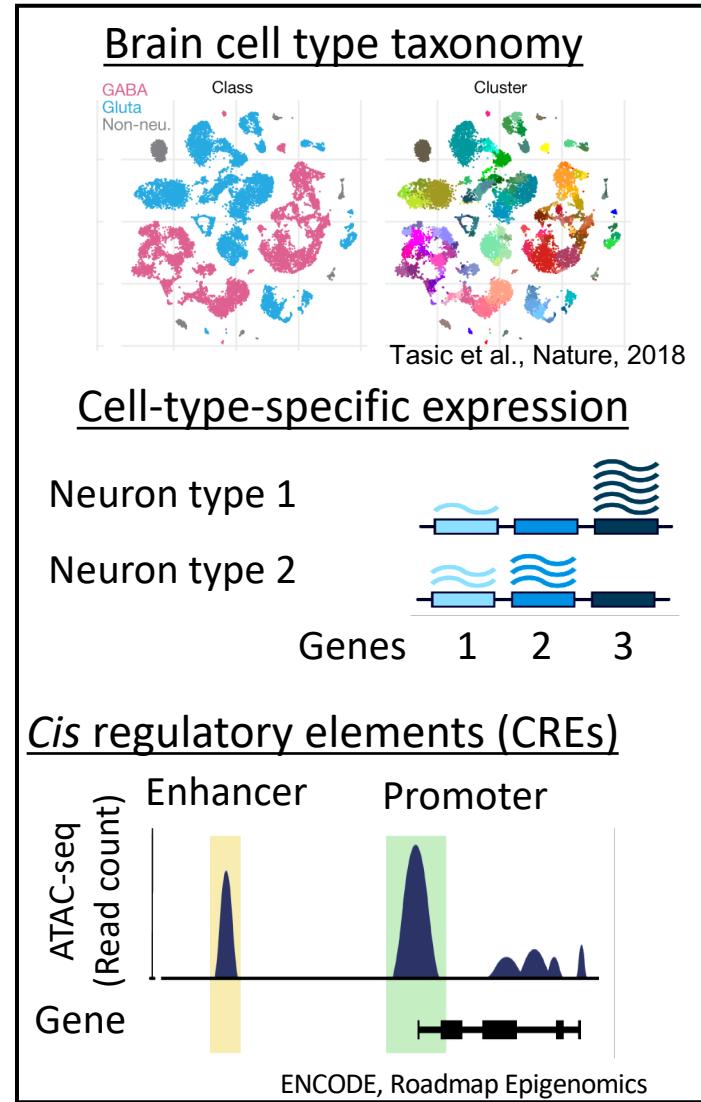
ATAC-seq identifies accessible DNA regions by probing **open chromatin** with hyperactive mutant **Tn5 Transposase** that inserts sequencing adapters into open regions of the genome.

Buenrostro, et al. Nature Methods, 2013

□ Droplet-based (e.g. 10x Genomics)



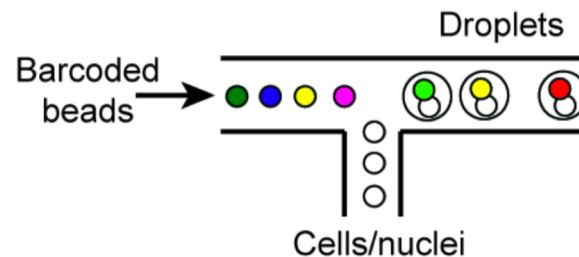
What gene regulatory programs are underlying the unique identity and function of cell types?



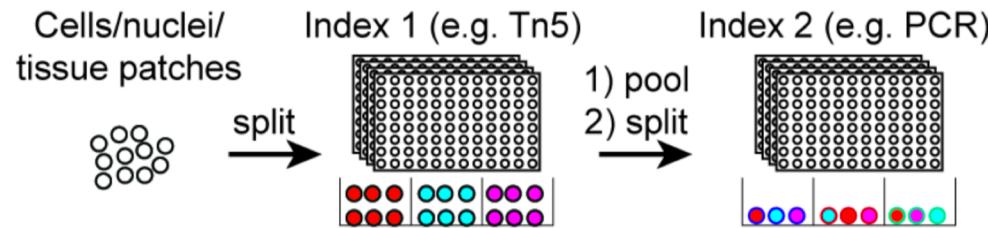
ATAC-seq identifies accessible DNA regions by probing **open chromatin** with hyperactive mutant **Tn5 Transposase** that inserts sequencing adapters into open regions of the genome.

Buenrostro, et al. Nature Methods, 2013

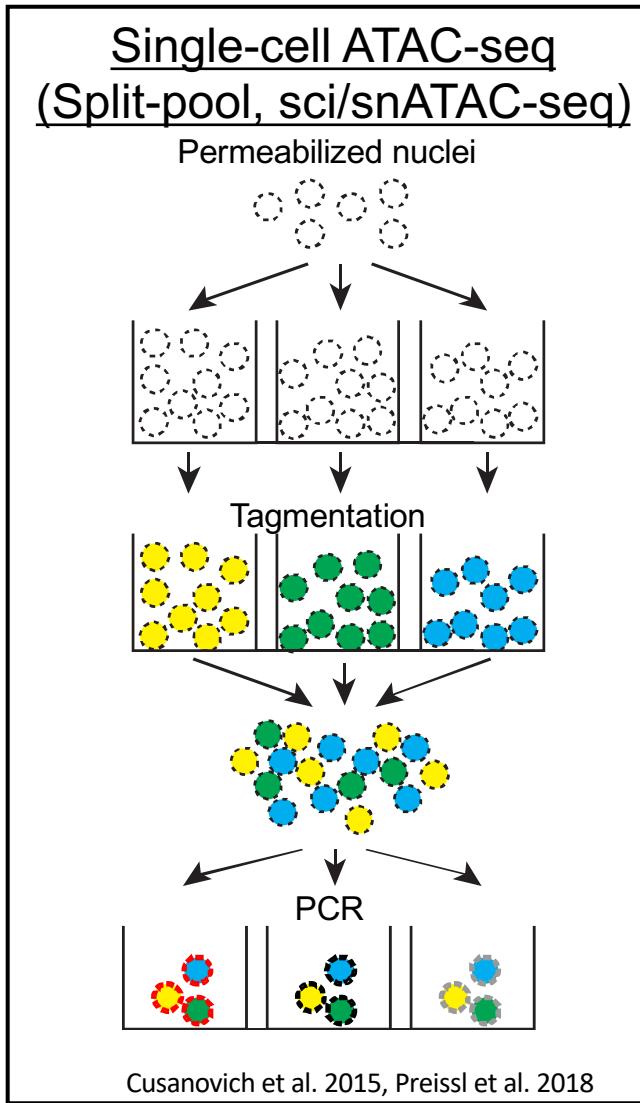
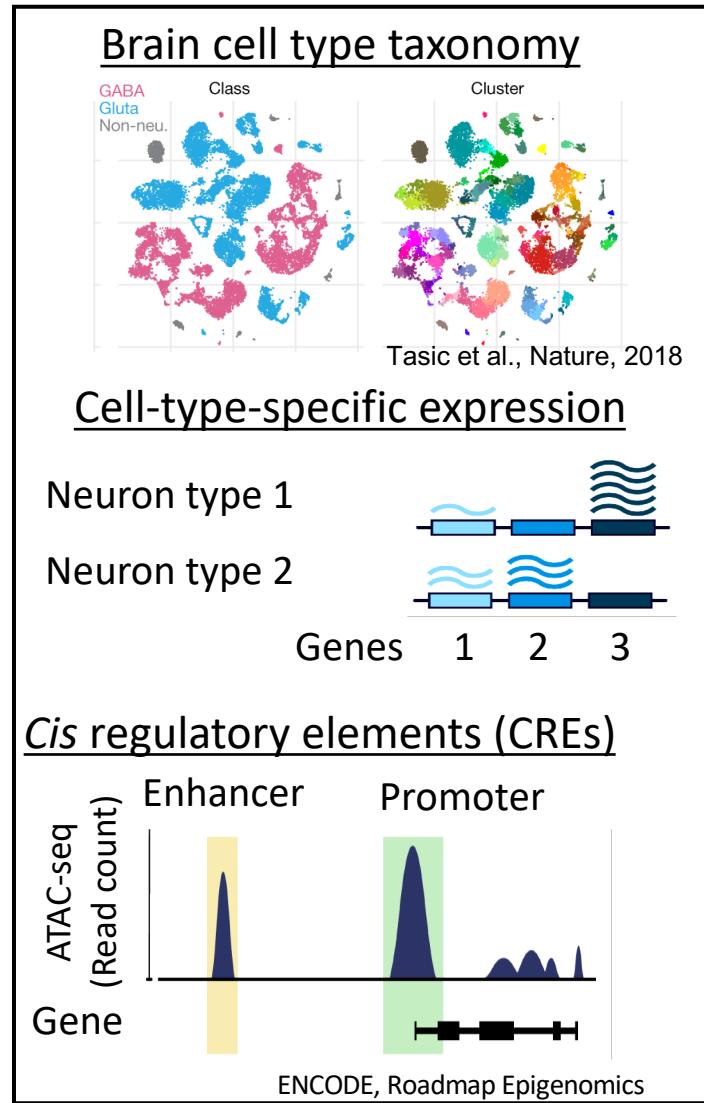
Droplet-based (e.g. 10x Genomics)



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



What gene regulatory programs are underlying the unique identity and function of cell types?



Dr. Sebastian Preissl

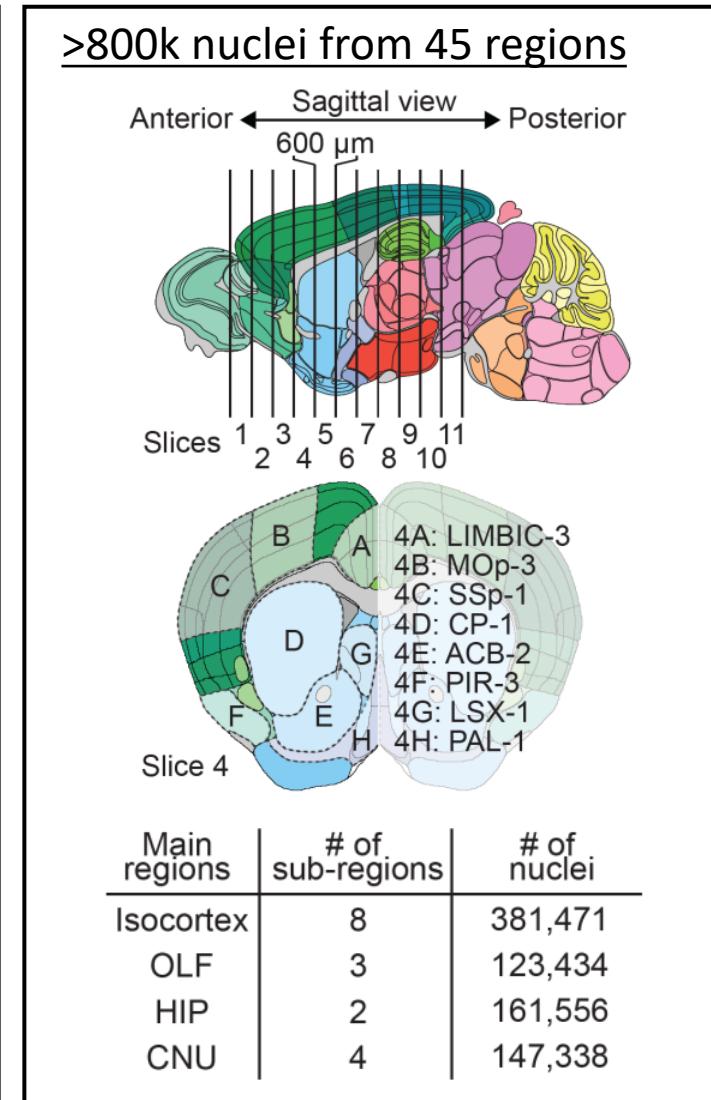
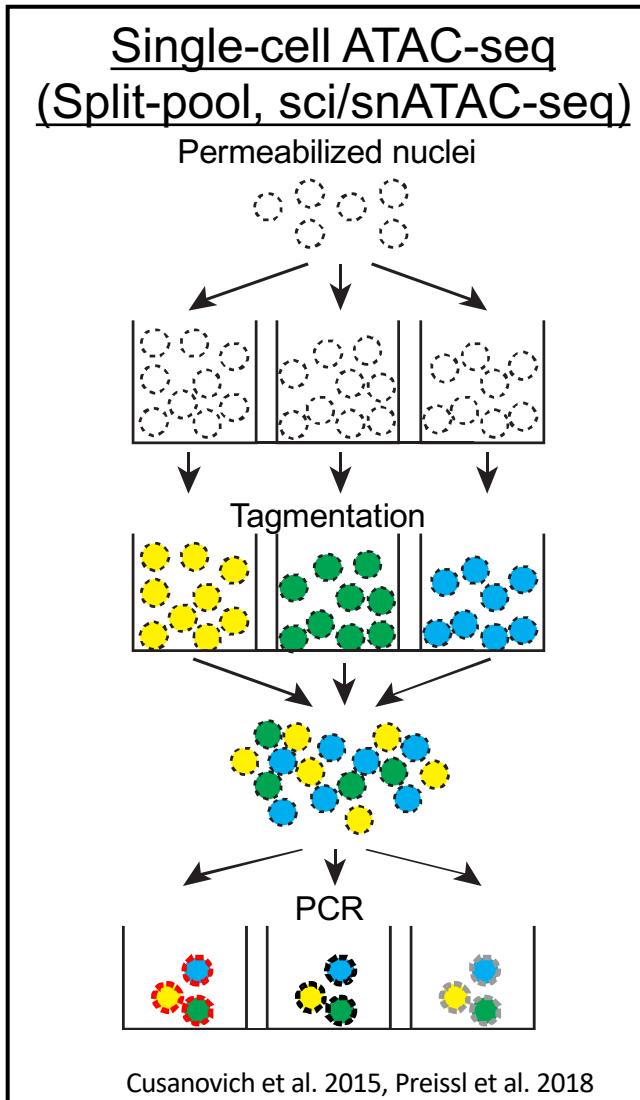
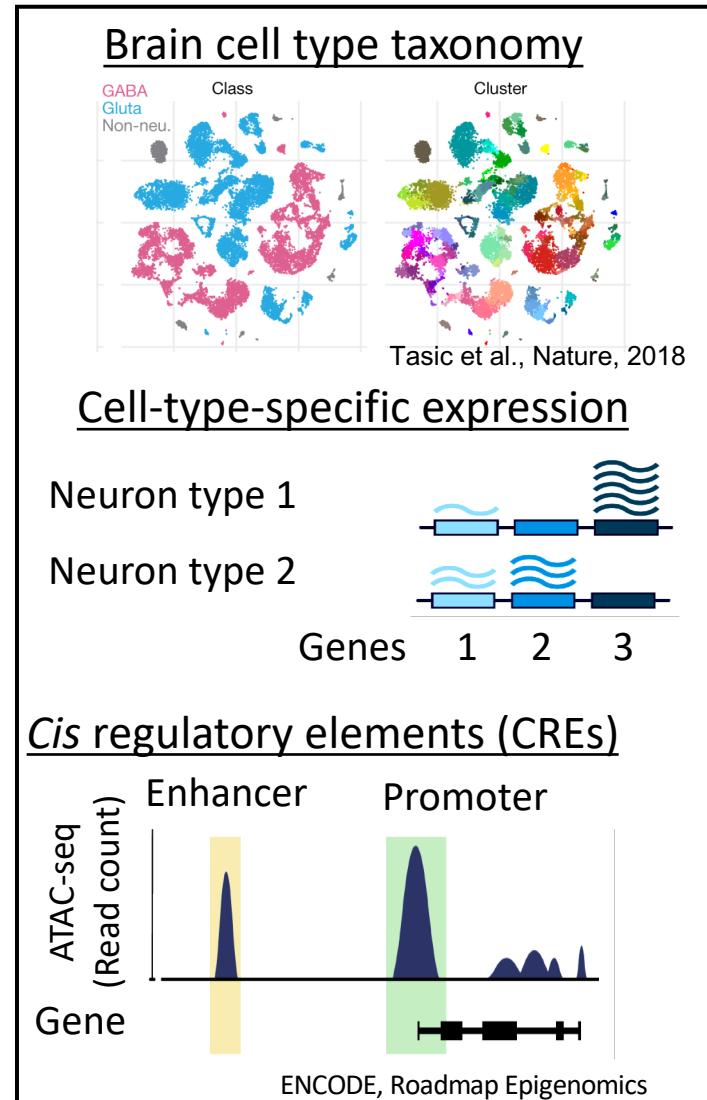
Associate Director
Single-cell Genomics



Xiaomeng Hou

Research Associate

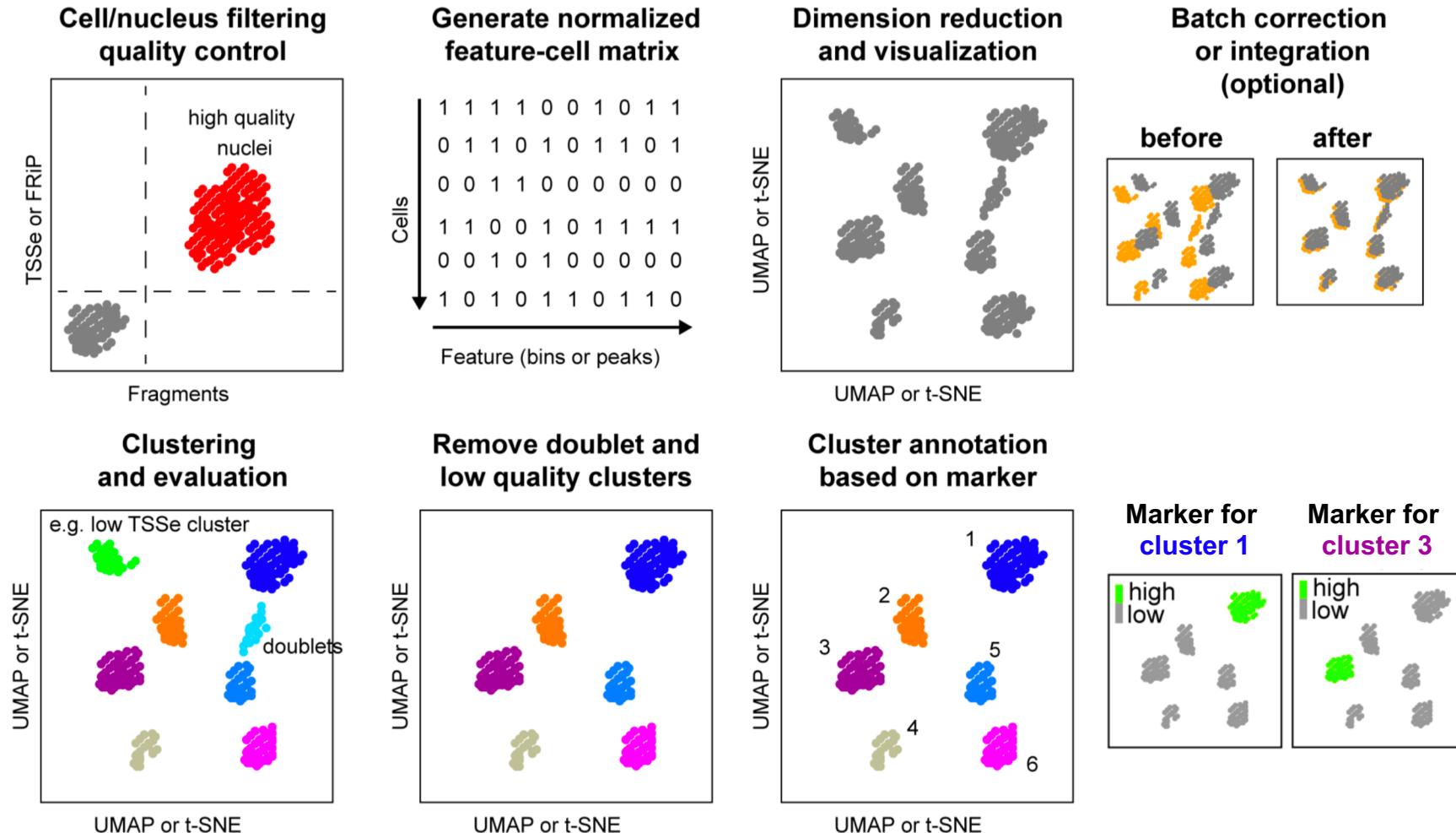
What gene regulatory programs are underlying the unique identity and function of cell types?



OLF: Olfactory bulb; HIP: Hippocampus; CNU: Cerebral nuclei (striatum, pallidum)

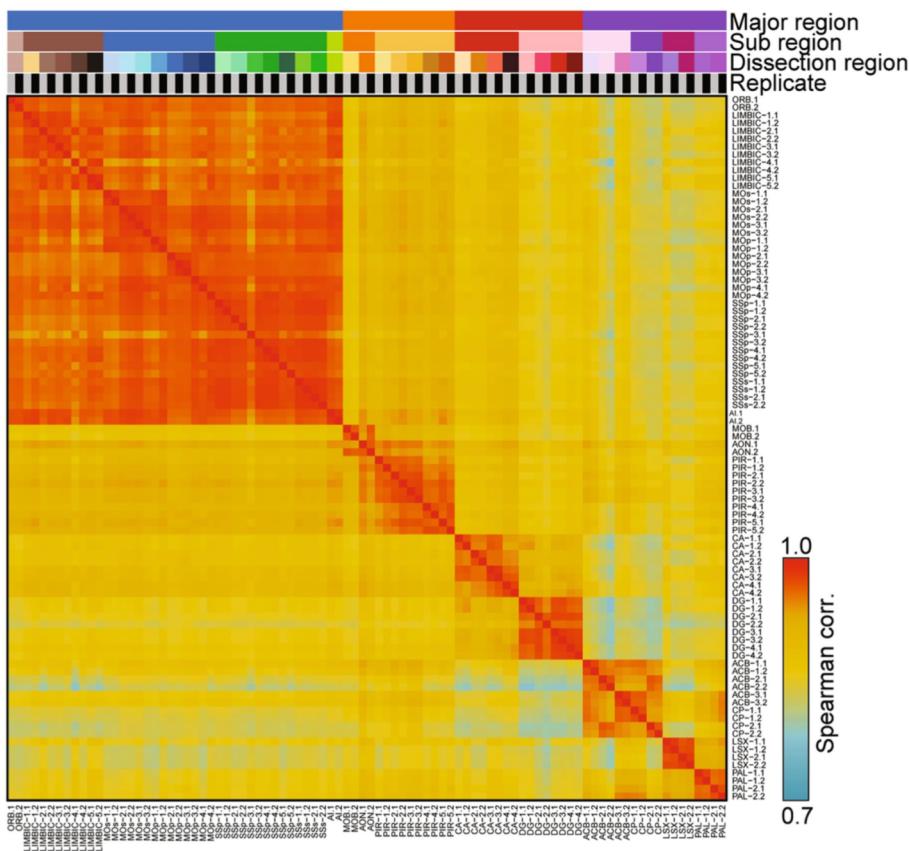
Overview of data analysis

Data Processing and Clustering

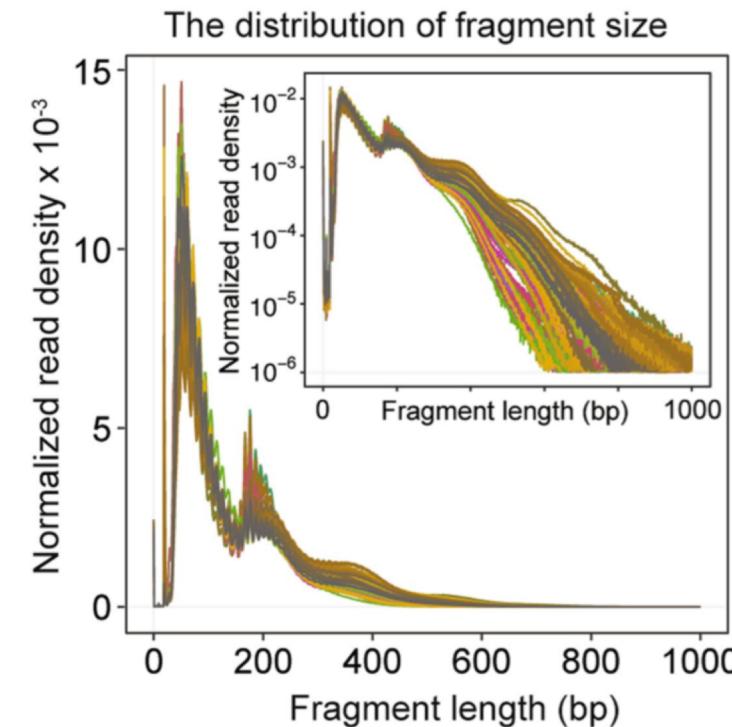


Quality control at bulk level

Pairwise Spearman Correlation Coefficients
between biological replicates



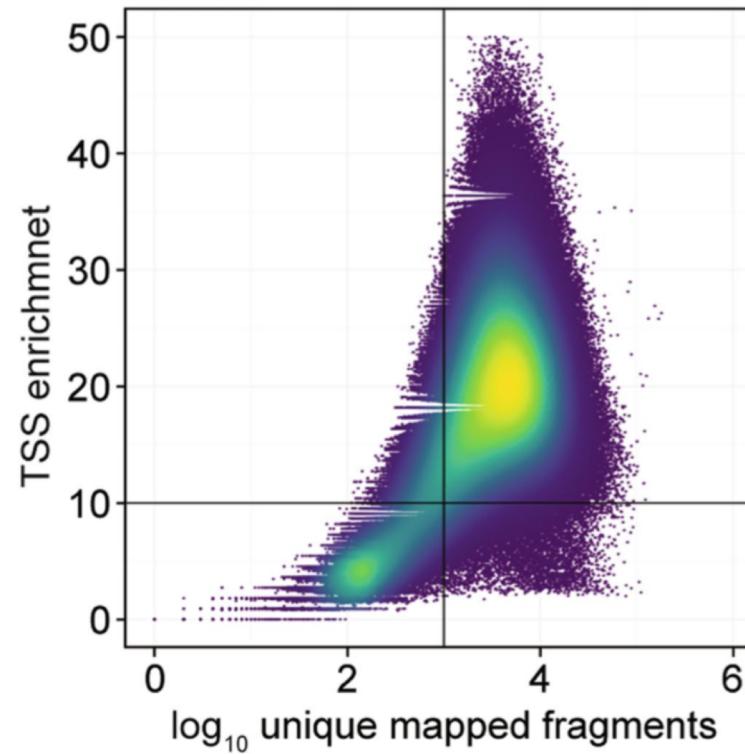
- A large proportion of reads with less than 100 bp, which represents the nucleosome-free region.
- The fragment size distribution should have a clear periodicity



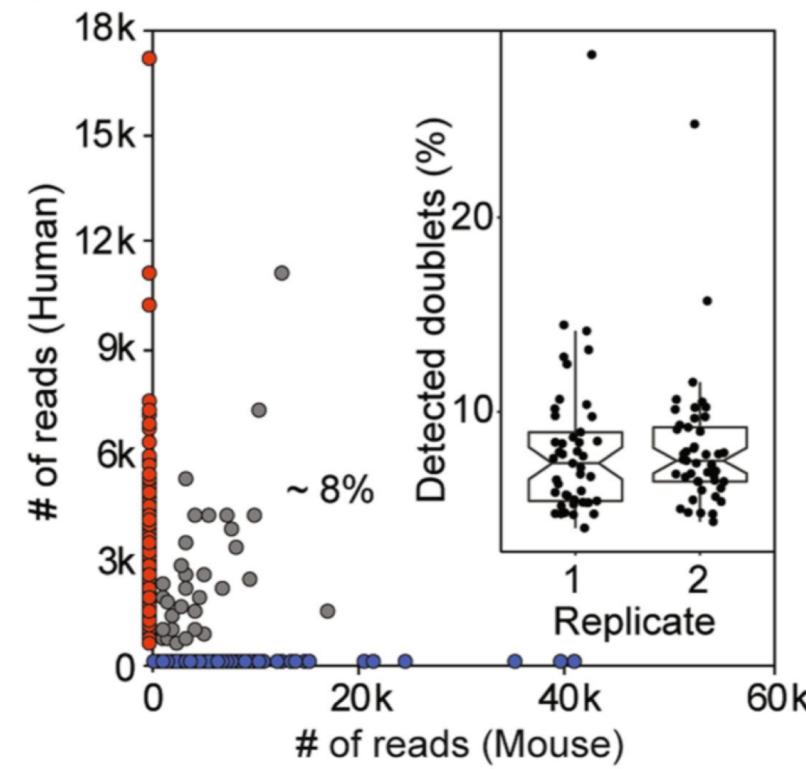
Quality control at single-cell/nucleus level

Keep single cell/nucleus has:

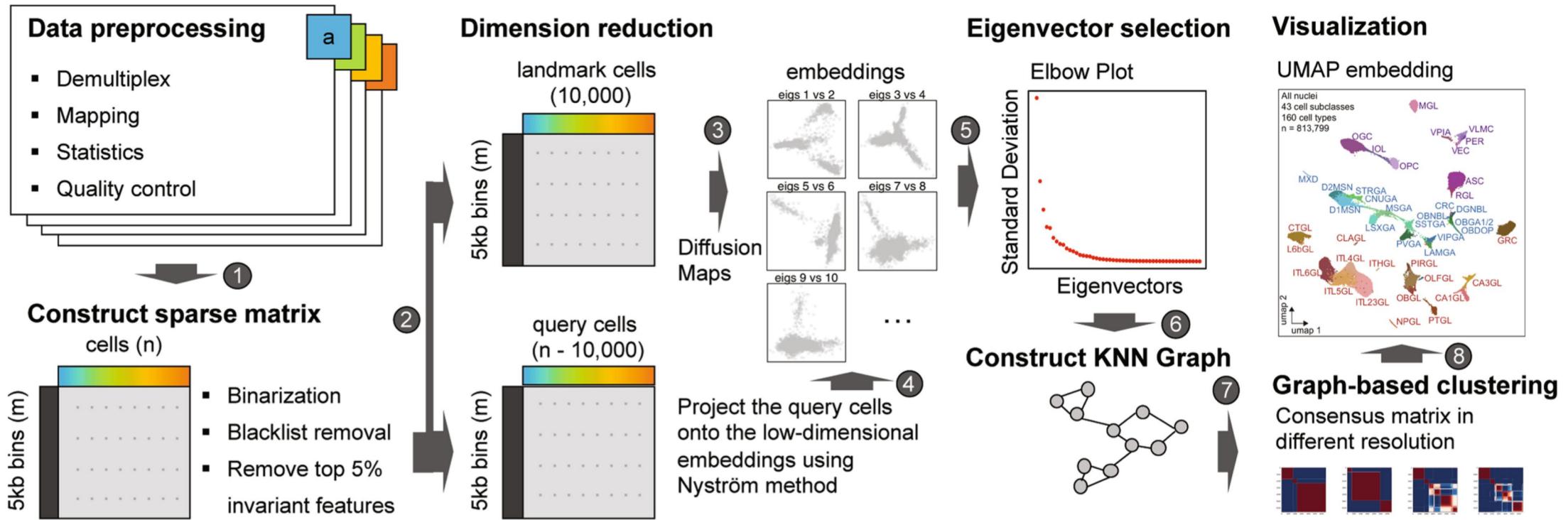
- more sequenced fragments, as well as
- high signal-to-noise ratio.



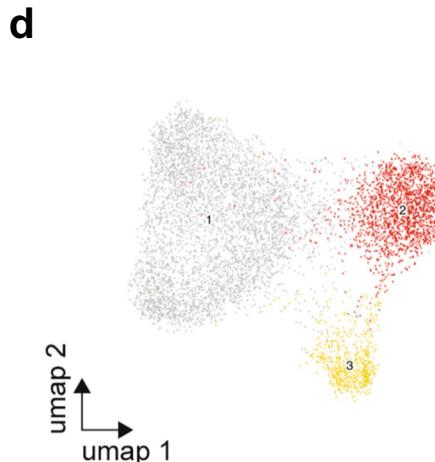
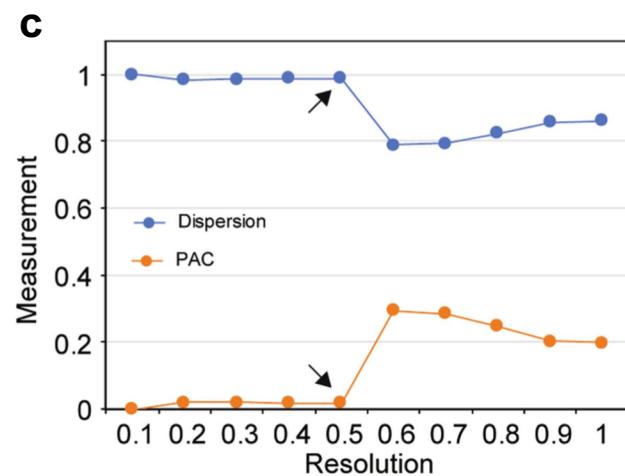
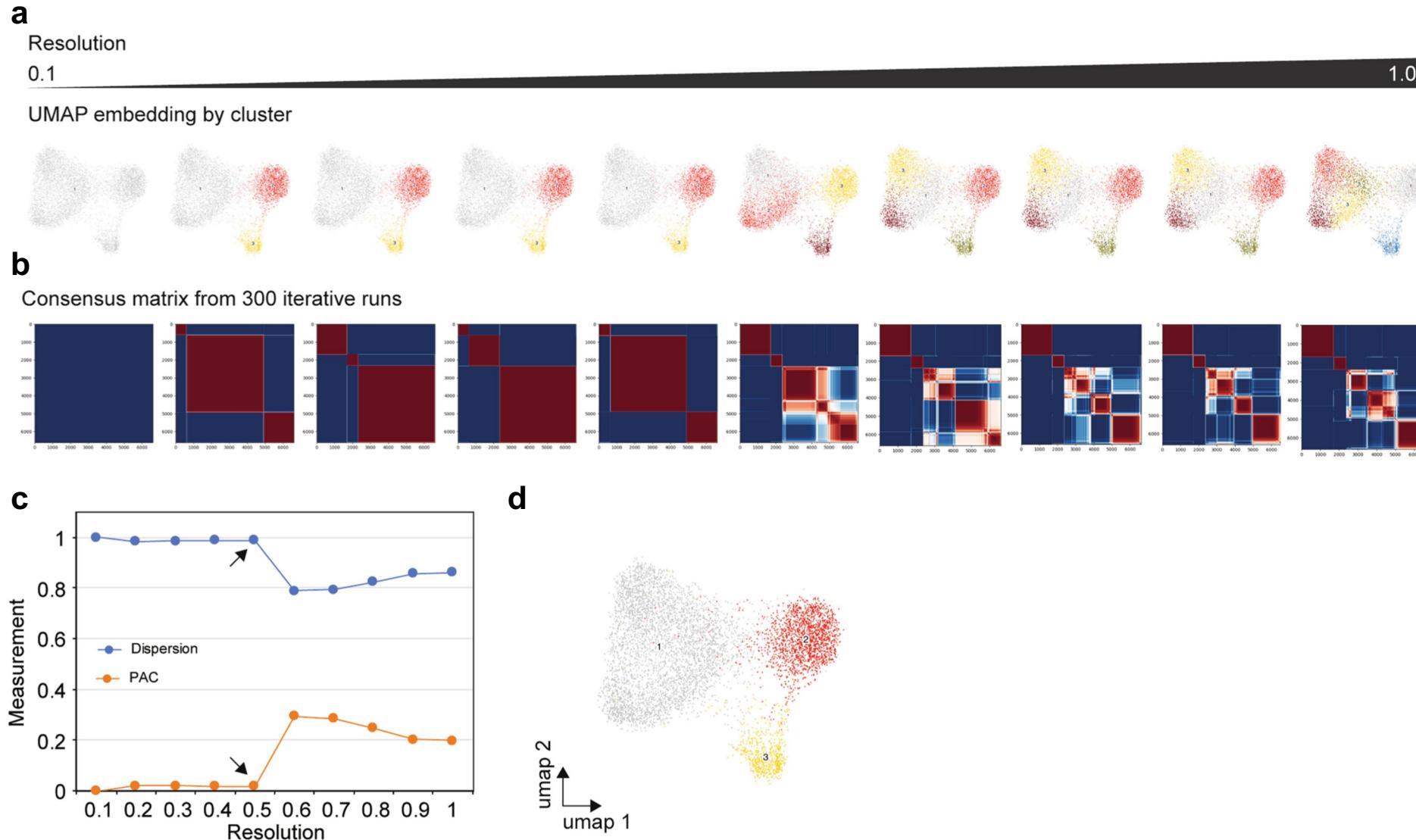
Potential doublets should be identified and further removed.



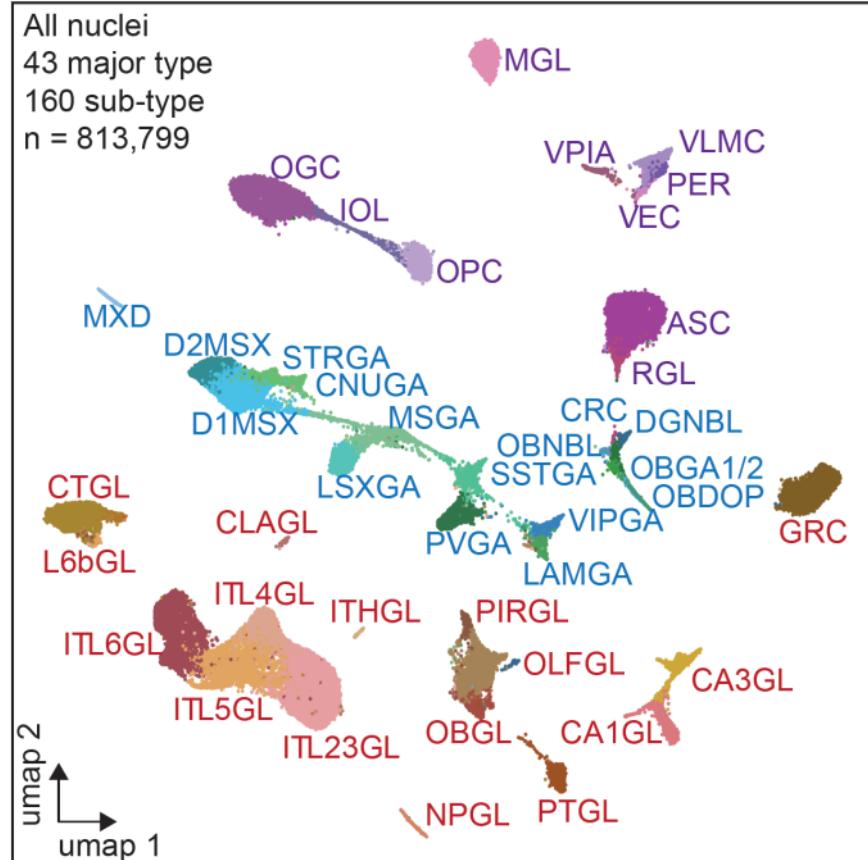
Cell clustering pipeline



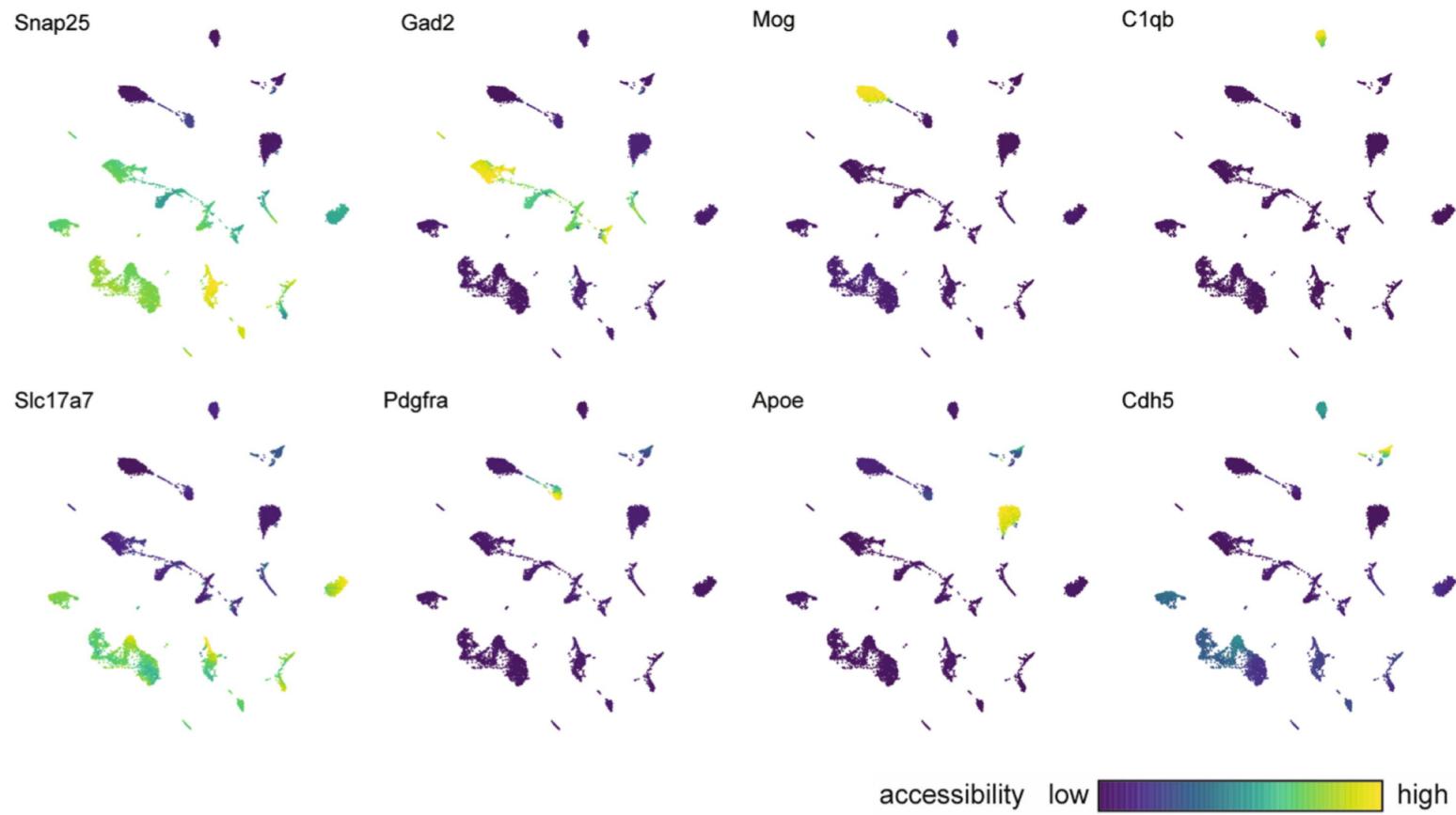
Cell clustering pipeline



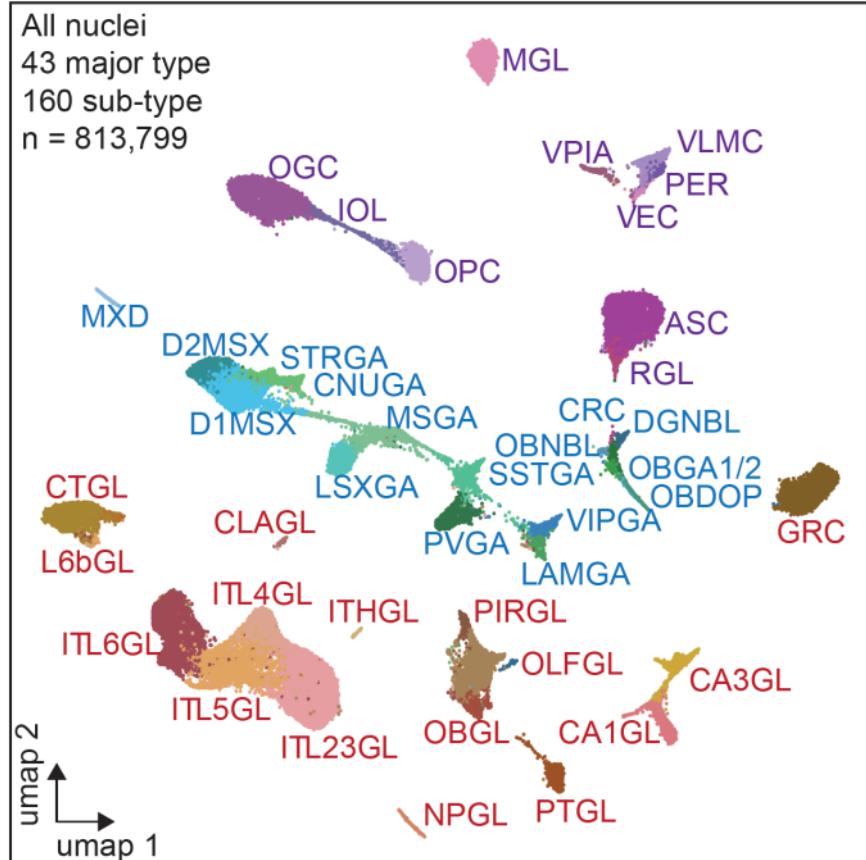
Single cell analysis of chromatin accessibility in the mouse cerebrum reveals 160 sub-types



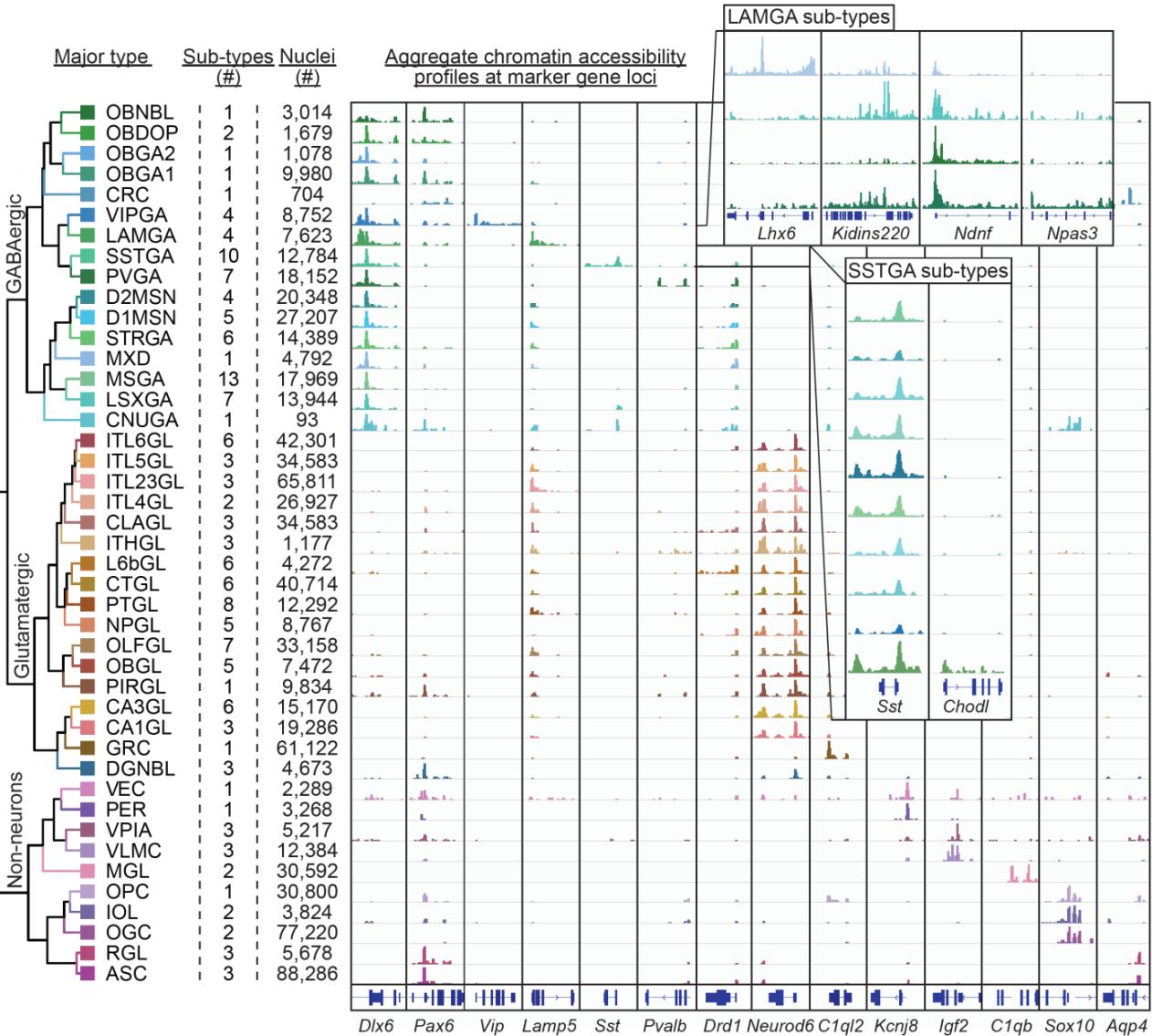
Non-neurons (21 sub-types) n = 259,558	GABAergic neurons (71 sub-types) n = 167,181	Glutamatergic neurons (68 sub-types) n = 387,060
--	--	--



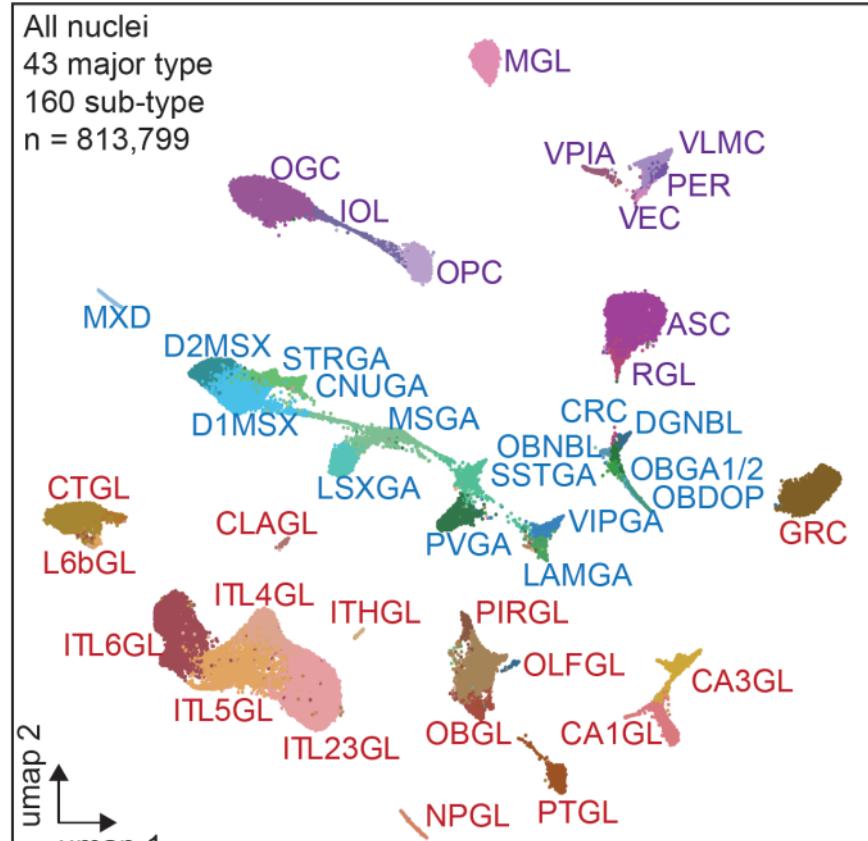
Single cell analysis of chromatin accessibility in the mouse cerebrum reveals 160 sub-types



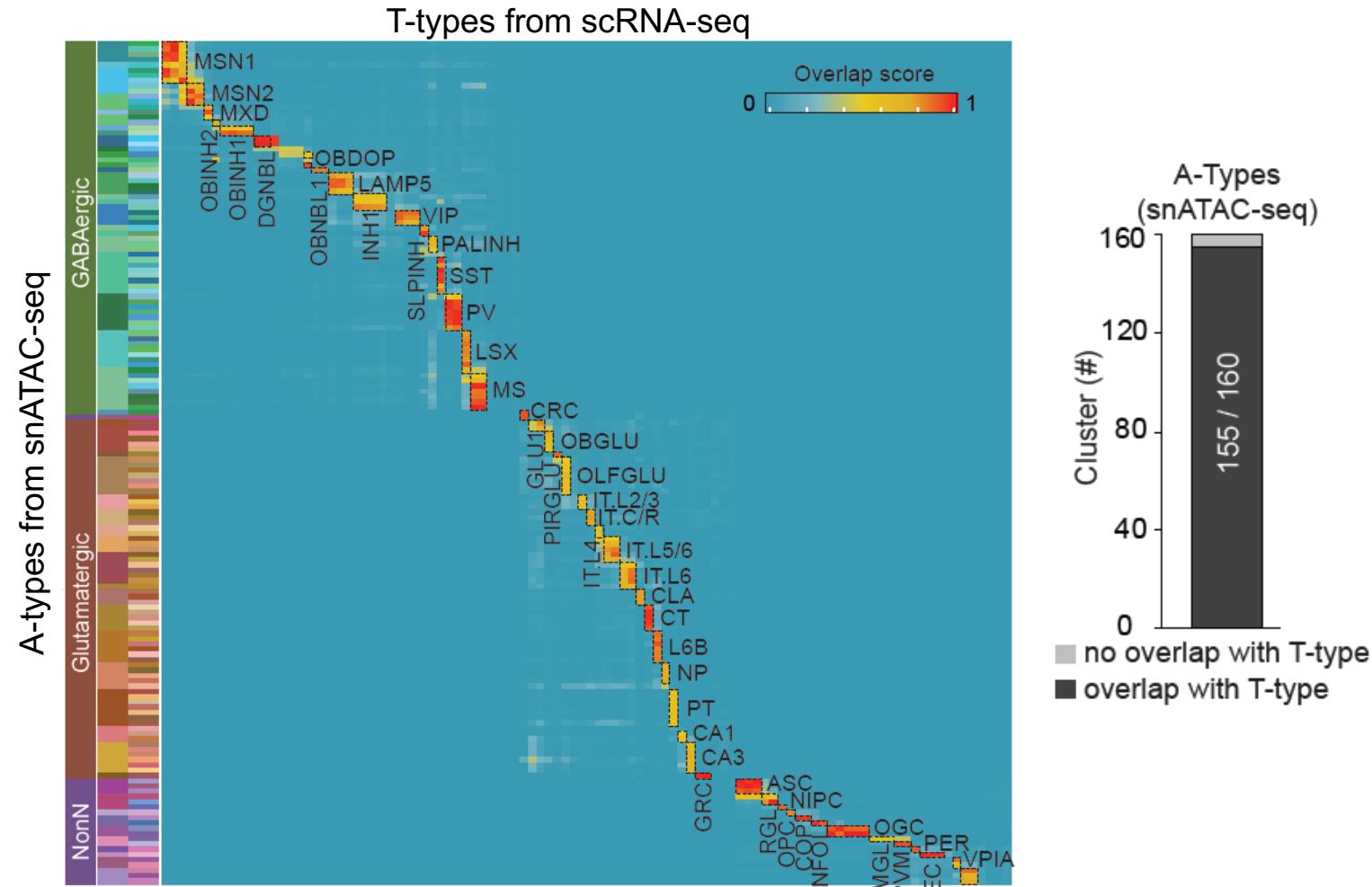
Non-neurons (21 sub-types) n = 259,558	GABAergic neurons (71 sub-types) n = 167,181	Glutamatergic neurons (68 sub-types) n = 387,060
--	--	--



Single cell analysis of chromatin accessibility in the mouse cerebrum reveals 160 sub-types



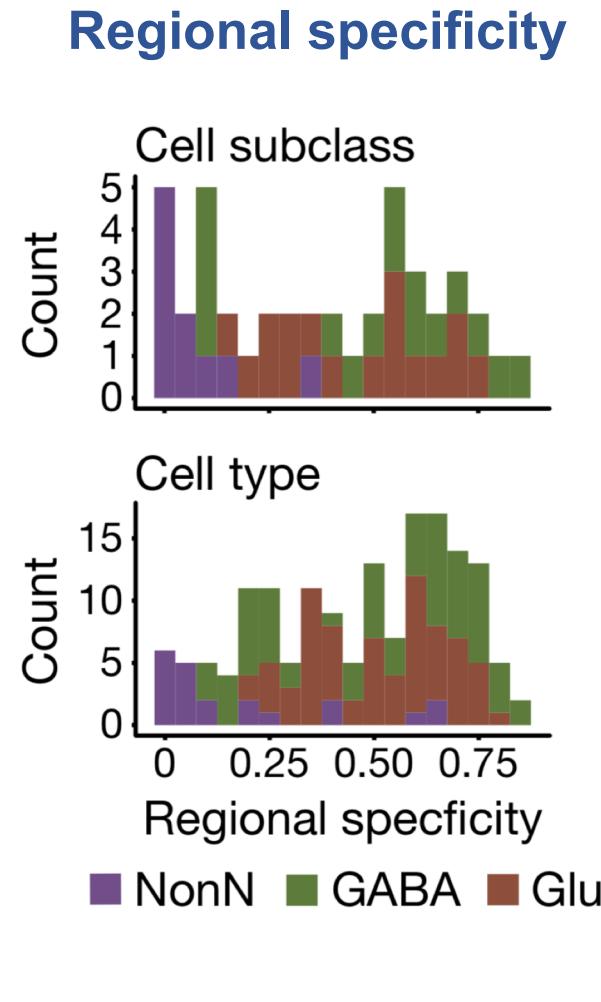
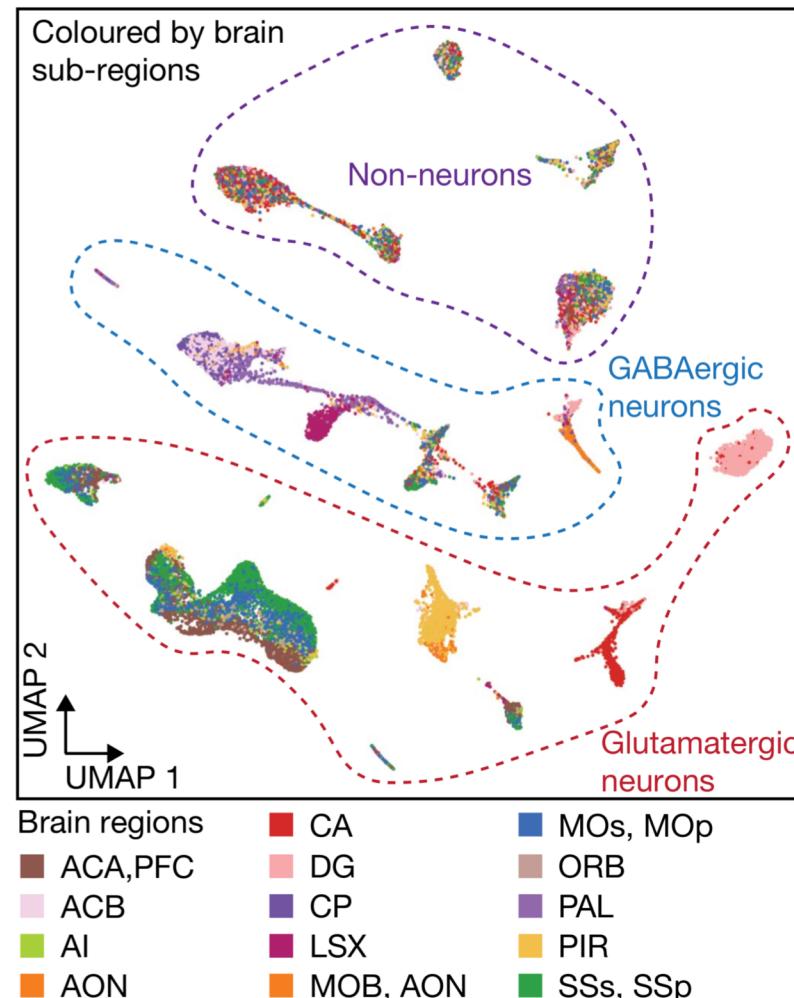
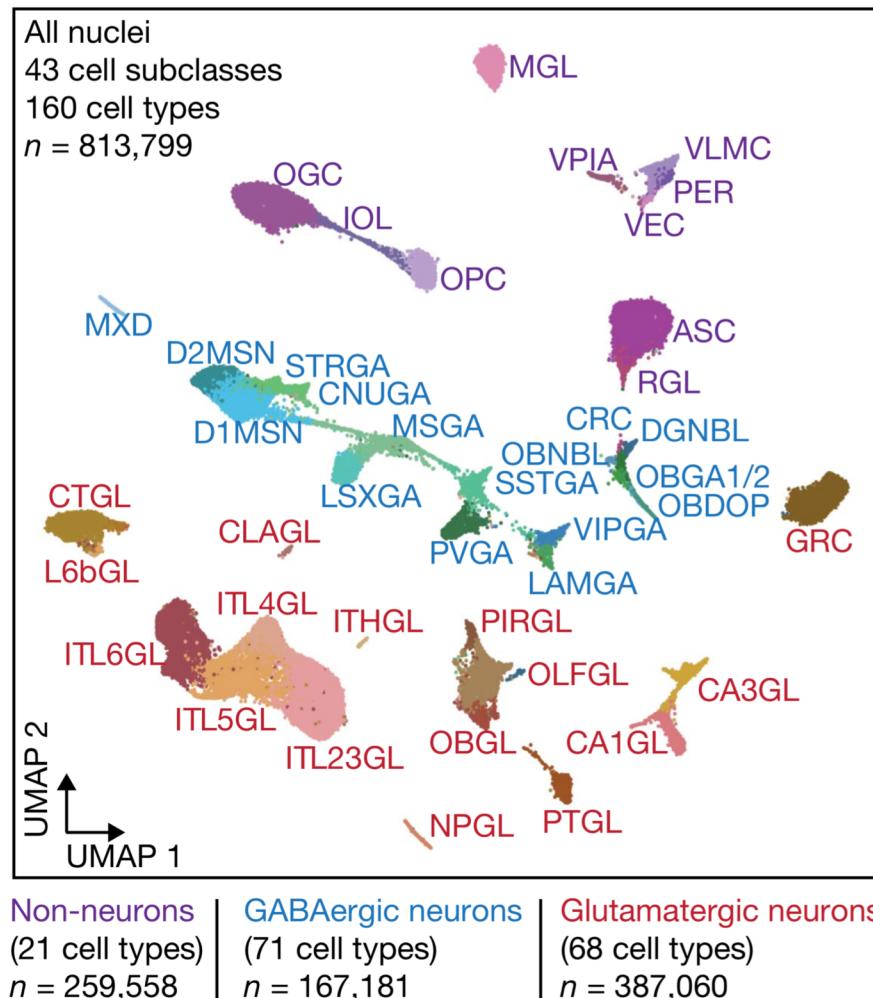
Non-neurons (21 sub-types) n = 259,558	GABAergic neurons (71 sub-types) n = 167,181	Glutamatergic neurons (68 sub-types) n = 387,060
--	--	--



Zeisel et al., Cell 2018

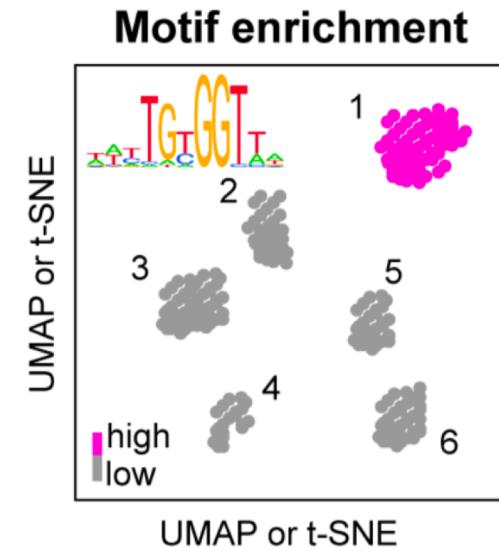
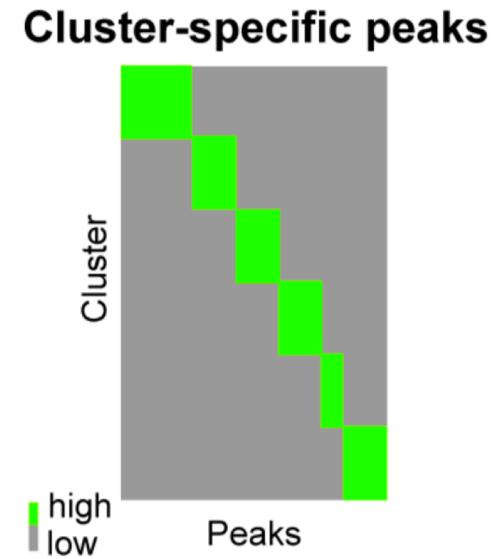
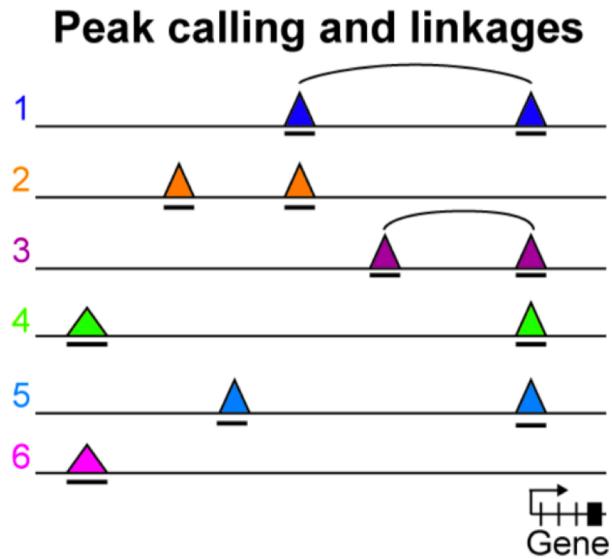
A-types: types derived from chromatin accessibility
T-types: cell types derived from transcriptomes

Regional Specificity of different brain cell types

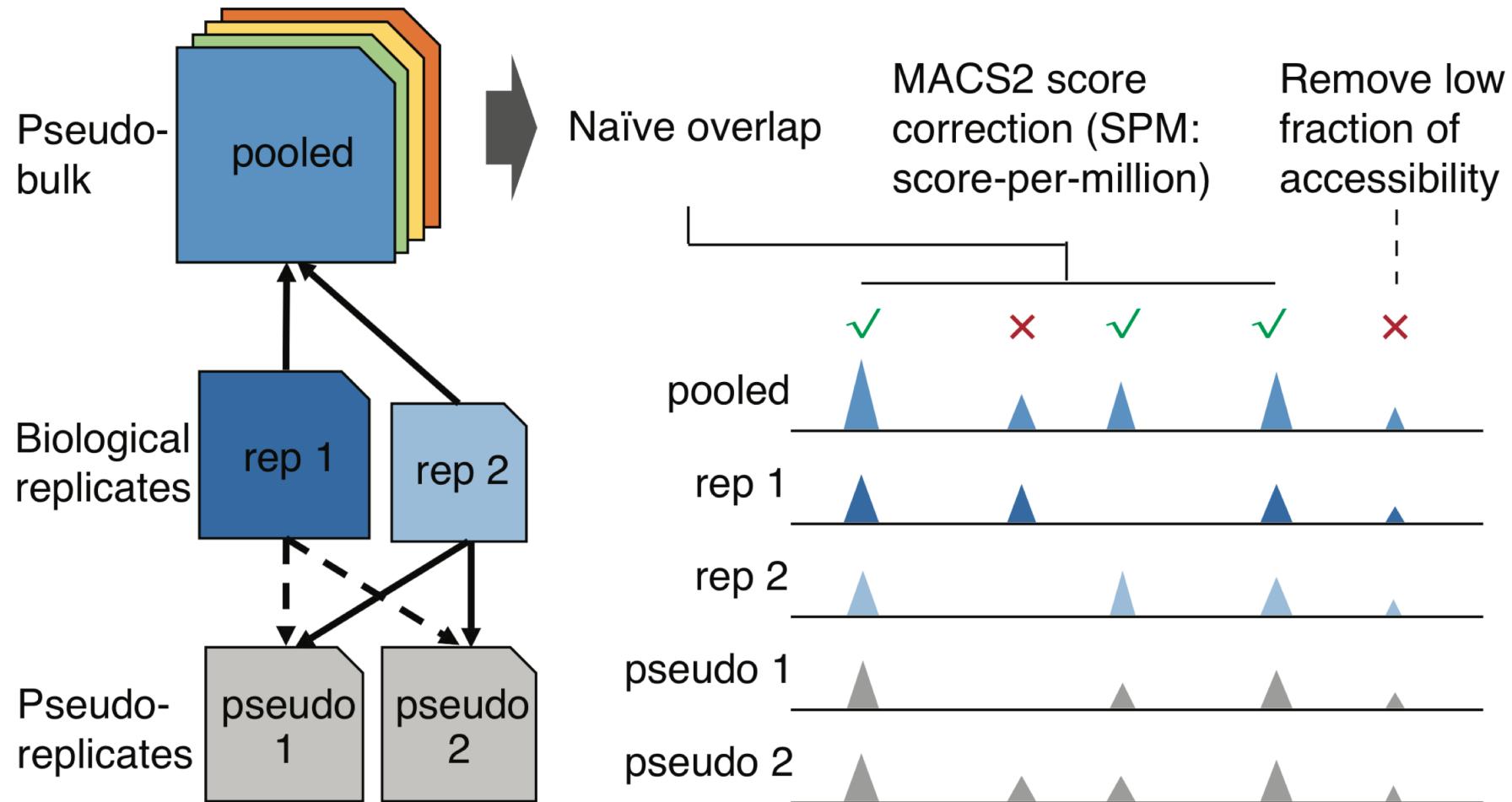


Overview of downstream analysis

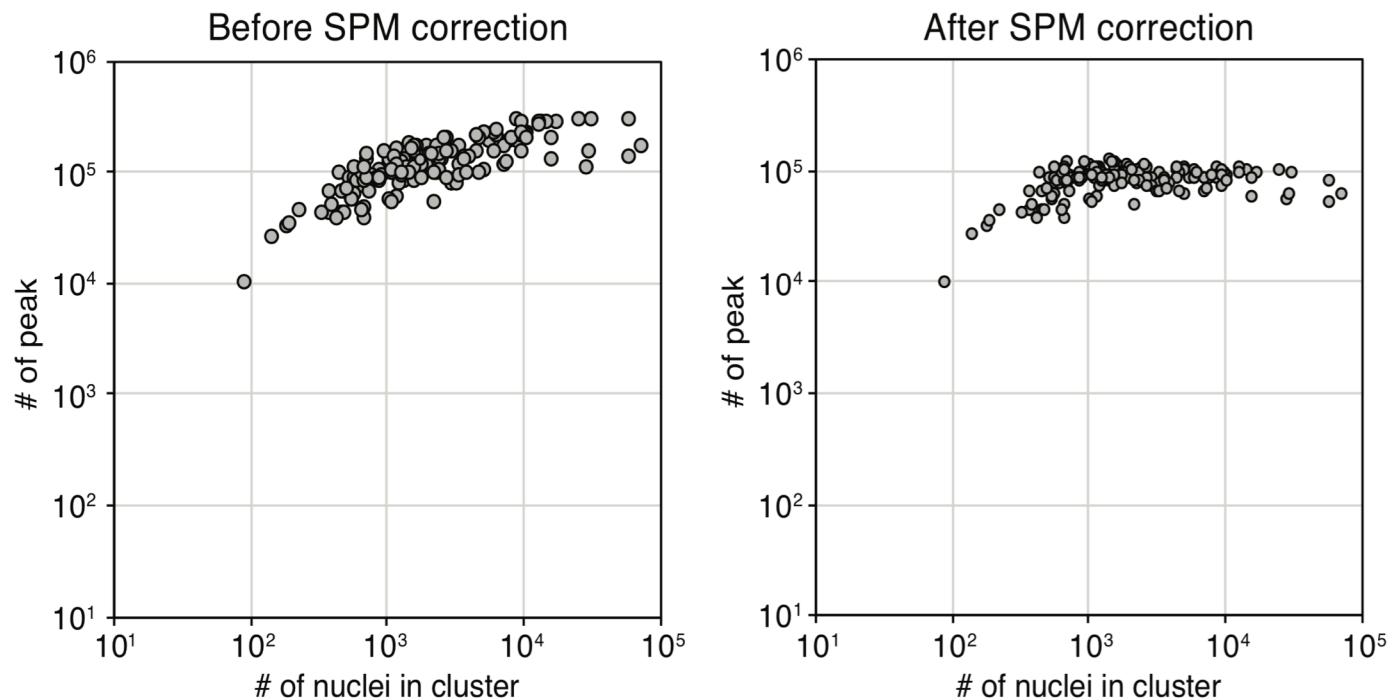
Downstream characterization



Identification of candidate cis-regulatory elements (cCREs)

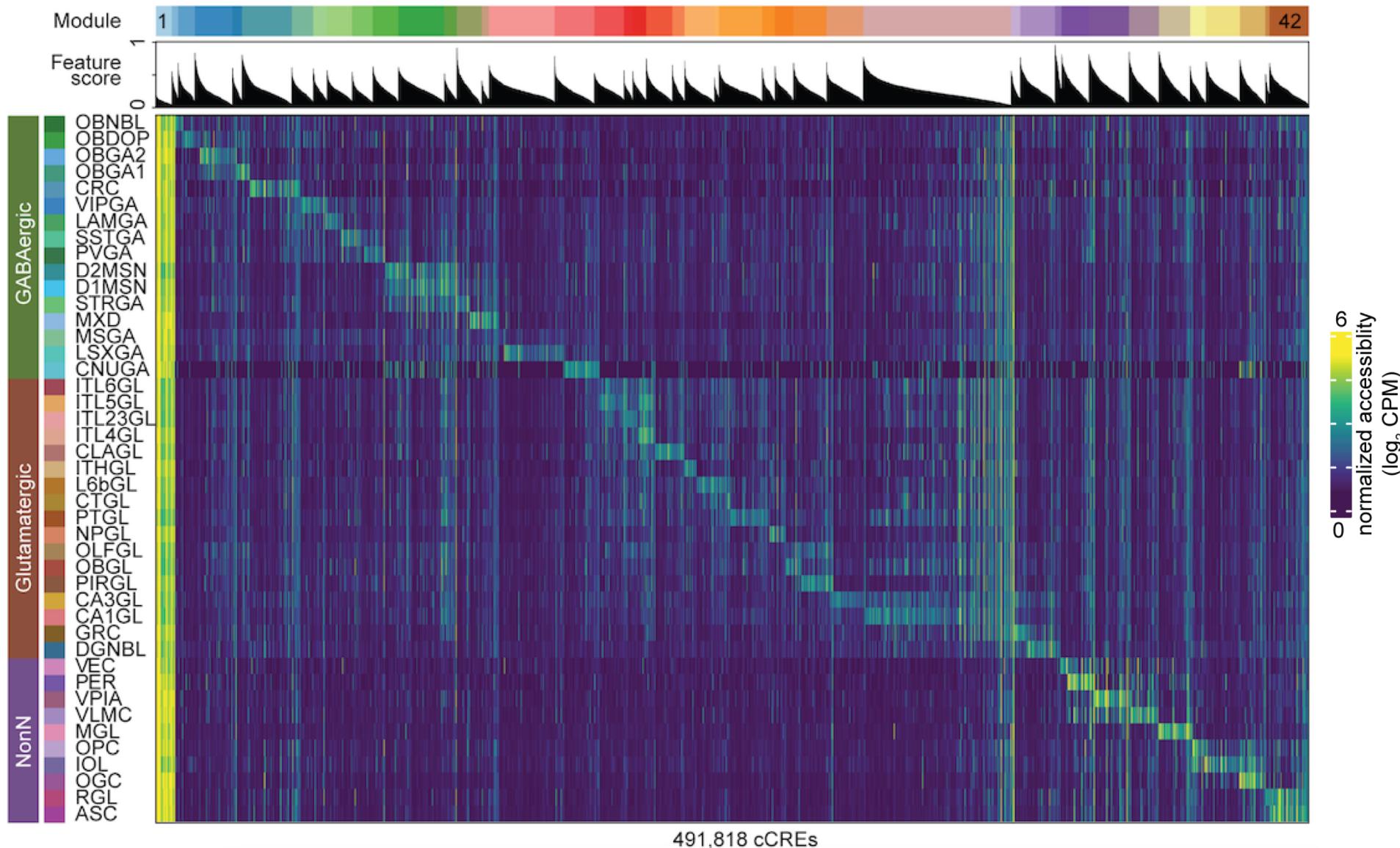


For single-cell dataset, we need to account for different sequencing depth and/or the number of nuclei in individual clusters



“score per million (SPM)”: corrected MACS2 peak scores ($-\log_{10}(q\text{-value})$)
First introduced in the analysis of bulk ATAC-seq from tumor samples
(Corces, M. R. et al. Science, 2018)

A catalogue of 491,818 cCREs in the mouse cerebrum with cell-type specific accessibility



Differential analysis in single cell level

Identify differential peaks

Full model: $\text{logit}(P_{ij}) = a_j + m_j + r_j + \varepsilon_j$

Reduced model: $\text{logit}(P_{ij}) = a_j + r_j + \varepsilon_j$

P: the probability that the i^{th} site is accessible in the j^{th} cell

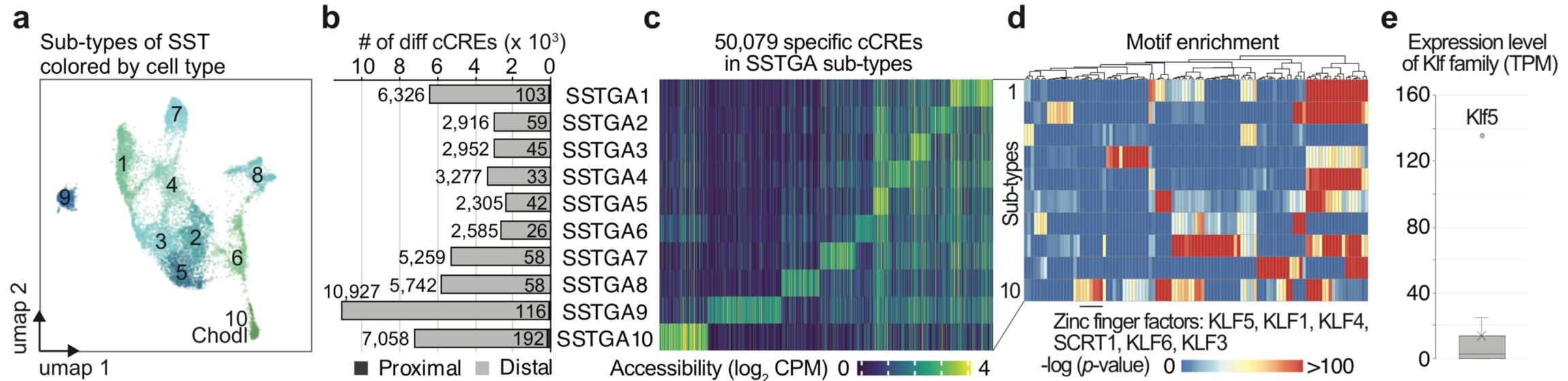
a: log10(total number of sites observed as accessible for the j^{th} cell)

m: membership of the j^{th} cell in the cluster/region being tested

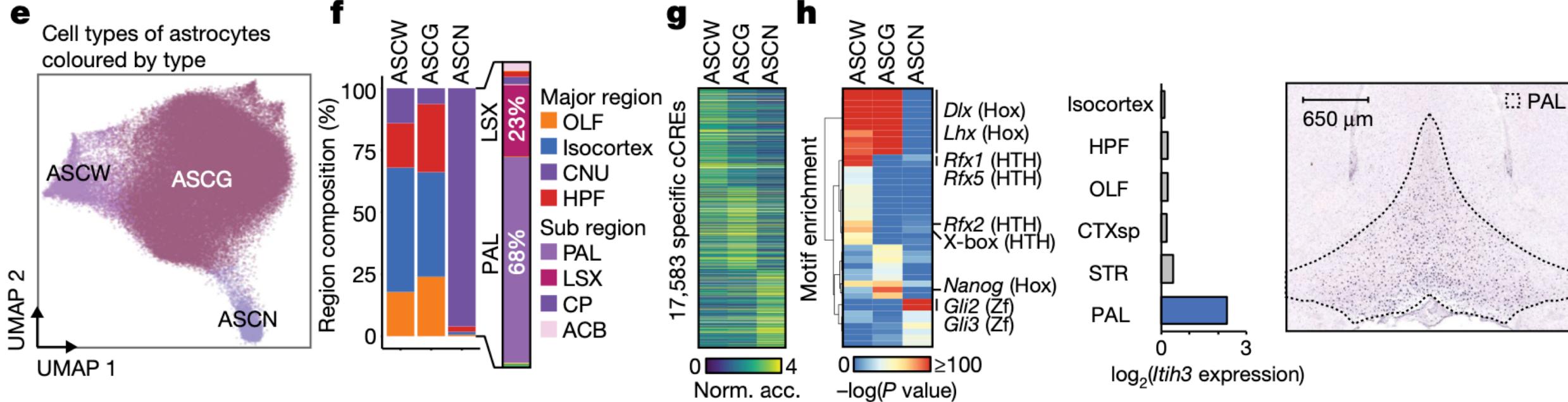
r: replicate label for j^{th} cell

Then, a likelihood ratio test is used to determine if the full model (including cell cluster membership) provided a significantly better fit of the data than reduced model

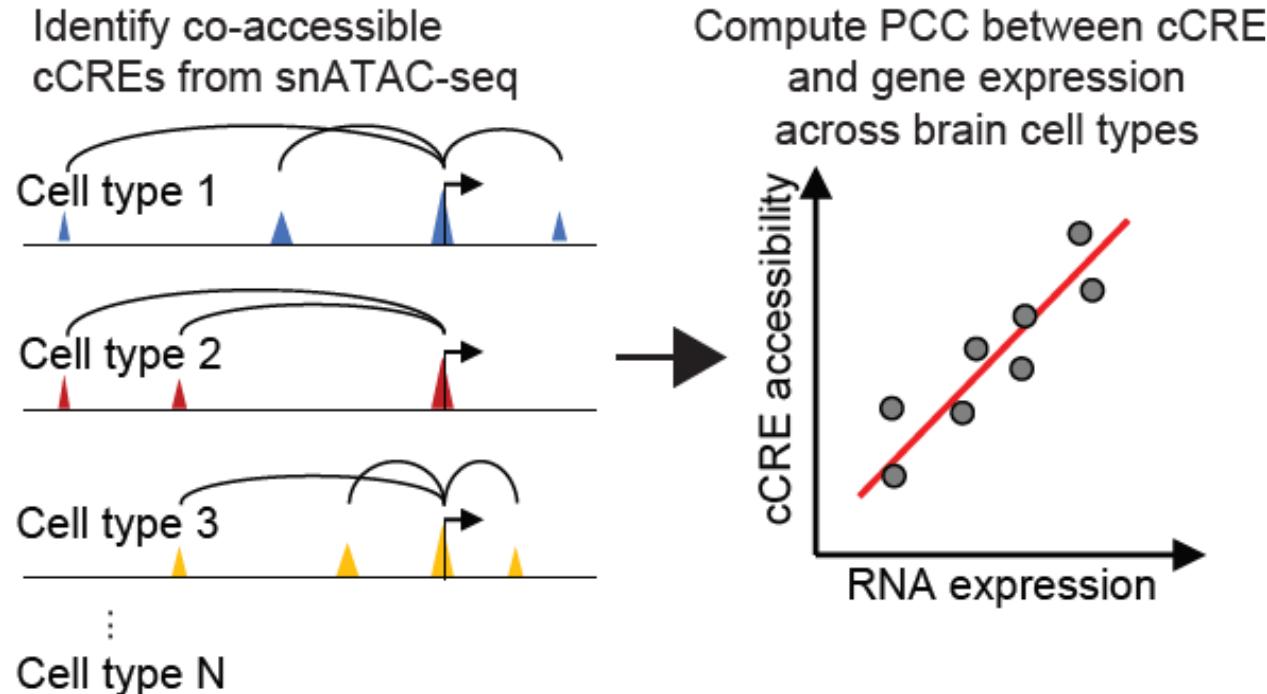
Open chromatin regions characterize distinct types of neuronal cells



Open chromatin regions characterize regional difference in cell types



Identification of putative enhancers and their target genes



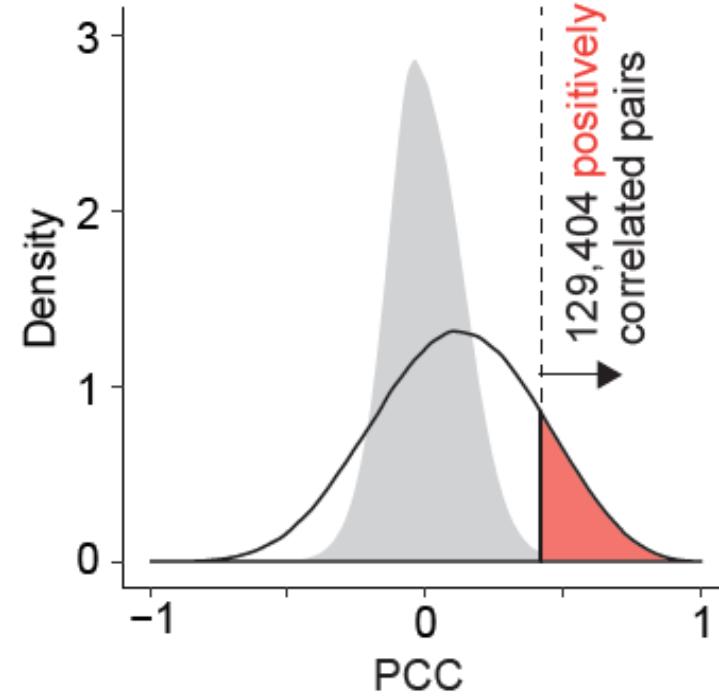
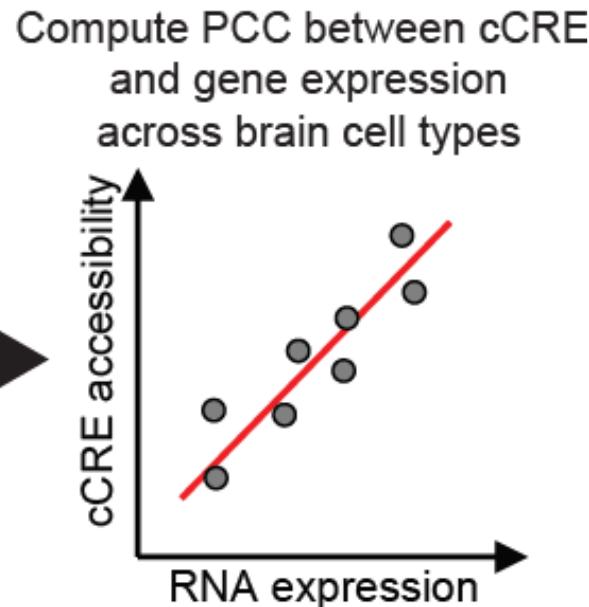
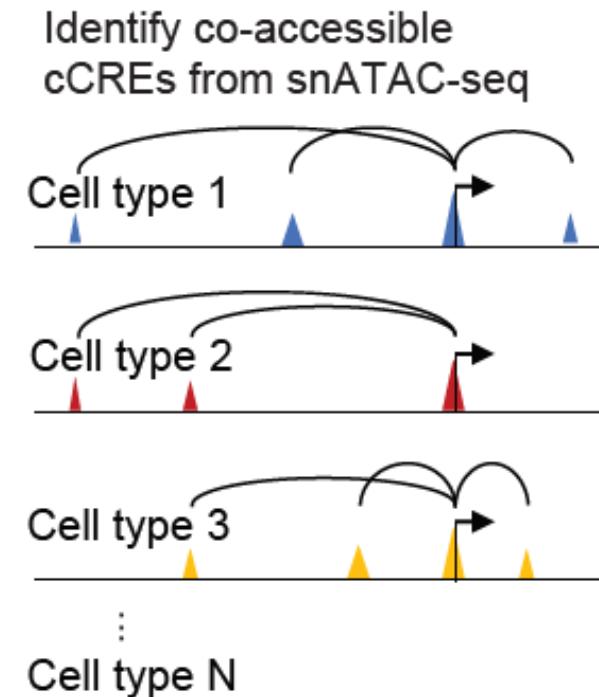
491,818 candidate *cis* regulatory elements (cCRE)

cCRE: candidate *cis* regulatory element

PCC: Pearson correlation coefficient

Co-accessible cCREs: identified by using Cicero (Trapnell lab).

Identification of putative enhancers and their target genes



491,818 candidate *cis* regulatory elements (cCRE)

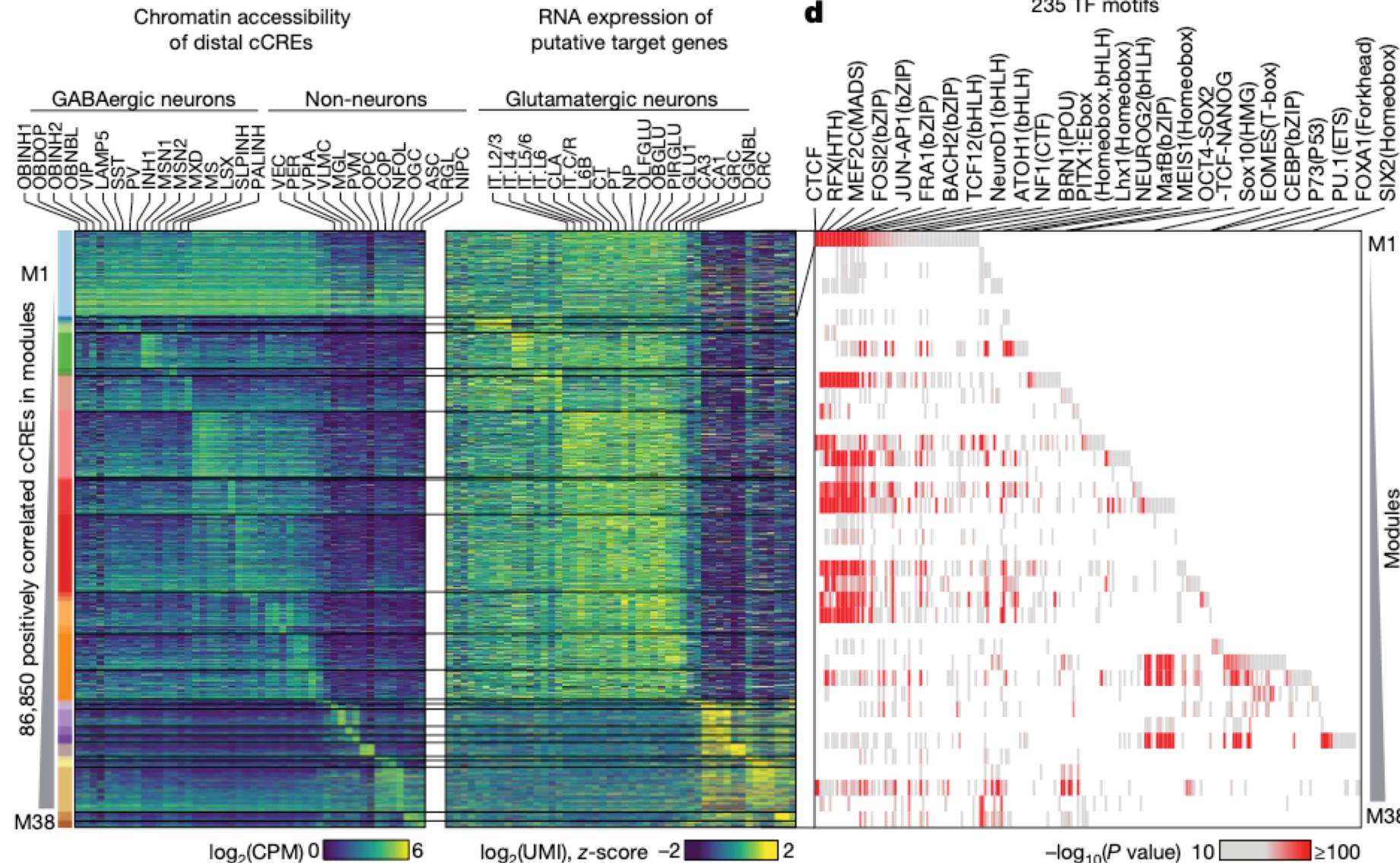
- 86,850 putative enhancers linked to 10,604 genes
- 7 enhancers per gene (average)
- 1-2 genes per enhancer
- Gene-enhancer distance: ~179 kbp (median)

cCRE: candidate *cis* regulatory element

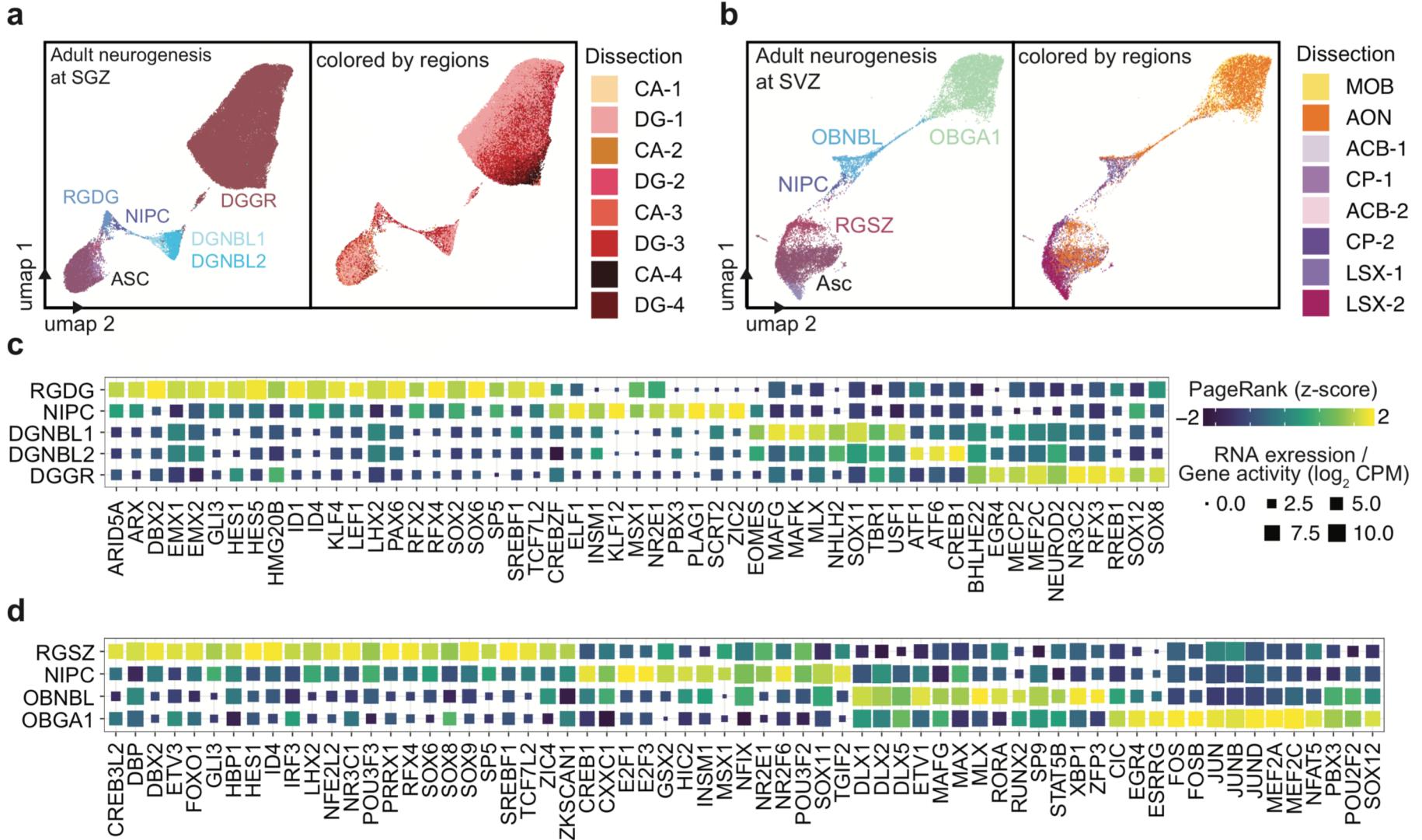
PCC: Pearson correlation coefficient

Co-accessible cCREs: identified by using Cicero (Trapnell lab).

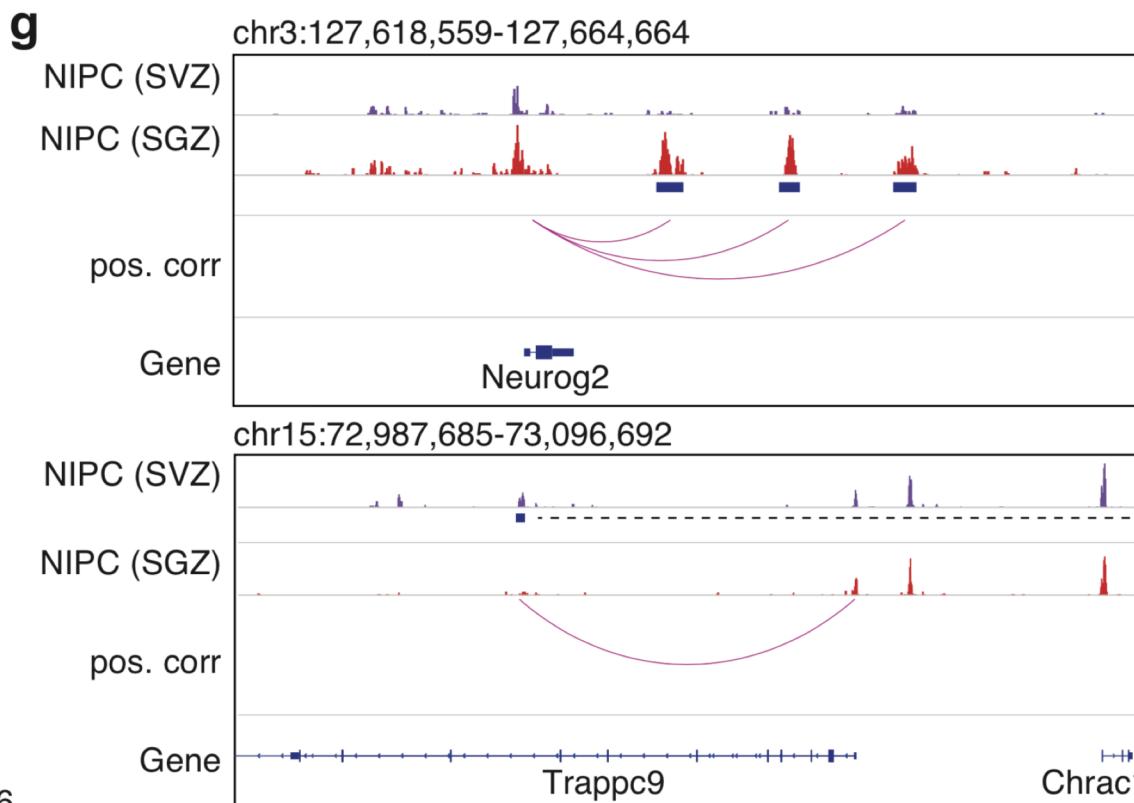
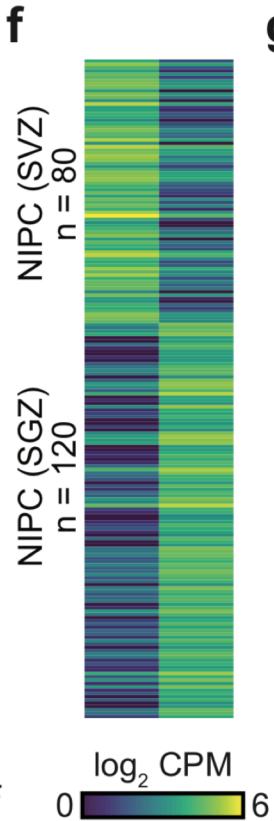
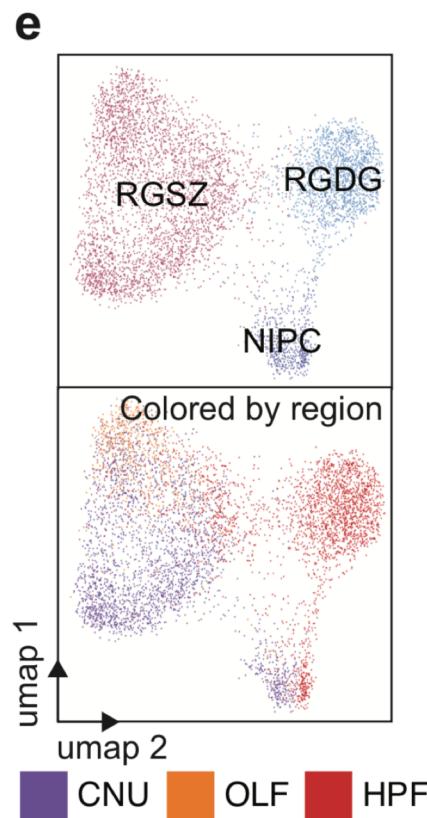
Putative enhancers reflect cell-type-specific gene expression patterns



Candidate TFs involved in adult neurogenesis in mouse brain

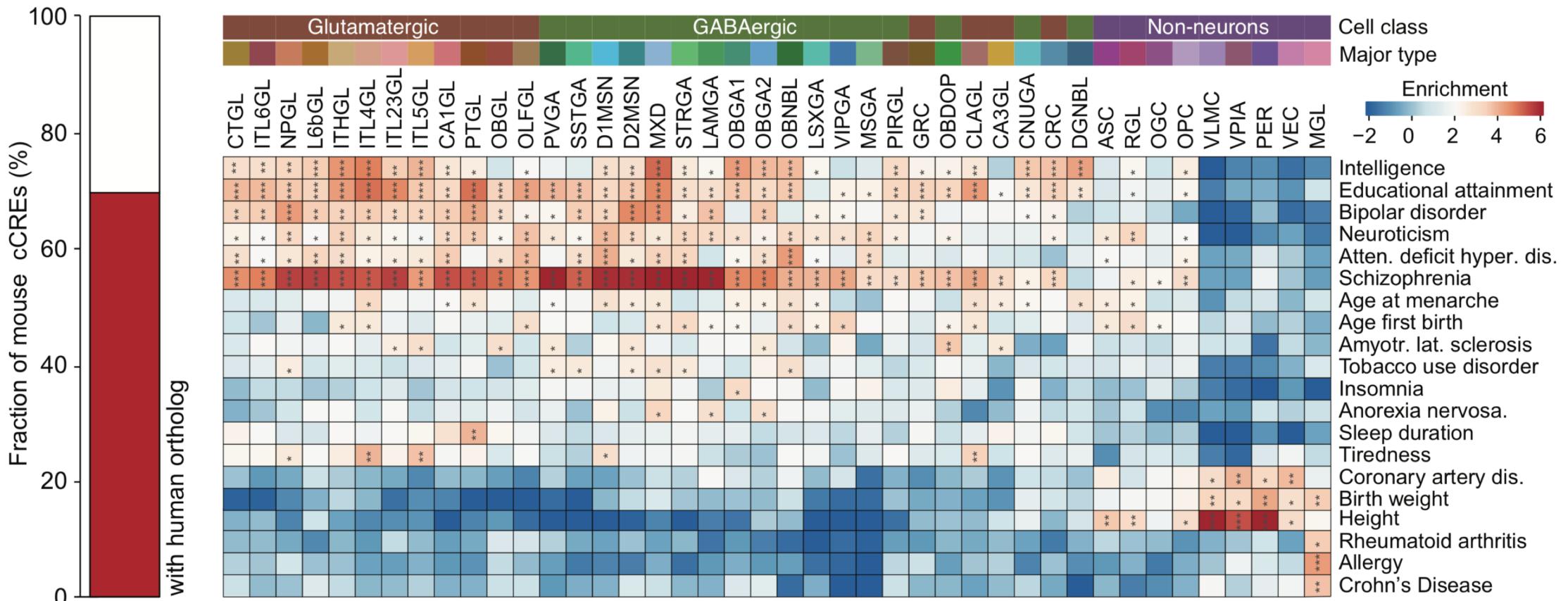


Candidate TFs involved in adult neurogenesis in mouse brain



Non-coding variants linked to Neurological Traits/Disorders are enriched in cCREs in distinct cerebral cell types

Fraction of mouse cCREs
mapped to the human genome
with high similarity >50%



Explore datasets using the CATlas web portal

The screenshot shows the CATlas web portal interface. At the top left is the CATlas logo with a cat silhouette and a DNA helix. Below the logo is a section titled "Mouse Brain" with the following details:
Sample: 45 brain sub-regions
Method: snATAC-seq
Nuclei count: > 800,000

On the left, there is a t-SNE plot (umap 1 vs umap 2) showing cell clusters, with labels for various regions like MGL, OGC, IOL, OPC, VPIA, VLMC, PER, VEC, ASC, RGL, GRC, etc. A legend indicates "All nuclei", "43 major clusters", "160 clusters", and "n = 813,799".

Below the t-SNE plot are three main sections: "Cell Clusters", "Cell Types", and "Explore Tracks".

- Cell Clusters:** Text: Explore cell clusters in different datasets. Last updated May, 2020. Button: Explore.
- Cell Types:** Text: Explore detailed information for every cell type. Button: Explore.
- Explore Tracks:** Text: Load signal tracks and compare between cell types/regions. Button: Explore.

A central panel displays a genomic tracks visualization (IGV) showing signal tracks across genomic regions for various cell types.

Article | Open Access | Published: 06 October 2021

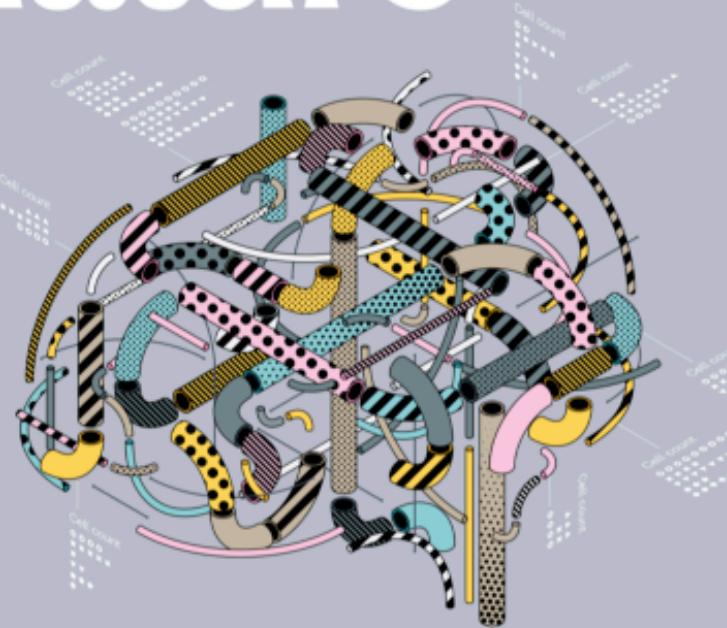
An atlas of gene regulatory elements in adult mouse cerebrum

Yang Eric Li, Sebastian Preissl, Xiaomeng Hou, Ziyang Zhang, Kai Zhang, Yunjiang Qiu, Olivier B. Poirion, Bin Li, Joshua Chiou, Hanqing Liu, Antonio Pinto-Duarte, Naoki Kubo, Xiaoyu Yang, Rongxin Fang, Xinxin Wang, Jee Yun Han, Jacinta Lucero, Yiming Yan, Michael Miller, Samantha Kuan, David Gorkin, Kyle J. Gaulton, Yin Shen, Michael Nunn, Eran A. Mukamel, M. Margarita Behrens, Joseph R. Ecker & Bing Ren [-Show fewer authors](#)

Nature 598, 129–136 (2021) | [Cite this article](#)

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

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



BRAIN CENSUS

A comprehensive cell atlas of the mammalian motor cortex

City heat
Urban governance must account for rising temperatures

Contrast adjustment
Plasmonic microscope slides add colour to tissue samples

Accessible science
Reflections on how to make research and results more inclusive

Vol. 598 No. 7410

Article | Open Access | Published: 06 October 2021

A multimodal cell census and atlas of the mammalian primary motor cortex

[BRAIN Initiative Cell Census Network \(BICCN\)](#)

Article | Open Access | Published: 06 October 2021

A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex

Zizhen Yao, Hanqing Liu, [...] Eran A. Mukamel [✉](#)

Article | Open Access | Published: 06 October 2021

Comparative cellular analysis of motor cortex in human, marmoset and mouse

Trygve E. Bakken [✉](#), Nikolas L. Jorstad, [...] Ed S. Lein [✉](#)

Article | Open Access | Published: 06 October 2021

DNA methylation atlas of the mouse brain at single-cell resolution

Hanqing Liu, Jingtian Zhou, [...] Joseph R. Ecker [✉](#)

Article | Open Access | Published: 06 October 2021

Epigenomic diversity of cortical projection neurons in the mouse brain

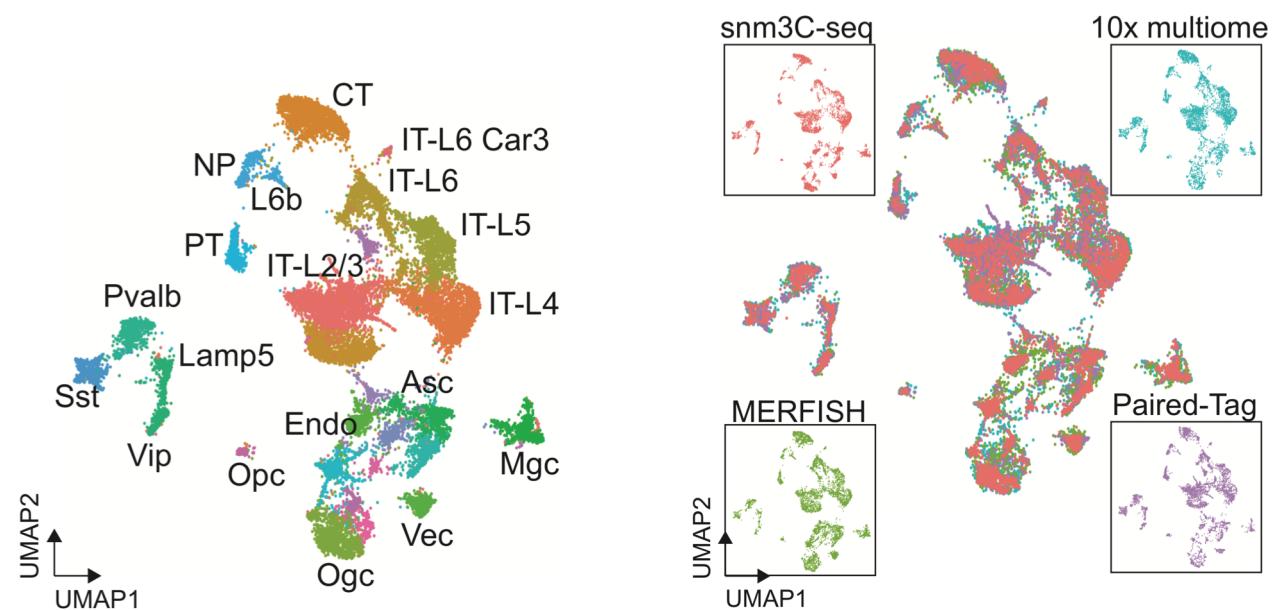
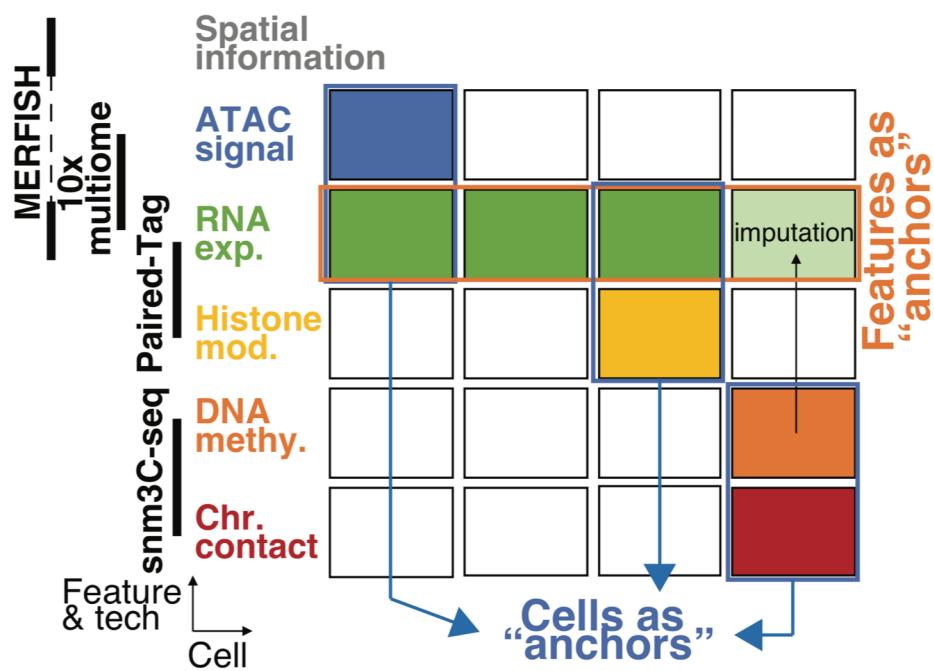
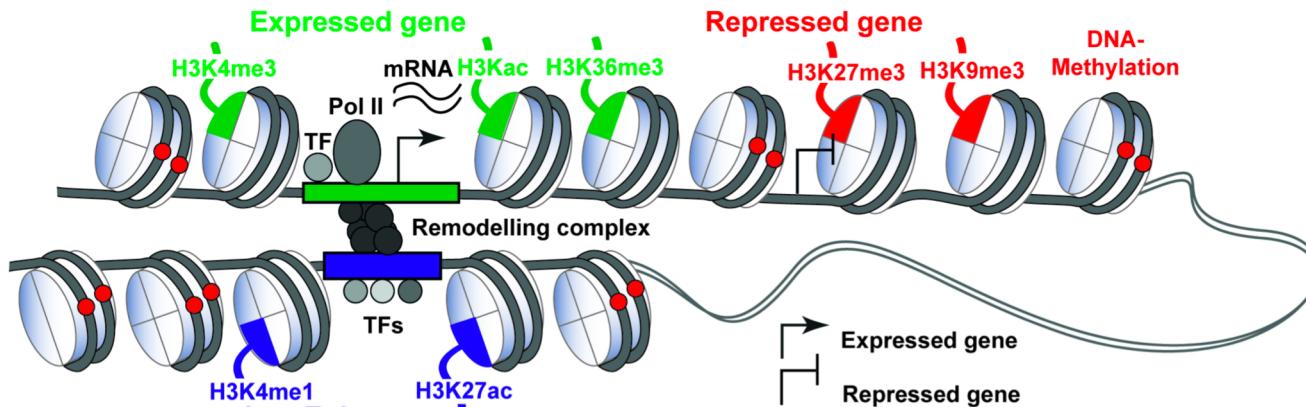
Zhuzhu Zhang, Jingtian Zhou, [...] Edward M. Callaway [✉](#)

Nature 598, 167–173 (2021) | [Cite this article](#)

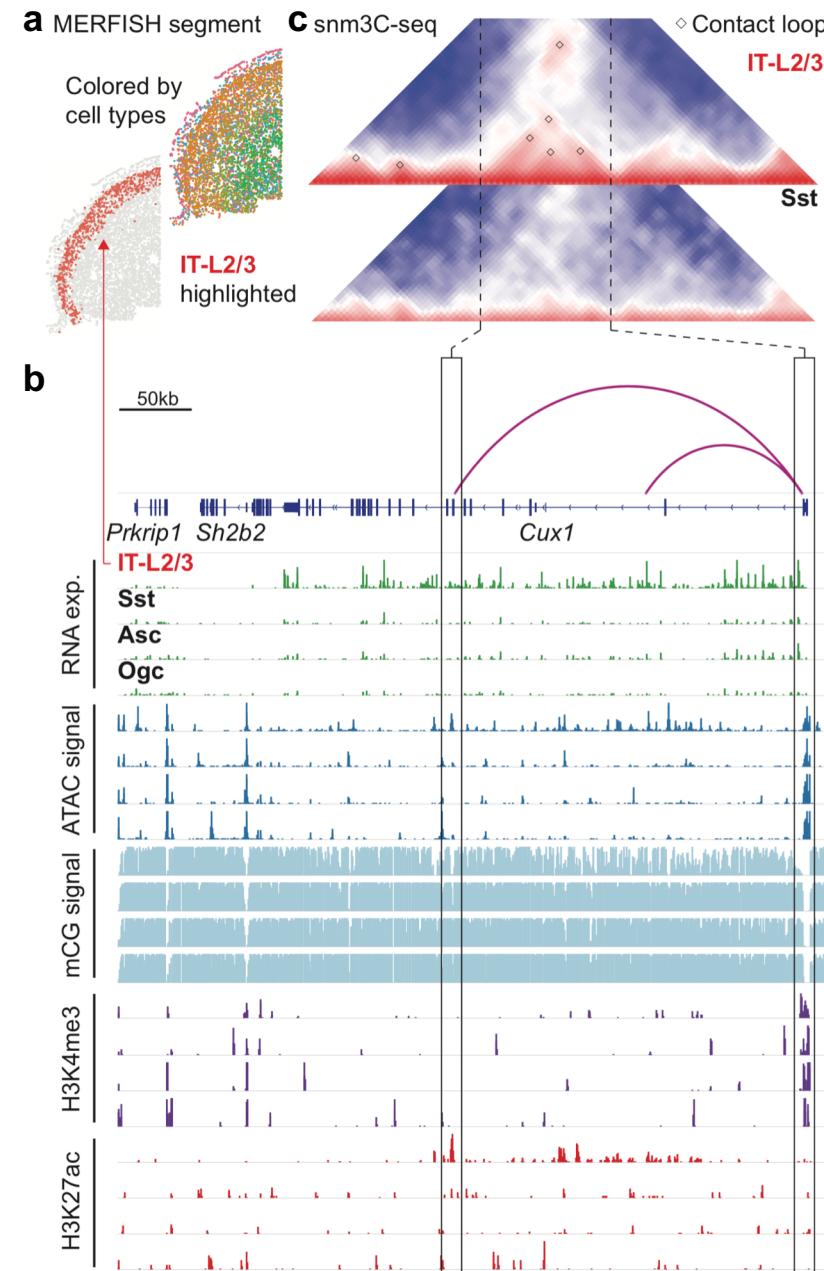
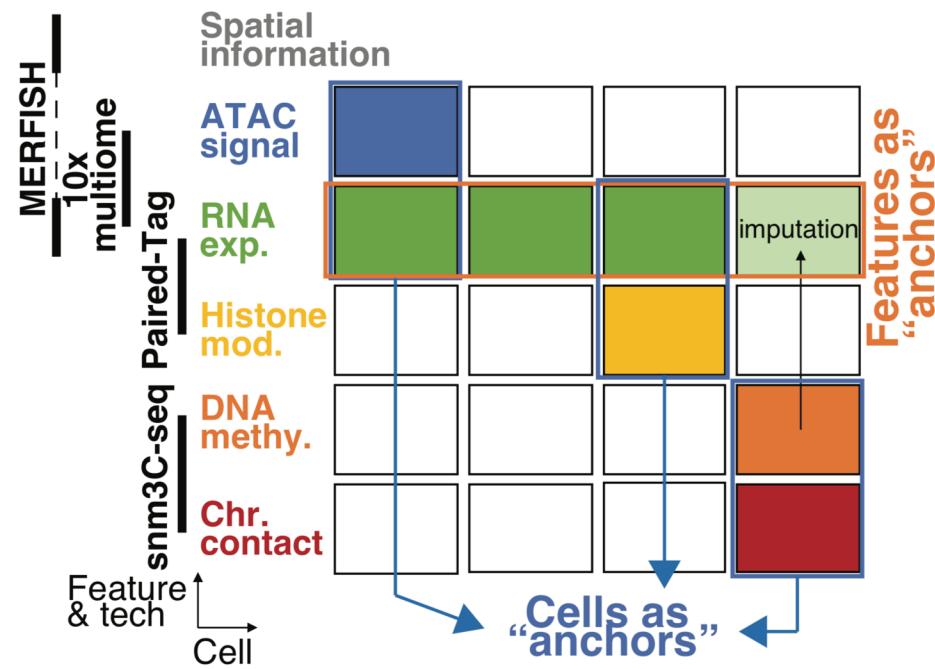
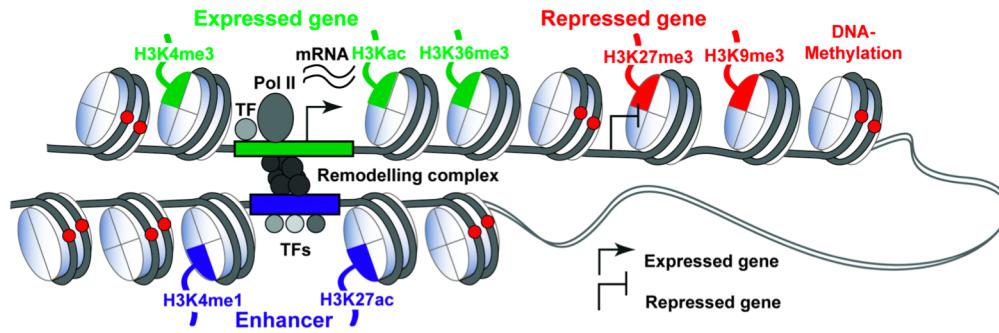
Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH

Meng Zhang, Stephen W. Eichhorn, Brian Zingg, Zizhen Yao, Kaelan Cotter, Hongkui Zeng, Hongwei Dong & Xiaowei Zhuang [✉](#)

Integrative analysis of single-cell multi-ome datasets



Integrative analysis of single-cell multi-ome datasets



Acknowledgment

Bing Ren Lab

Dr. Bing Ren
Ziyang Zhang
Bin Li
Kai Zhang
Yunjiang Qiu
Naoki Kubo
Rongxin Fang
James Hocker
Ramya Raviram
Yanxiao Zhang
Guoqiang Li

Other members

Center for epigenomics

Dr. Bing Ren
Dr. Sebastian Preissl
Dr. David Gorkin
Dr. Olivier Poirion
Xiaomeng Hou
Jee Yun Han
Xinxin Wang

Yin Shen lab

Xiaoyu Yang

Kyle Gaulton lab

Joshua Chiou

Margarita Behrens lab

Dr. Margarita Behrens
Jacinta Lucero
Antonio Pinto-Duarte

Joseph Ecker lab

Dr. Joe Ecker
Dr. Chongyuan Luo
Hanqing Liu
Jingtian Zhou

Eran Mukamel lab

Dr. Eran Mukamel

