Supplemental Material

The accessible chromatin landscape of the murine hippocampus at single-cell resolution

John R. Sinnamon[1,a,b], Kristof A. Torkenczy[2,a], Michael W. Linhoff[1], Sarah Vitak[2], Ryan M. Mulqueen[2], Hannah Pliner[3], Cole Trapnell[3], Frank J. Steemers[4], Gail Mandel[1], Andrew C. Adey[2,5,6,b]

1) The Vollum Institute, Portland, OR, USA
2) Oregon Health & Science University, Department of Molecular and Medical Genetics, Portland, OR, USA
3) University of Washington, Department of Genome Sciences, Seattle, WA, USA
4) Illumina, Inc., San Diego, CA, USA
5) Knight Cardiovascular Institute, Portland, OR, USA
6) Knight Cancer Institute, Portland, OR, USA
7) Knight Center for Early Detection Advanced Research, Oregon Health & Science University, Portland, OR, USA

[a] Denotes equal contribution
[b] To whom correspondence should be addressed (adey@ohsu.edu, sinnamon@ohsu.edu)
Lead contact: adey@ohsu.edu

Table of Contents

	This study	Preissl et al. 2018	Cusanovich et al. 2015
Single cell strategy	Combinatorial indexing	Combinatorial indexing	Combinatorial indexing
Species	Mouse	Human	Cell lines
Reads per cell/nucleus	Fresh: 5,021-97,083 (23,512 mean) Frozen: 3,455-151,534 (32,870 mean) In Vitro: 2,265-327,904 (43,532 mean)	9,375-18,397	1,390-3,094
Fraction of reads in peaks	Fresh: 25.0-57.4 % (30.8 % median) Frozen: 25.0-58.2 % (33.9 % median) In Vitro: 25.0-64.8 % (39.5 % median)	15.7-37.4%	41-59 %

	Buenrostro et al. 2015	Cusanovich et al. 2018	Cusanovich et al. 2018
Single cell strategy	Microfluidics	Combinatorial indexing	Combinatorial indexing
Species	Cell lines	Drosophila	Mouse
Reads per cell/nucleus	73,000 (mean)	7,181-8,024	8,743-23,456
Fraction of reads in peaks	> 15%	n.r.	36%

Supplemental Table 1: Data comparison with other single-cell ATAC-seq studies
Comparison of QC metrics for our study broken down by sample and other single-cell ATAC-seq publications based on the reporting by those publications.

Cluster	Tissue	Cells	Percent (of tissue)
Noise	Frozen	5	n/a
	Fresh	10	n/a
Granule Cells	Frozen	238	29.86%
	Fresh	499	34.13%
Interneurons	Frozen	66	8.28%
	Fresh	61	4.17%
Neurons1	Frozen	104	29.86%
	Fresh	169	11.56%
Neurons2	Frozen	251	31.49%
	Fresh	496	33.93%
Collisions	Frozen	25	n/a
	Fresh	47	n/a
Astrocytes	Frozen	33	4.14%
	Fresh	89	6.09%
Microglia	Frozen	25	3.14%
	Fresh	18	1.23%
Oligodendrocytes	Frozen	60	7.53%
	Fresh	91	6.22%
OPCs	Frozen	20	2.51%
	Fresh	39	2.67%

	Total Frozen	827	35.25%
	Total Fresh	1519	64.75%

Supplemental Table 2: Cell Type Composition
Proportion of cells assigned to each cluster for the *In Vivo* dataset along with the fresh vs. frozen breakdown.

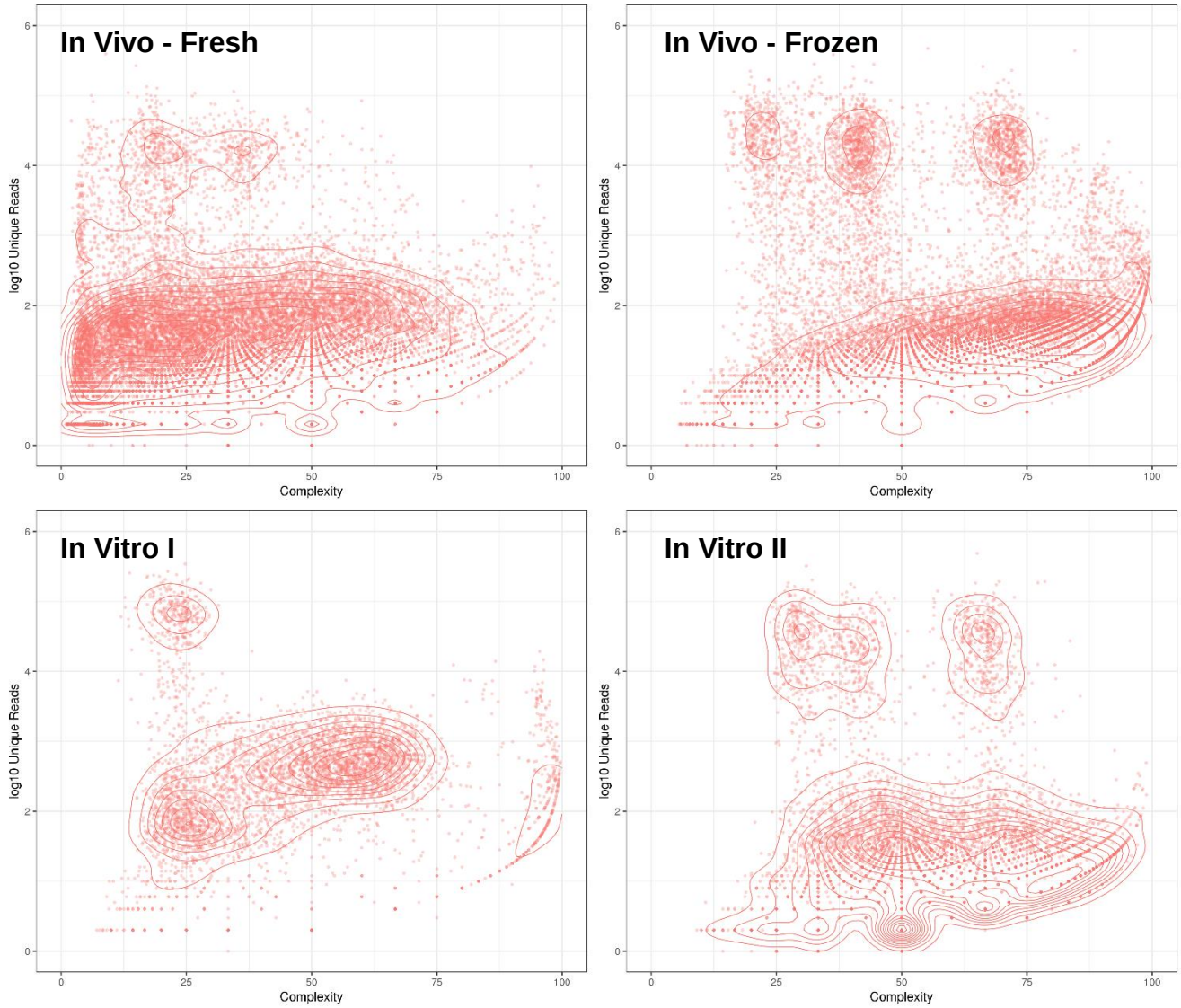
	Chrom.	Start	End
Granule Cells	chr2	12,805,030	12,805,581
	chr18	13,157,877	13,158,511
	chr19	24,665,688	24,666,194
	chr8	24,303,072	24,303,600
	chr14	11,231,939	11,232,458
	chr1	190,306,260	190,306,838
	chr1	190,642,627	190,643,179
	chr5	88,830,139	88,830,687
	chr1	55,828,229	55,828,748
	chr10	48,010,760	48,012,230
Interneurons	chr2	158,610,479	158,611,273
	chr2	71,529,369	71,529,901
	chr11	88,028,257	88,028,867
	chr2	22,621,845	22,622,763
	chr6	55,398,876	55,399,381
	chr5	124,206,256	124,206,918
	chr5	31,589,699	31,590,705
	chr2	36,011,161	36,011,981
	chr5	35,985,506	35,986,014
	chr19	15,052,684	15,053,503
Neurons 1 (CA1)	chr18	15,008,128	15,008,636
	chr9	22,030,329	22,030,831
	chr8	40,455,798	40,456,317
	chr7	84,576,178	84,576,680
	chr13	73,912,658	73,913,254
	chr2	46,330,878	46,331,408
	chrX	145,341,895	145,342,781
	chr6	89,050,828	89,051,367
	chr10	54,551,634	54,552,521
	chr3	74,734,781	74,735,299

	Chrom.	Start	End
Neurons 2 (CA3)	chr19	23,330,204	23,331,209
	chr2	169,049,869	169,050,512
	chr15	94,011,340	94,012,637
	chr19	18,368,387	18,369,139
	chr9	97,883,149	97,883,765
	chr6	55,836,835	55,837,355
	chr18	78,243,805	78,244,339
	chr3	65,884,483	65,885,067
	chr10	111,308,849	111,309,852
	chr1	127,541,771	127,542,290
Astrocytes	chr3	50,421,595	50,422,393
	chr10	18,359,980	18,360,766
	chr2	102,620,556	102,621,235
	chr14	65,889,584	65,890,606
	chr14	121,113,651	121,114,302
	chr1	154,139,439	154,140,079
	chr16	18,083,199	18,083,957
	chr1	182,483,531	182,484,443
	chr5	33,705,059	33,705,607
	chr12	25,239,950	25,241,190
Microglia	chr7	67,445,504	67,446,168
	chr7	126,746,073	126,747,014
	chr15	27,702,515	27,703,332
	chr7	80,078,688	80,079,367
	chr15	59,728,185	59,728,766
	chr7	145,111,444	145,112,140
	chr4	129,782,344	129,783,082
	chr13	83,571,978	83,572,659
	chr17	88,581,421	88,581,940
	chr7	120,851,508	120,852,400

	Chrom.	Start	End
Oligodendrocytes	chr10	83,906,604	83,907,477
	chr12	100,073,051	100,073,764
	chr18	75,544,174	75,544,677
	chr13	56,581,977	56,583,211
	chr18	75,541,370	75,541,889
	chr9	25,230,456	25,231,009
	chr17	8,350,898	8,351,641
	chr2	127,627,327	127,628,521
	chr14	34,590,699	34,591,715
	chr9	25,240,906	25,241,521
Oligo. Progenitors	chr5	65,434,477	65,435,230
	chr2	125,723,709	125,725,049
	chr9	56,868,193	56,868,694
	chr6	30,572,647	30,573,201
	chr7	79,729,178	79,729,709
	chr14	57,534,993	57,535,499
	chr4	106,317,513	106,318,014
	chr3	51,749,422	51,750,382
	chr1	172,064,528	172,065,243
	chr6	146,955,683	146,956,269

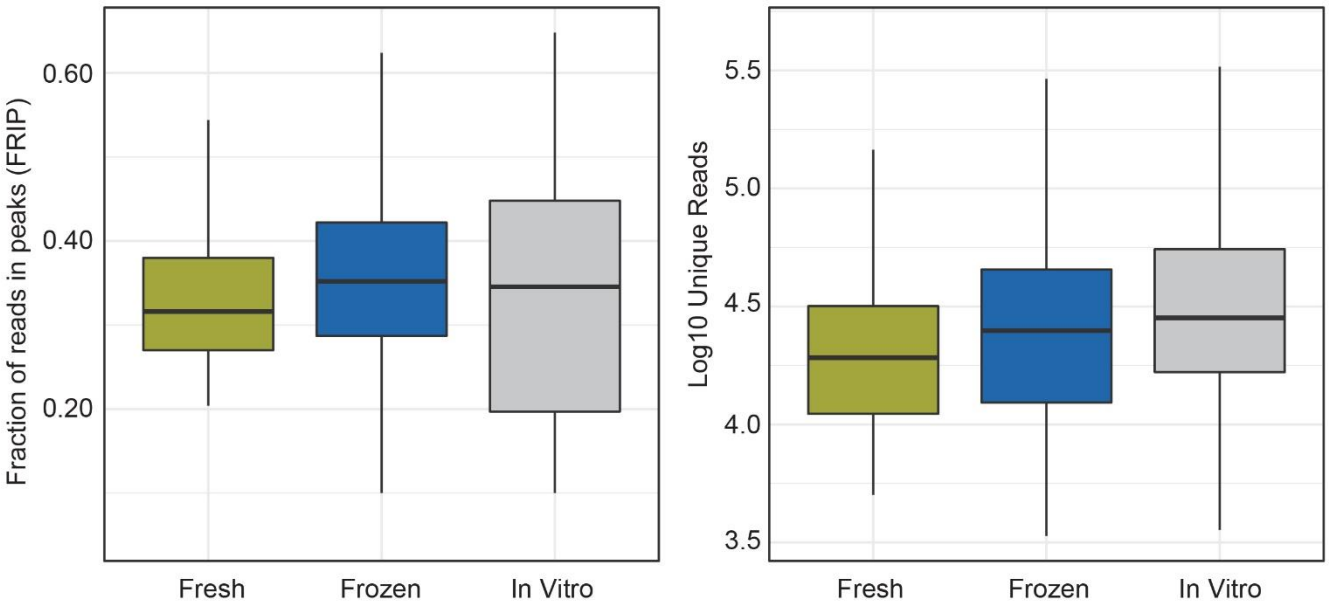
Supplemental Table 3: Differentially Accessible Peaks

The top 10 differentially accessible (see Methods) peaks corresponding to each cluster. Plots for these can be found in Supplemental Fig. 4.



Supplemental Figure 1: Library complexity

The Log10 total unique reads per cell (y-axis) is plotted with respect to the fraction of unique reads per cell (x-axis) for the four sets of libraries. Populations of true cells are the clusters of points above 1,000 unique reads with the remaining points comprising barcode combinations that do not contain cells. For plots with multiple populations of cells, the different populations represent different sets of PCR wells that were pooled together as opposed to biological replicates.



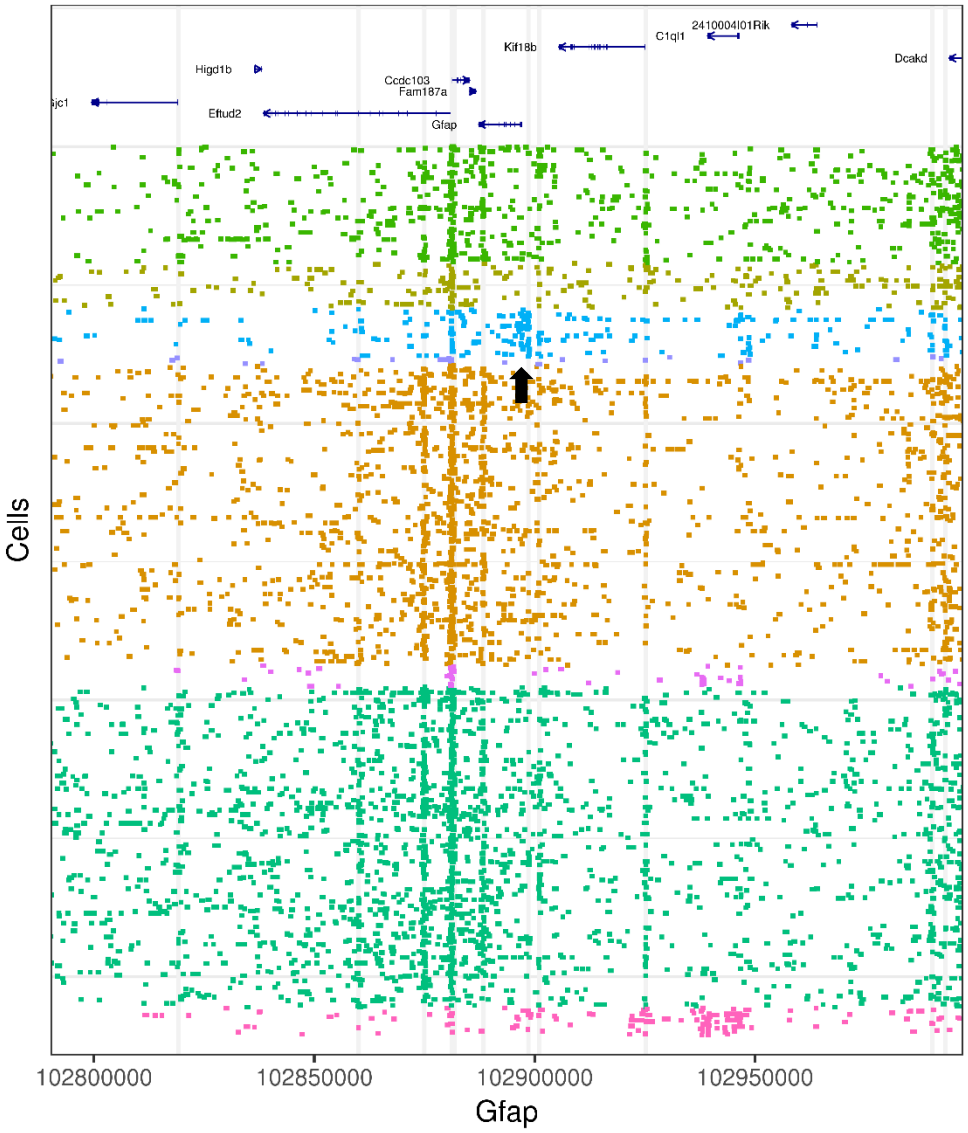
Supplemental Figure 2: QC between fresh, frozen, and in vitro libraries
Comparison between the fraction of reads in peaks (right, FRIP, pre- 25% minimum filter) and the total unique read counts (left) for the fresh, frozen, and in vitro preparations.

Supplemental Figure 3: Marker Gene Dashboards

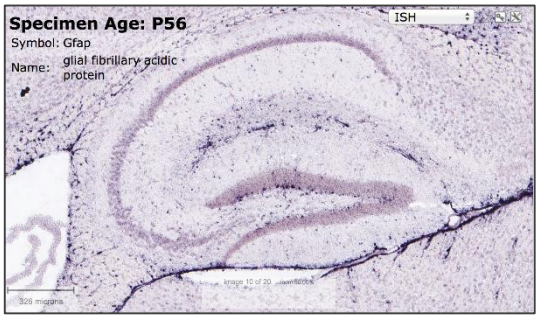
Marker gene dashboards (A-N) contain three plots and additional information. For each dashboard the gene name can be found in the upper right with the specified cell type below in parentheses. The literature reference for why the gene was selected is included below the gene title. The top left plot is a ‘read plot’ of the region around the gene (+/- 100,000 bp) with genes in the region plotted at the top followed by rows for each cell with unique reads in the region colored by the cluster identity. To the right of the read plot is an *in situ* hybridization image from the Allen Brain Atlas for the specified gene. Lastly, the bottom panel is a genome browser view showing the aggregated cluster sci-ATAC-seq profiles at a zoomed in region around the marker gene along with mouse hippocampus H3K4me3 and H3K27ac ChiP-seq peaks obtained from the ENCODE project. Black arrows on the read plot and genome viewer screenshot indicate cluster-specific signal and the corresponding cluster in the legend of the read plot.

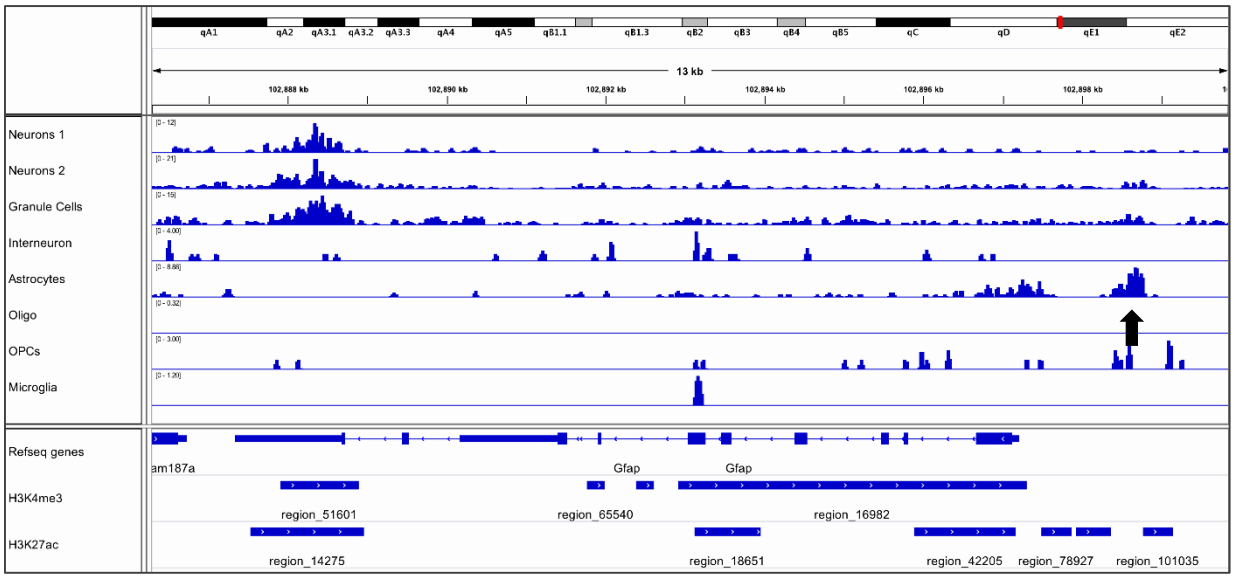
A) *Gfap*
(Astrocytes)

PMID: 7952264
Brain Pathol. 1994 Jul;4(3):229-37.



AST
GRN
INT
MRG
NR1
NR2
OLI
OPC

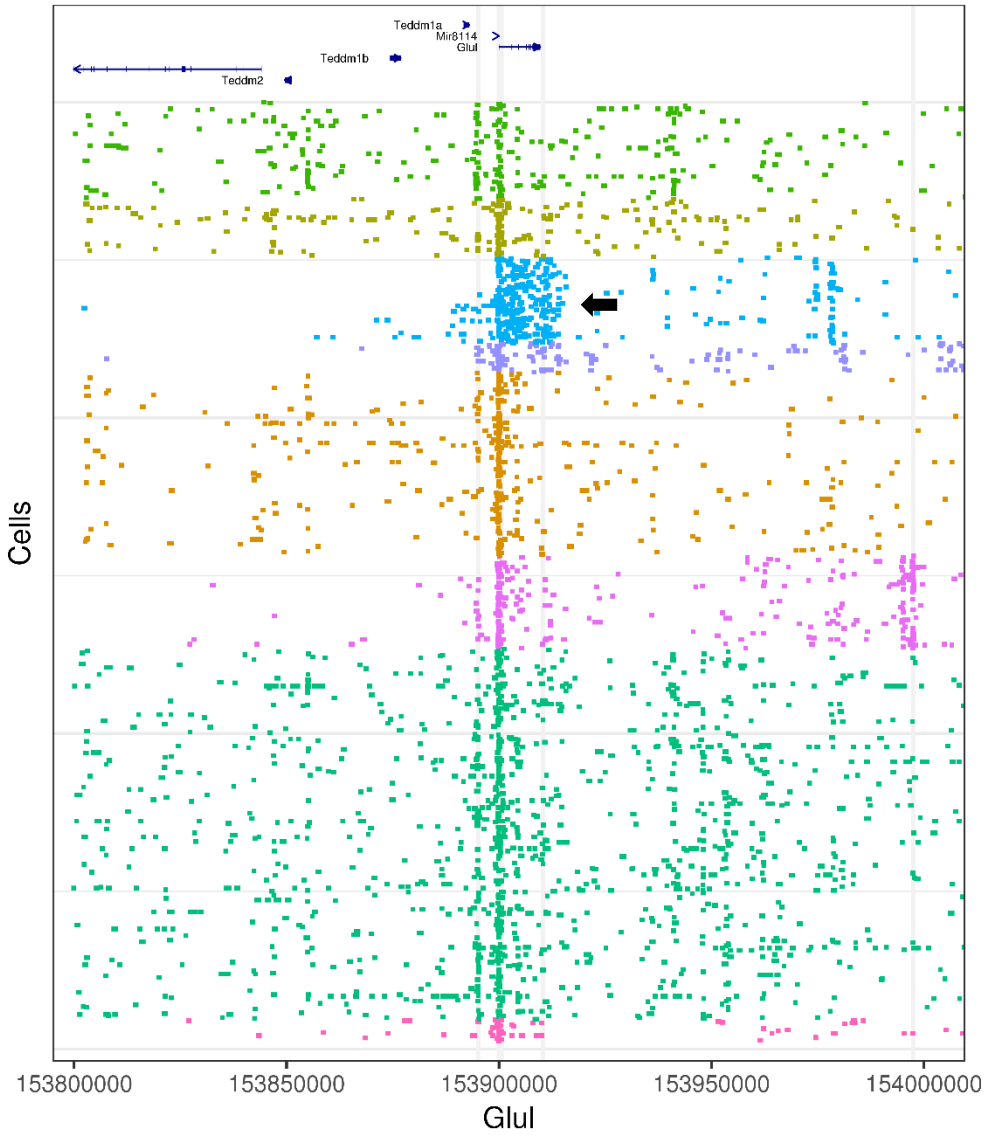




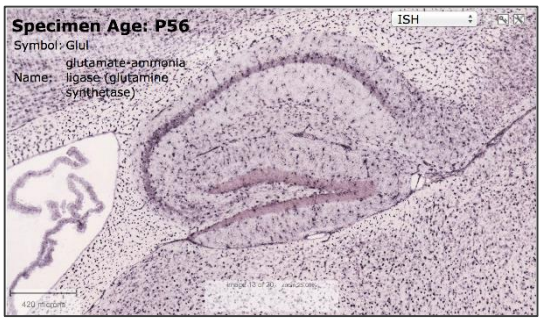
B) *Glul*

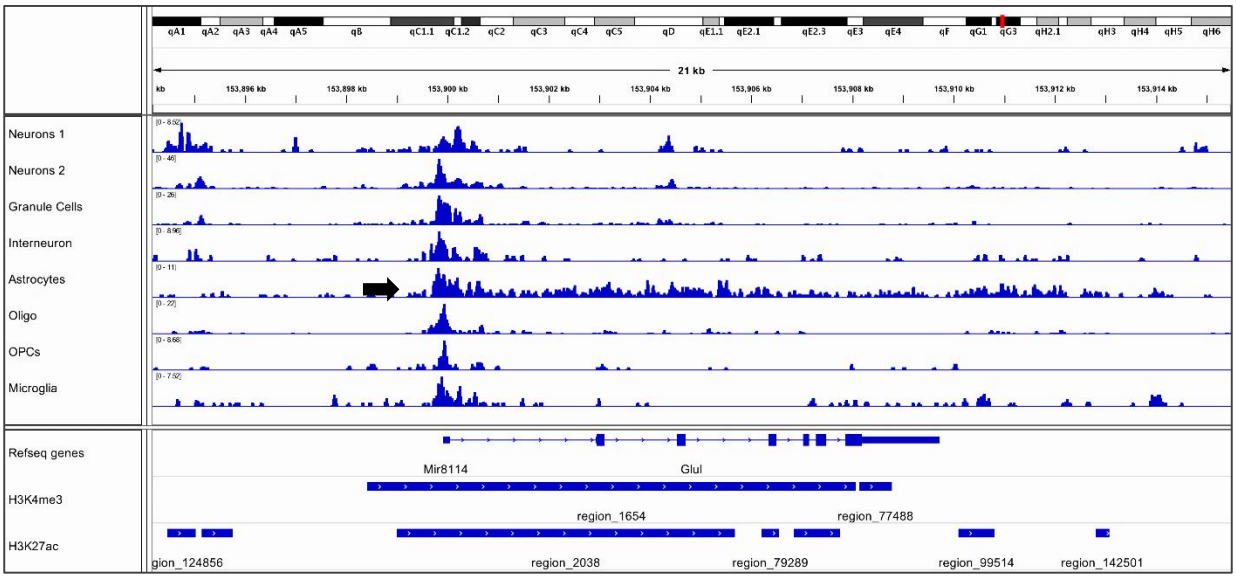
(Astrocytes)

PMID: 14400
Science. 1977 Mar 25;195(4284):1356-8
PMID: 24137157
Front Endocrinol (Lausanne). 2013 Oct 16;4:144.



- AST
- GRN
- INT
- MRG
- NR1
- NR2
- OLI
- OPC

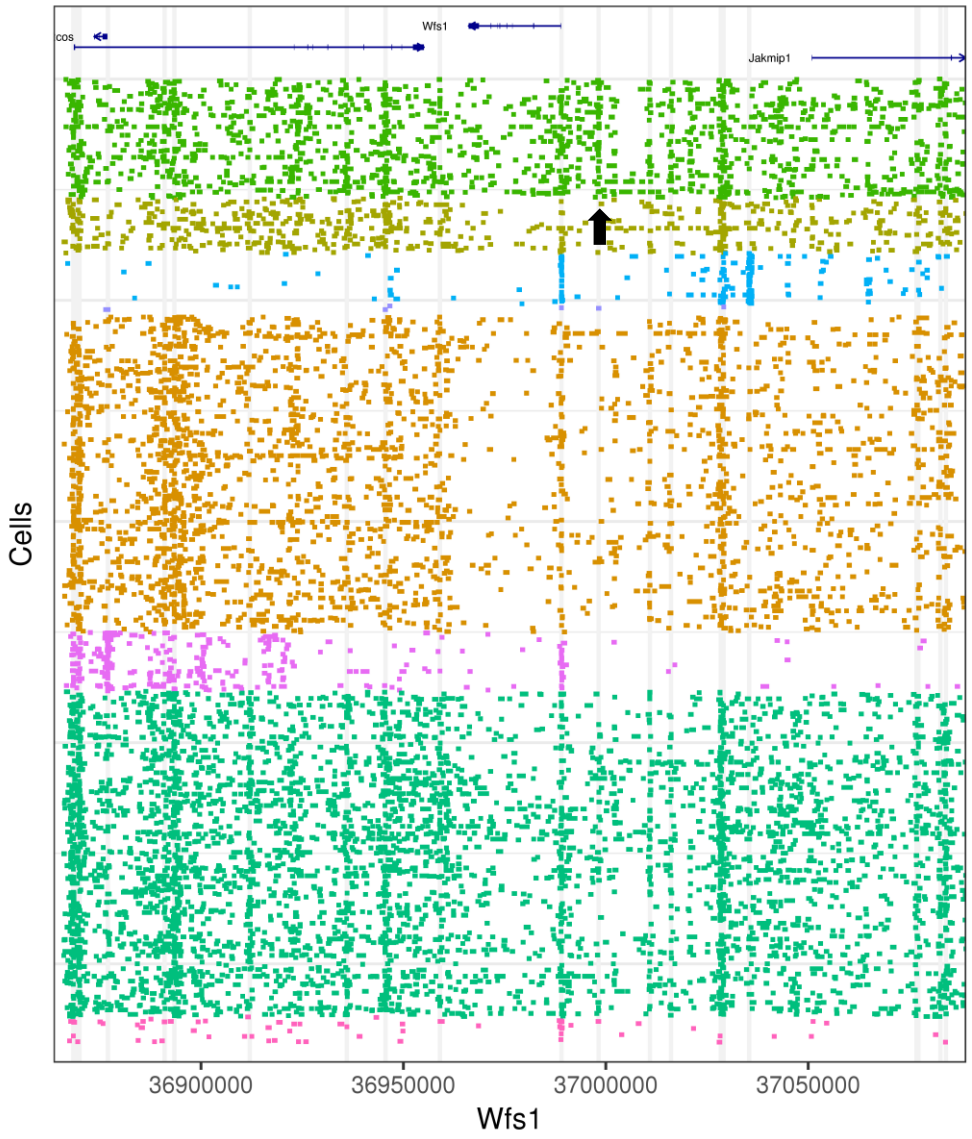




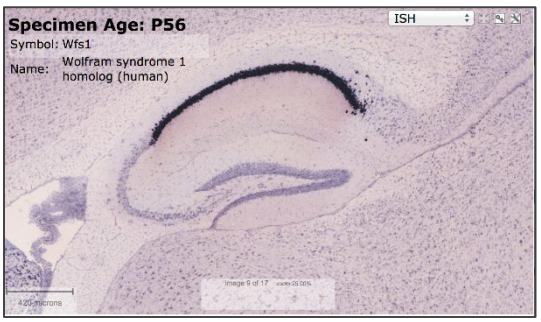
C) *Wfs1*

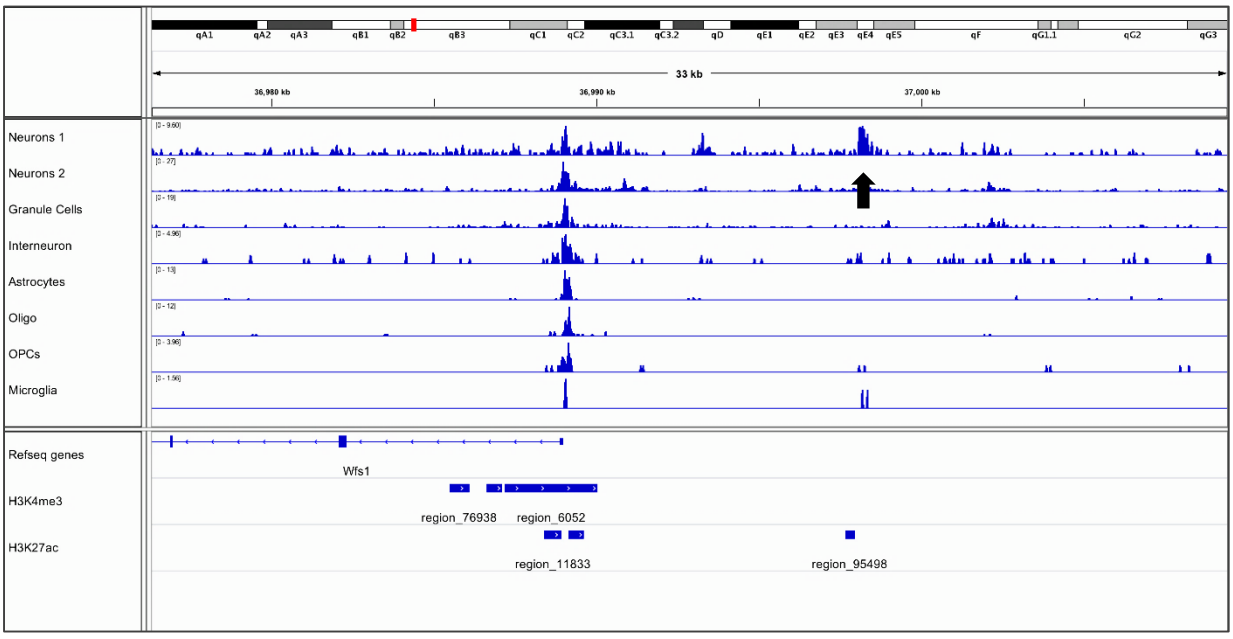
(CA1 Neurons)

PMID: 24336151
Nat Neurosci. 2014 Feb;17(2):269-79. doi: 10.1038/nn.3614



- AST
- GRN
- INT
- MRG
- NR1
- NR2
- OLI
- OPC

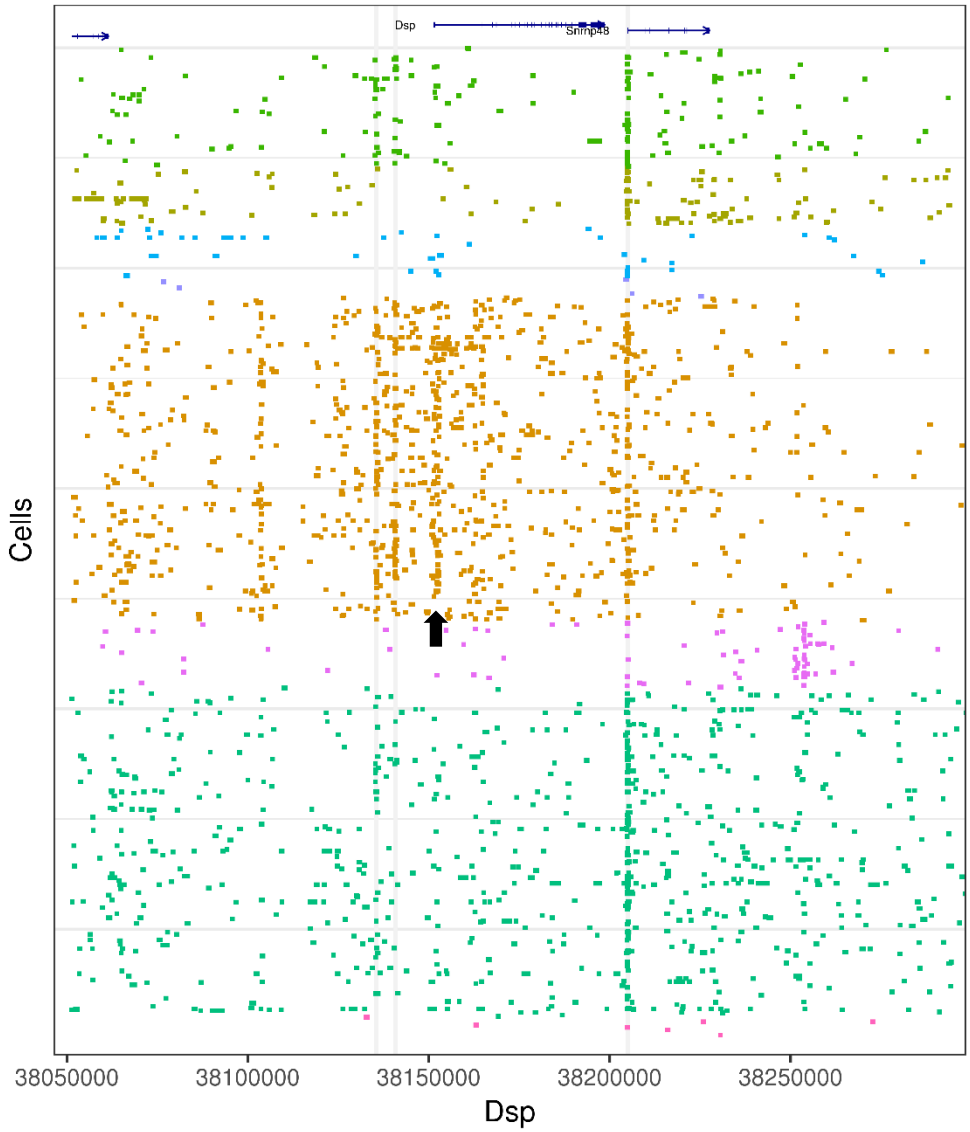




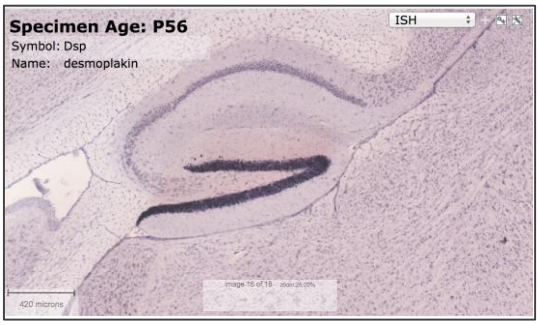
D) *Dsp*

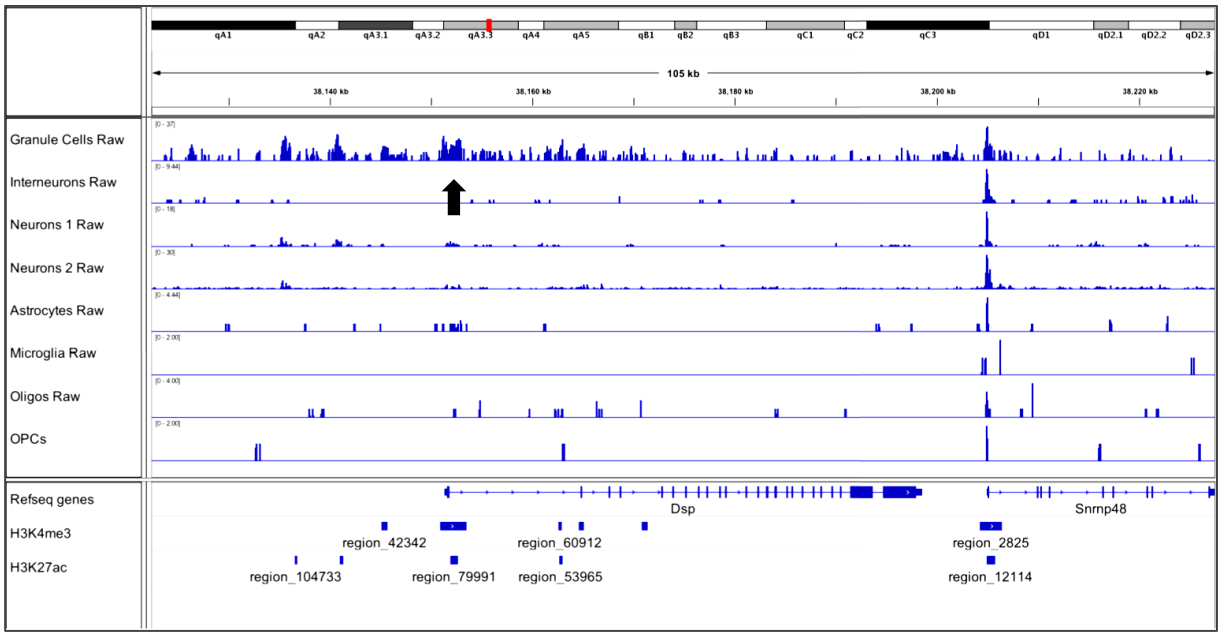
(Dentate Granule Cells)

PMID: 15084669
DOI: 10.1523/JNEUROSCI.4710-03.2004



- AST
- GRN
- INT
- MRG
- NR1
- NR2
- OLI
- OPC

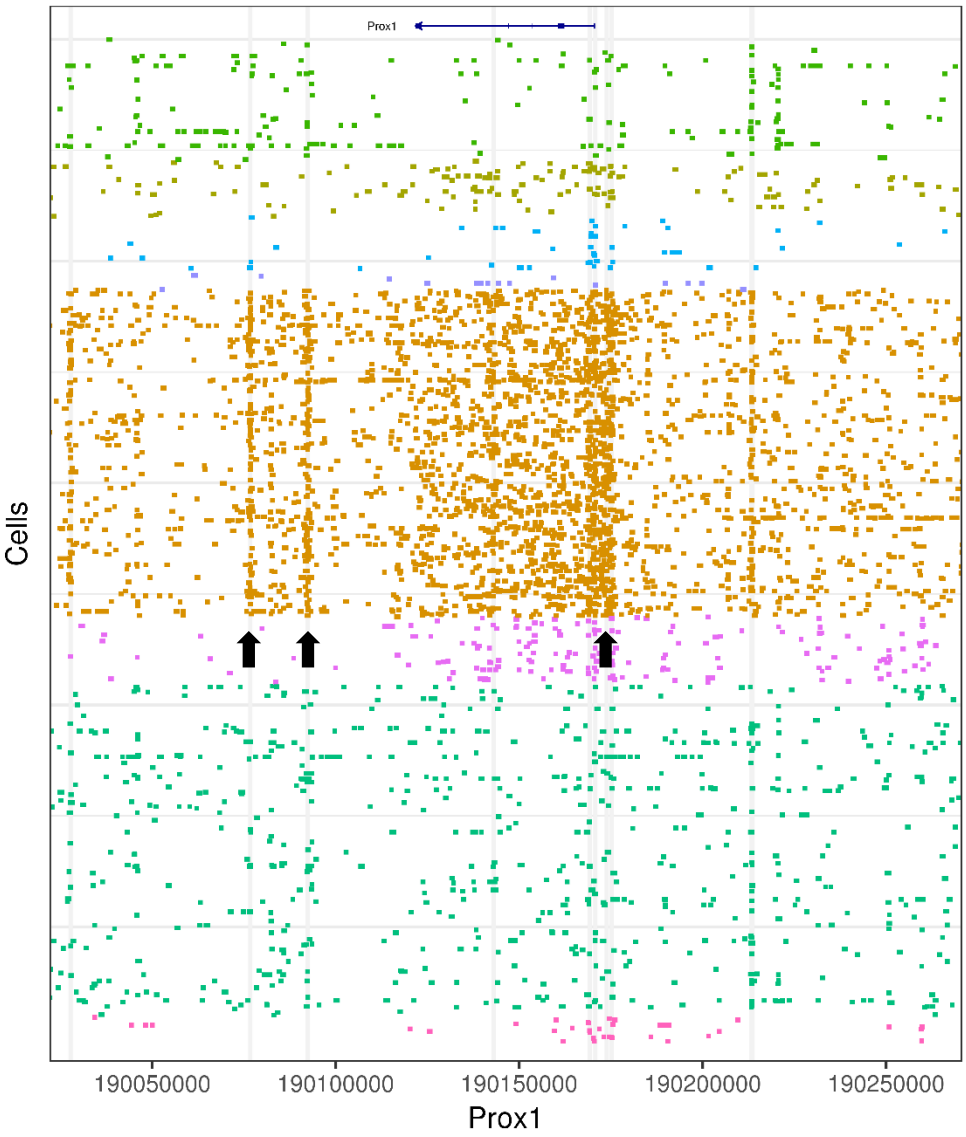




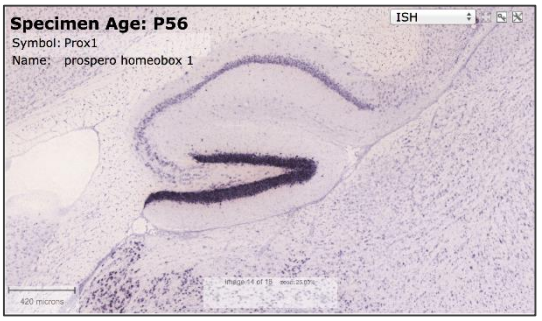
E) *Prox1*

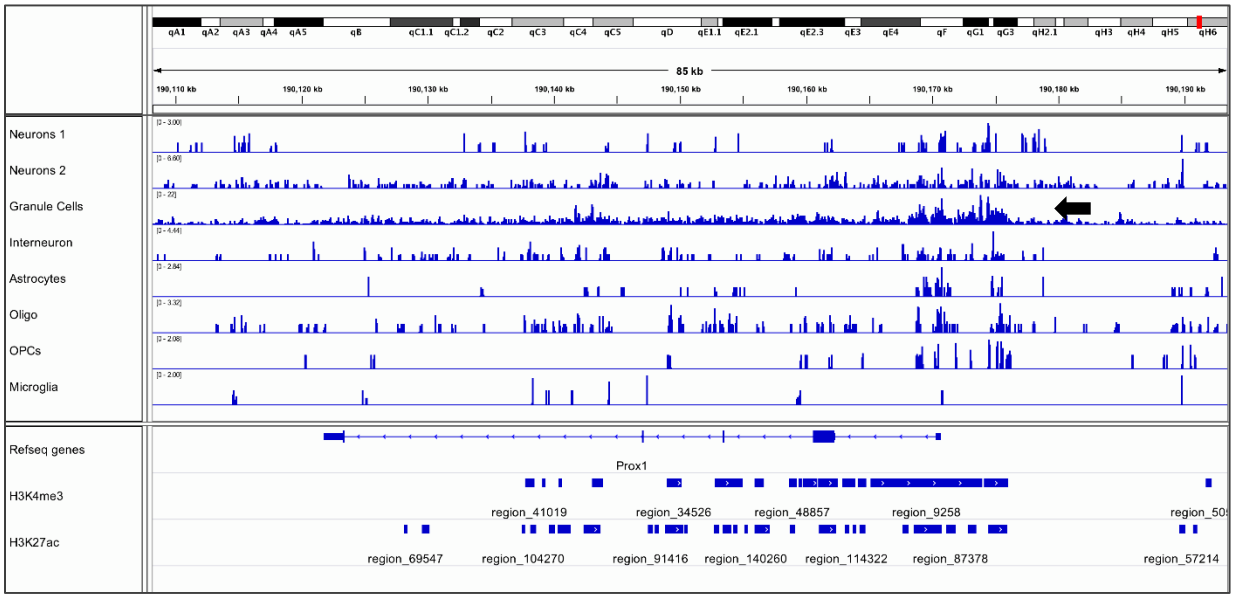
(Dentate Granule Cells)

PMID: 22791897
Development. 2012 Aug;139(16):3051-62. doi: 10.1242/dev.080002



- AST
- GRN
- INT
- MRG
- NR1
- NR2
- OLI
- OPC

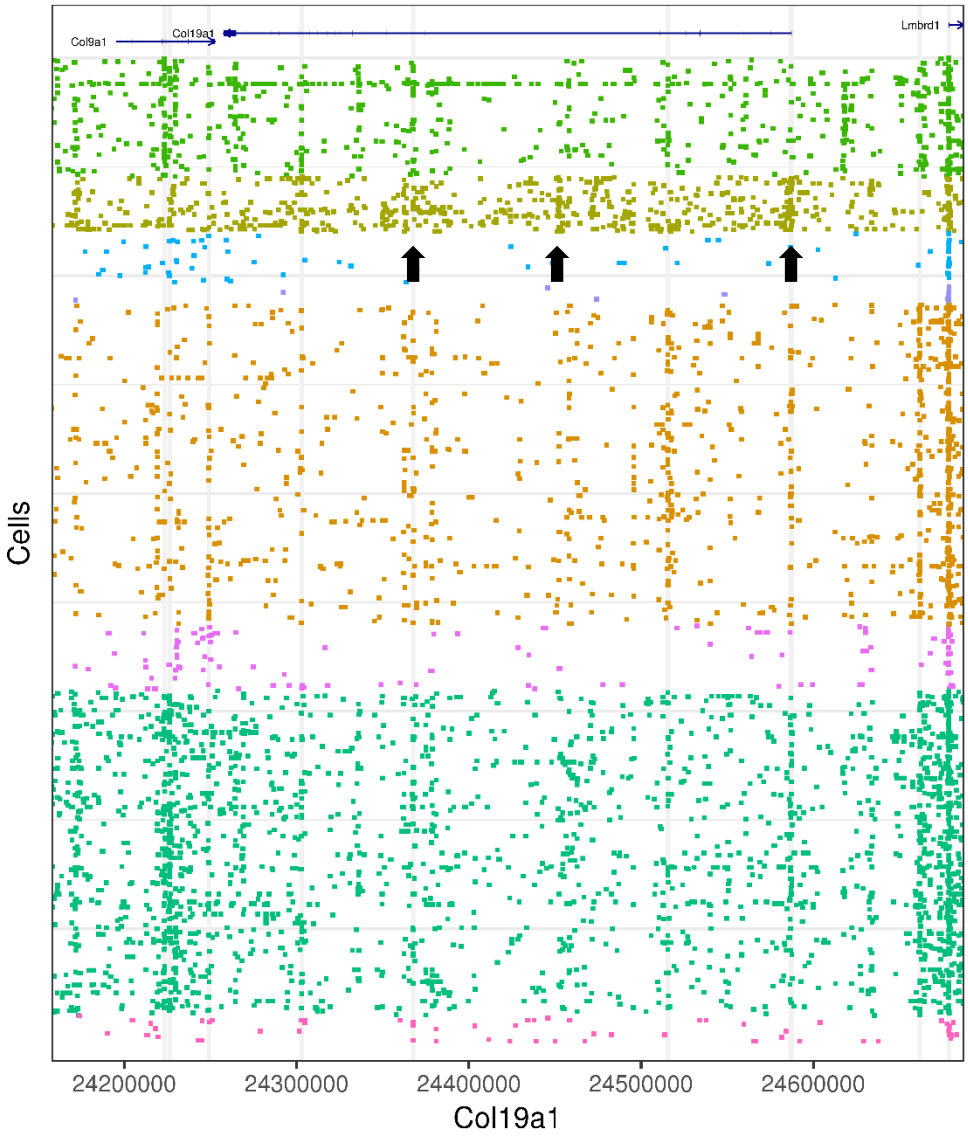




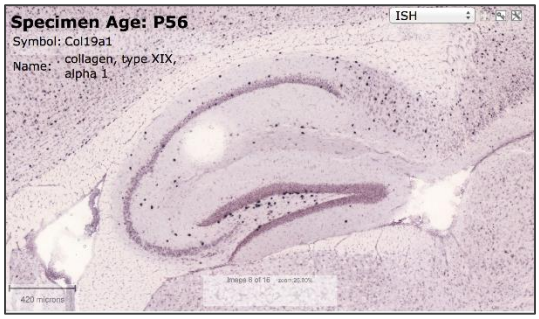
F) *Col19a1*

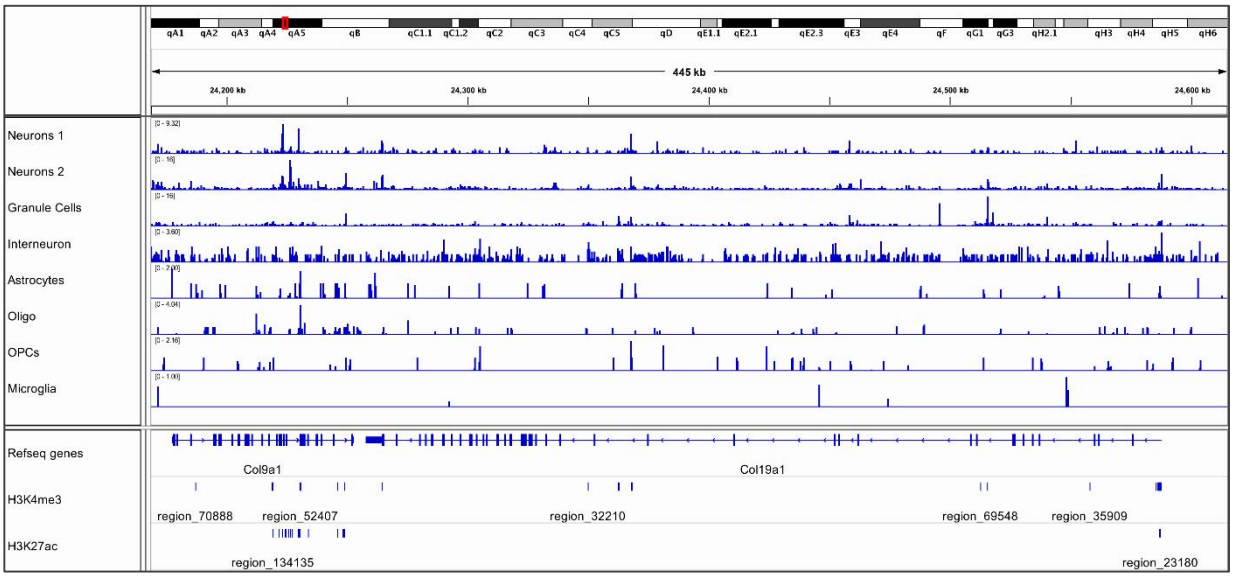
(Interneurons)

PMID: 19937713
J Comp Neurol. 2010 Jan 10;518(2):229-53



- AST
- GRN
- INT
- MRG
- NR1
- NR2
- OLI
- OPC

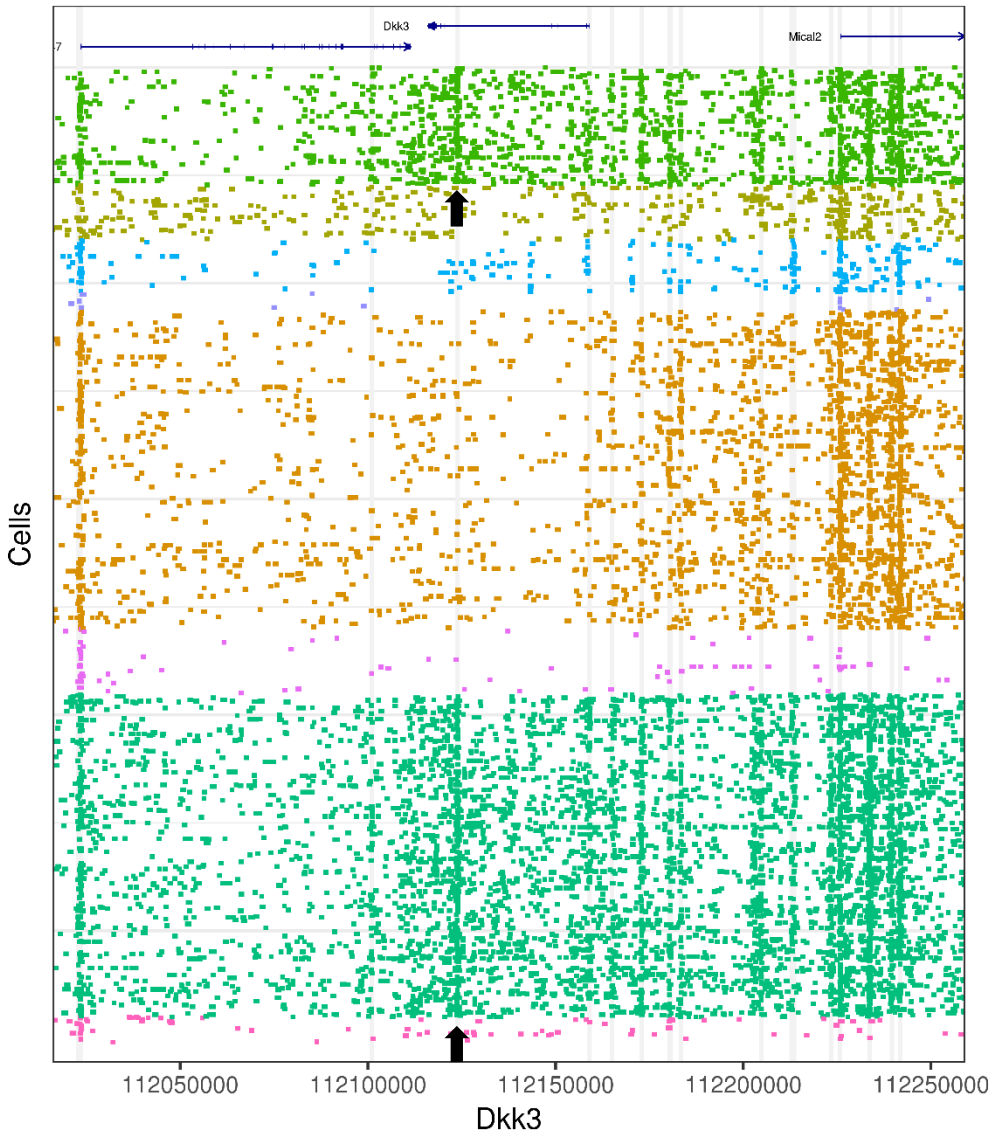




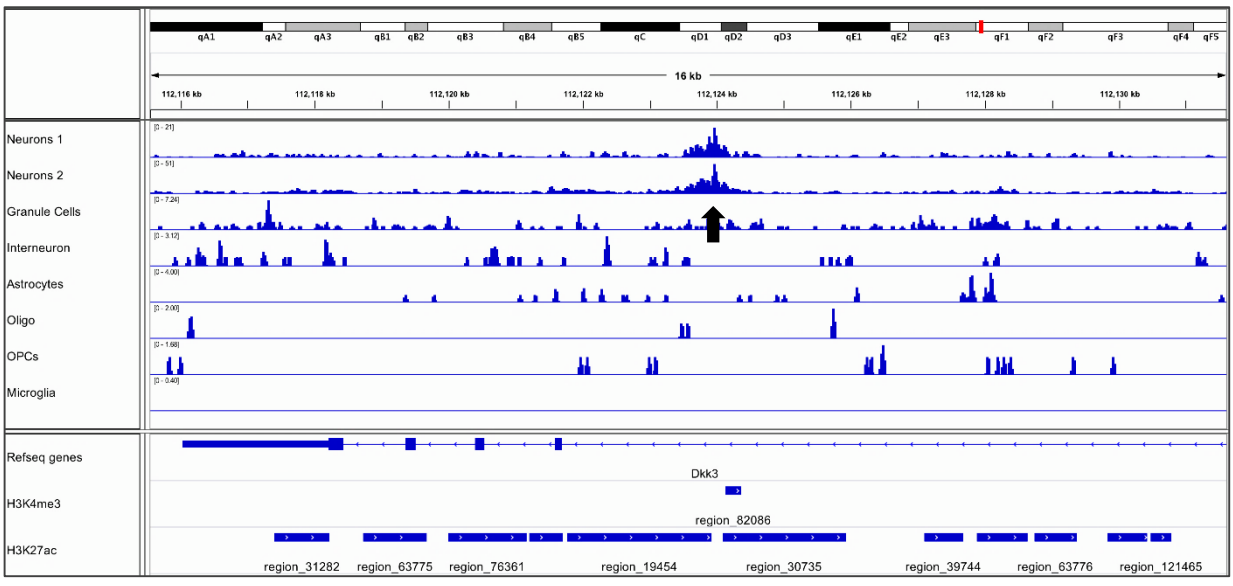
G) *Dkk3*

(Pyramidal Neurons)

PMID: 27113915
Elife. 2016 Apr 26;5:e14997.
doi: 10.7554/eLife.14997



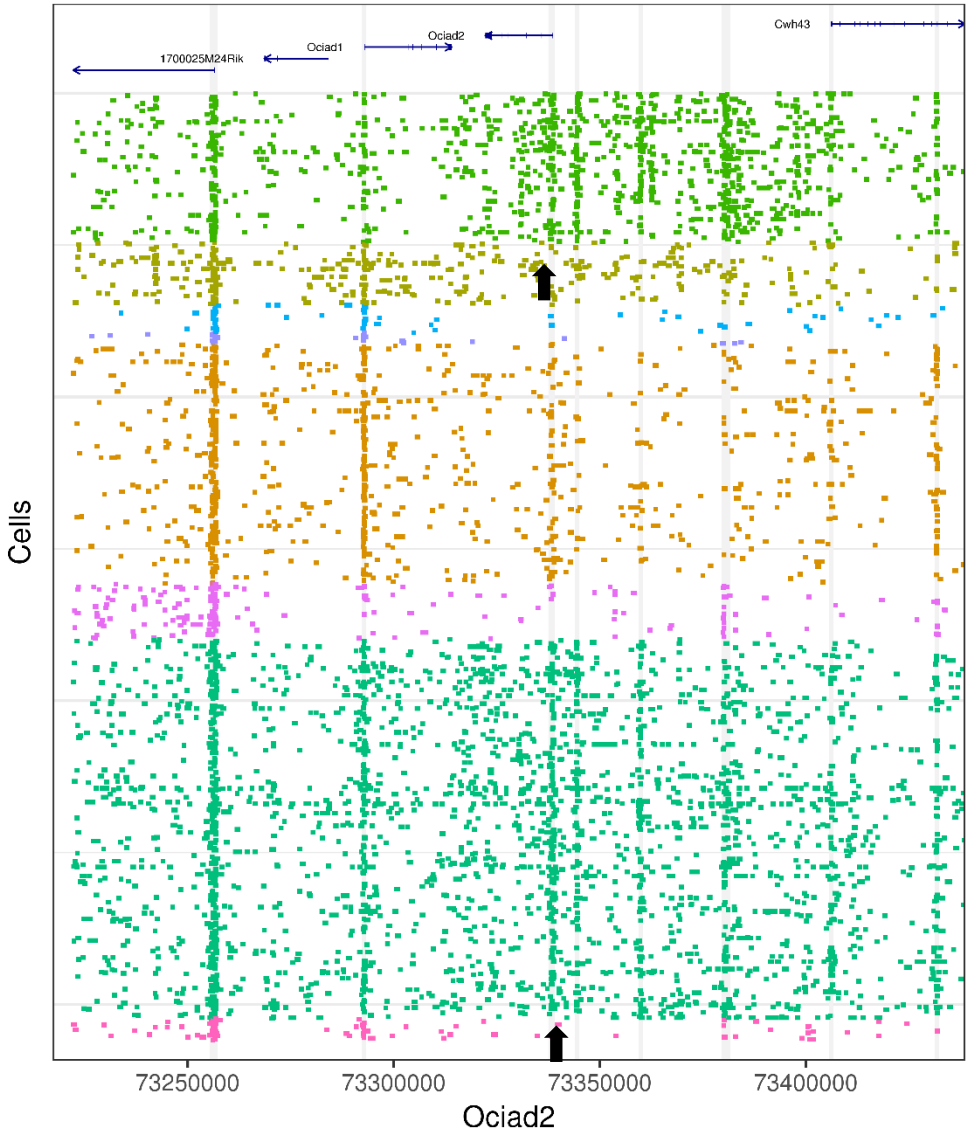
- AST
- GRN
- INT
- MRG
- NR1
- NR2
- OLI
- OPC



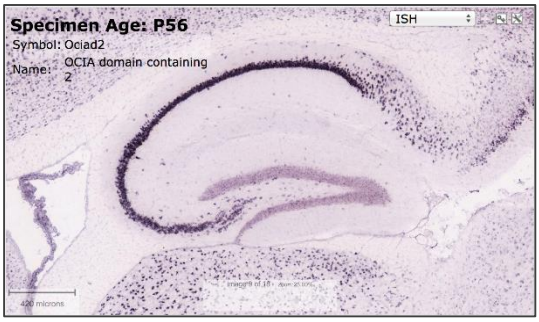
H) *Ociad2*

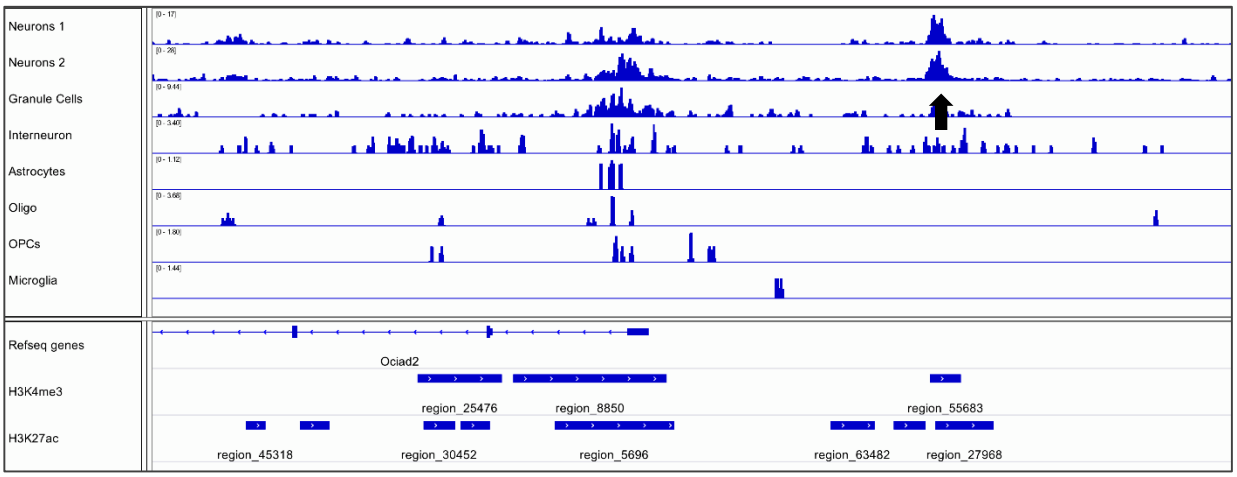
(Pyramidal Neurons)

PMID: 27113915
Elife. 2016 Apr 26;5:e14997.
doi: 10.7554/eLife.14997



- AST
- GRN
- INT
- MRG
- NR1 ⬅
- NR2 ⬅
- OLI
- OPC

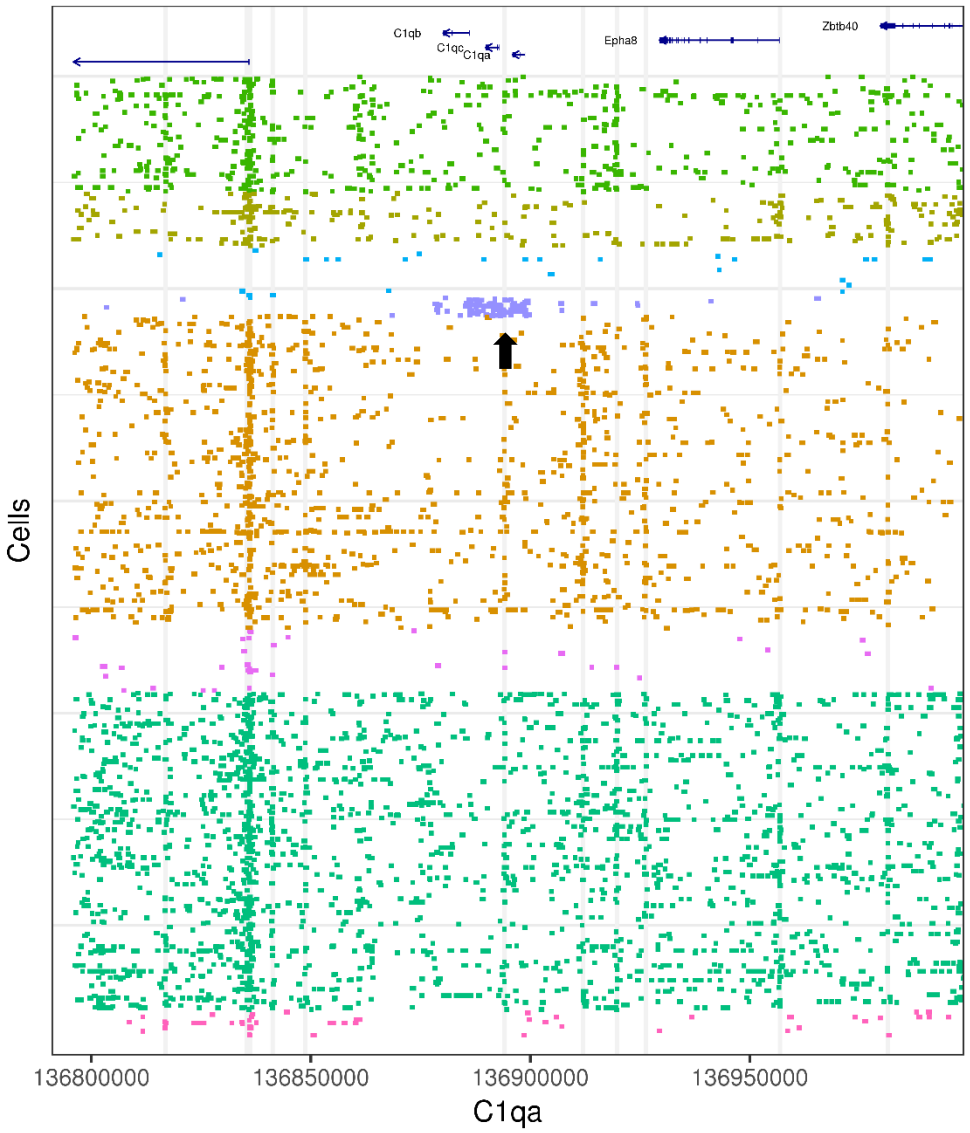




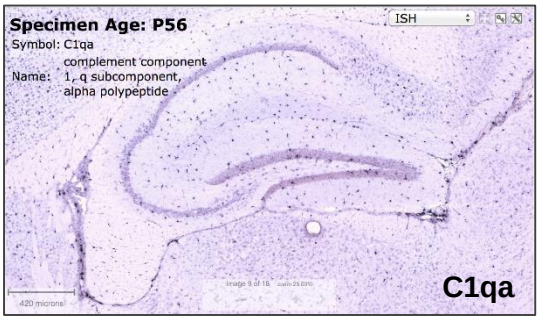
I) *C1qa/c*

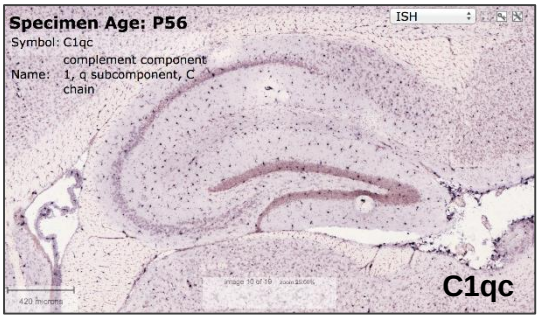
(Microglia)

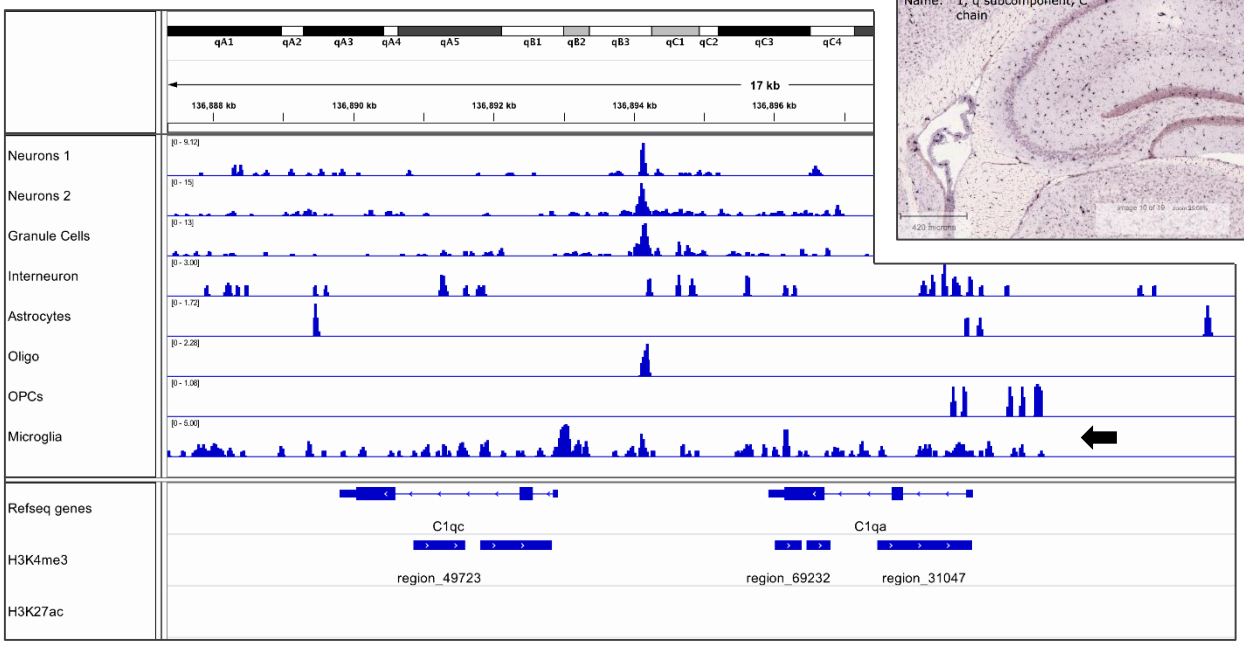
PMID: 26982357
Immunity. 2016 Mar 15;44(3):505-515



- AST
- GRN
- INT
- MRG
- NR1
- NR2
- OLI
- OPC

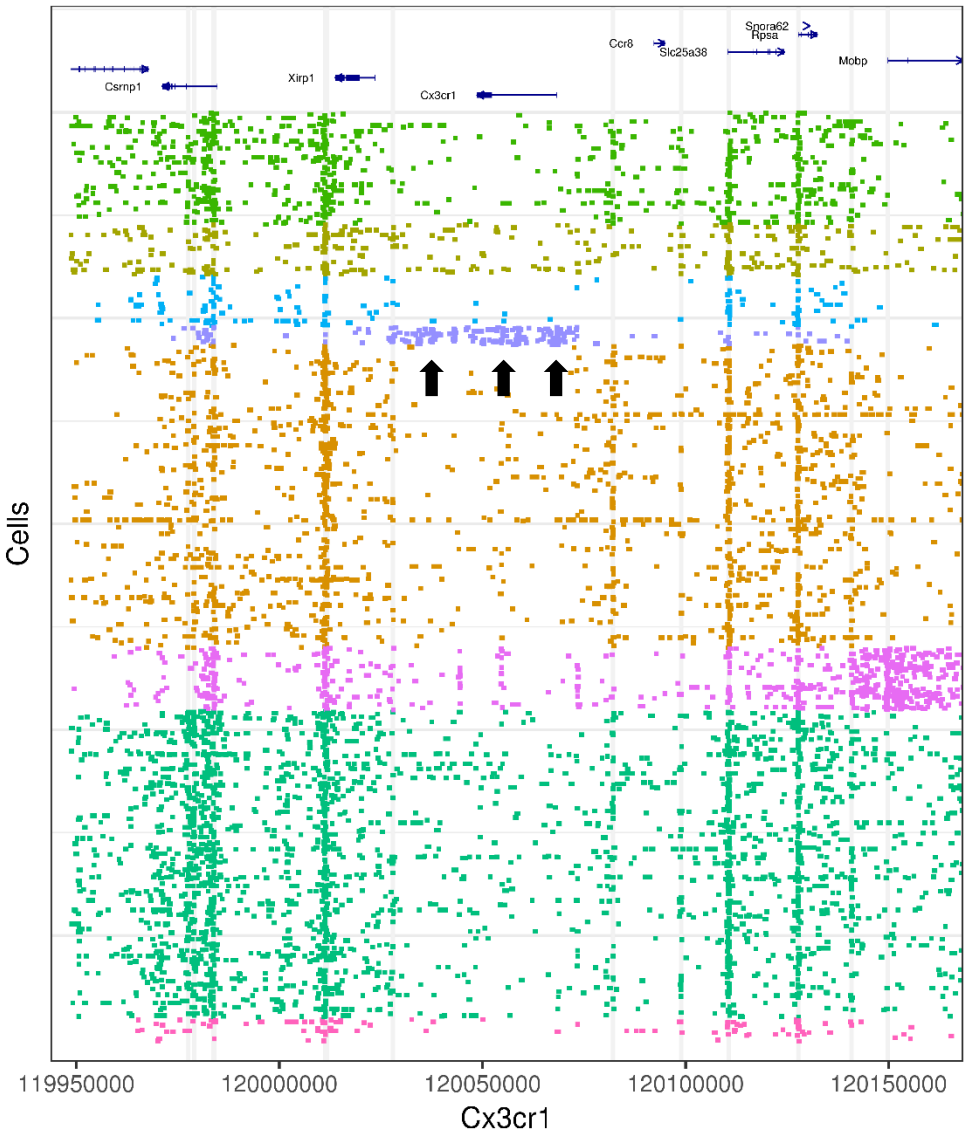




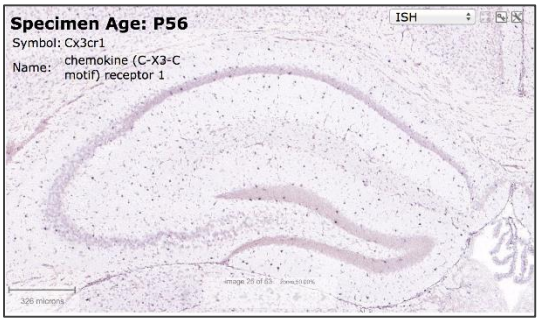


J) *Cx3cr1*
(Microglia)

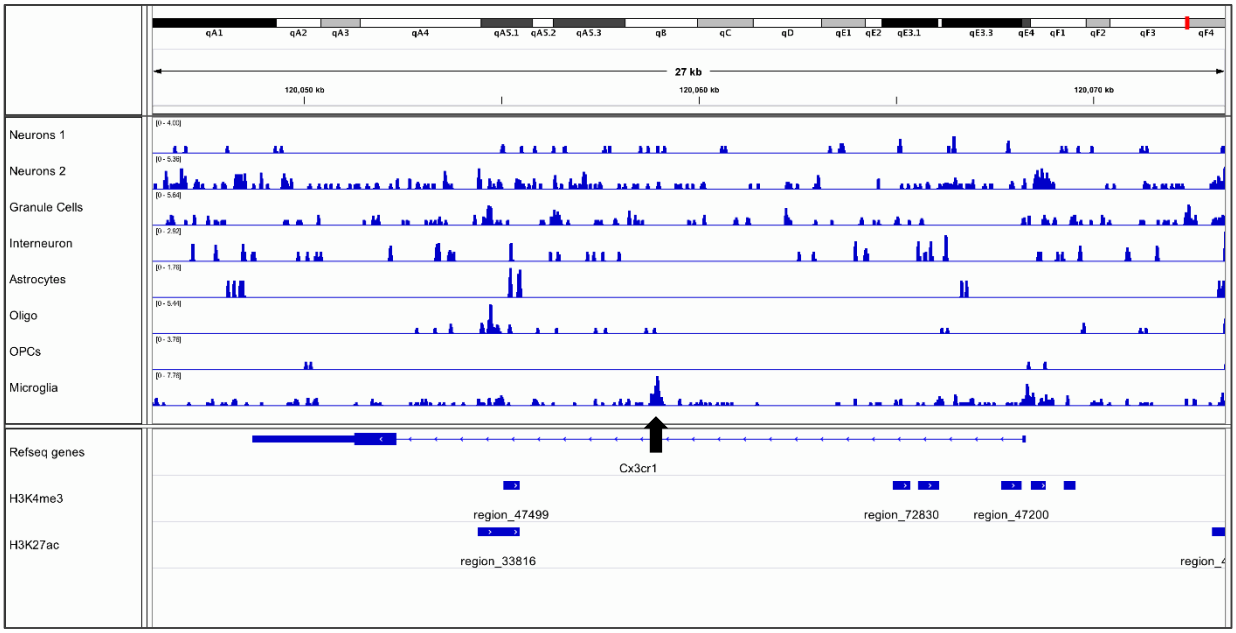
PMID: 25258010
Glia. 2015 Apr;63(4):531-548



AST
GRN
INT
MRG
NR1
NR2
OLI
OPC



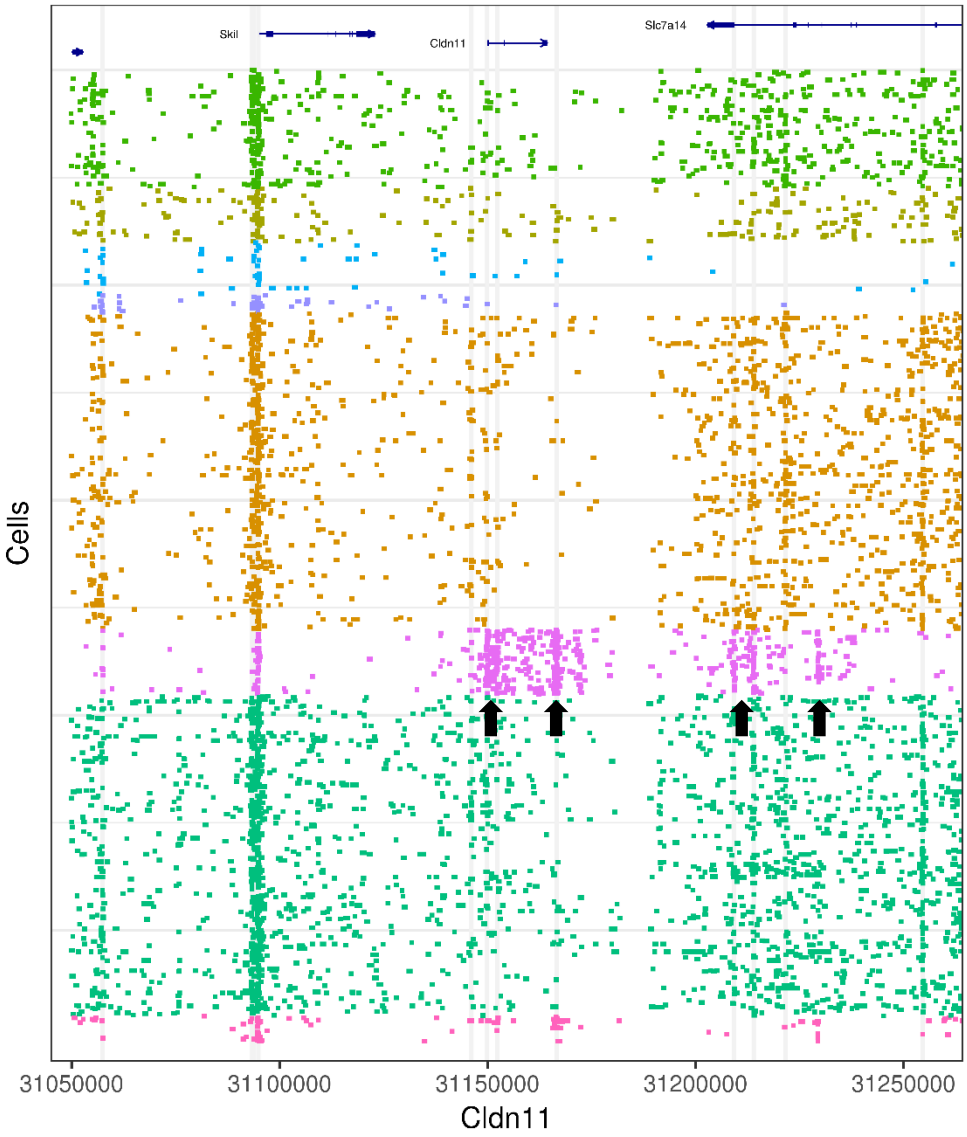
Note: Coronal section (no sagittal available)



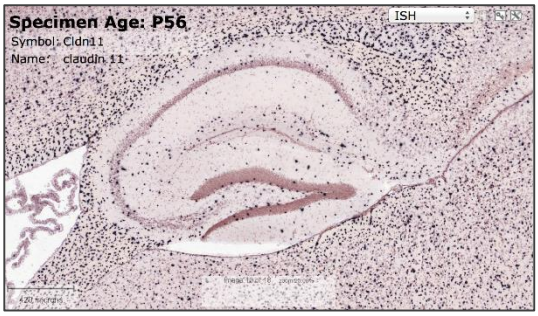
K) *Cldn11*

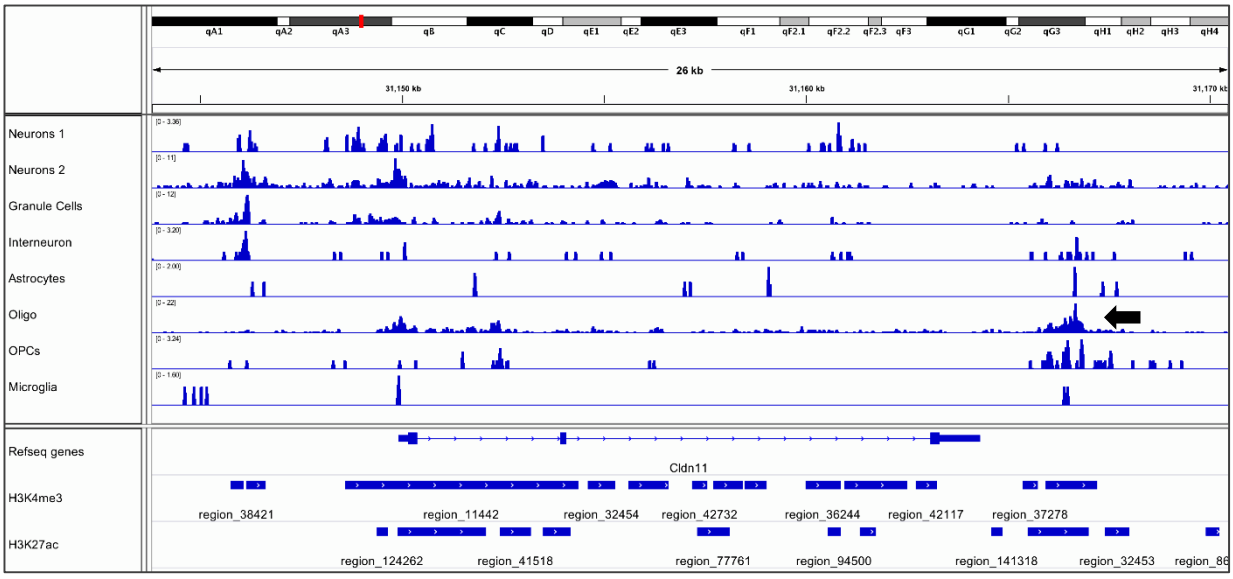
(Oligodendrocytes)

PMID: 10225958
J Cell Biol. 1999 May 3;145(3):579-88



- AST
- GRN
- INT
- MRG
- NR1
- NR2
- OLI
- OPC

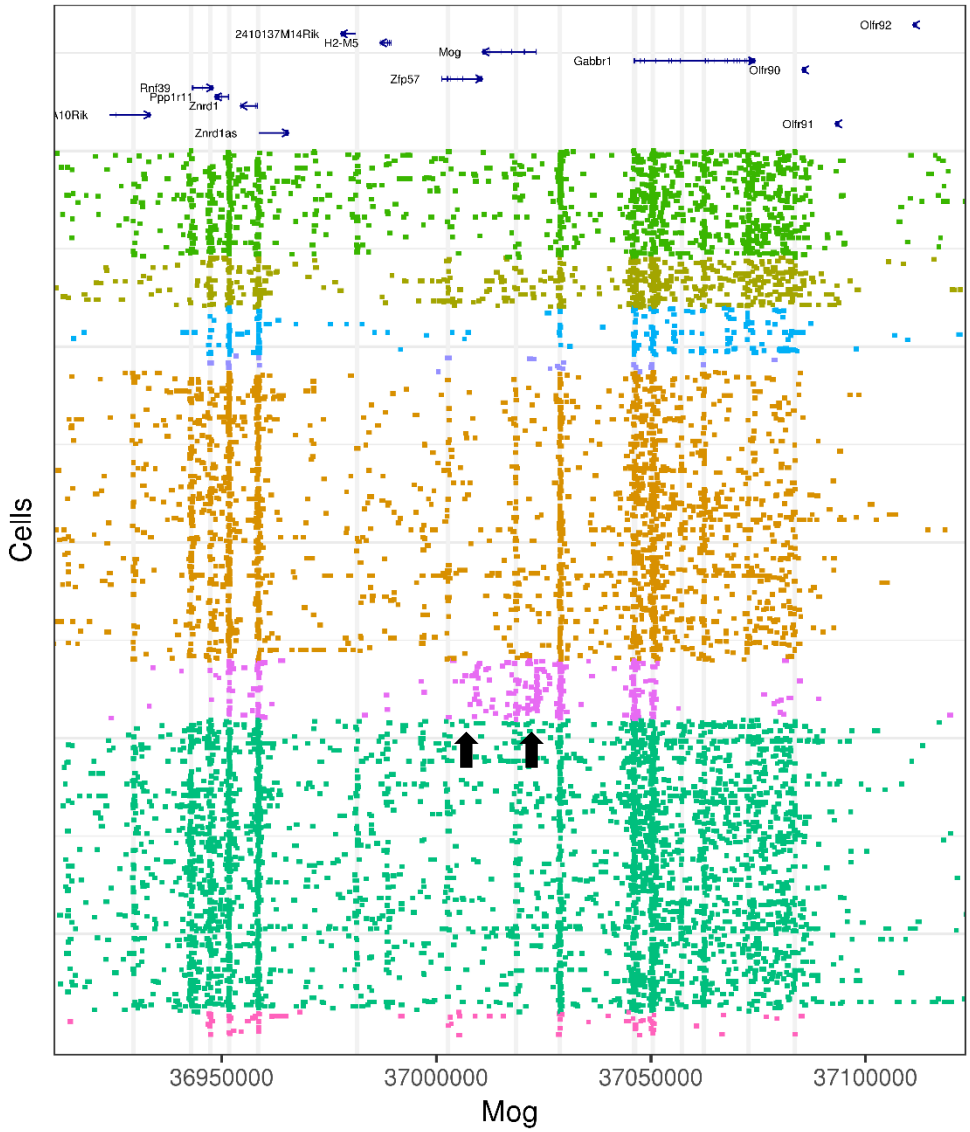




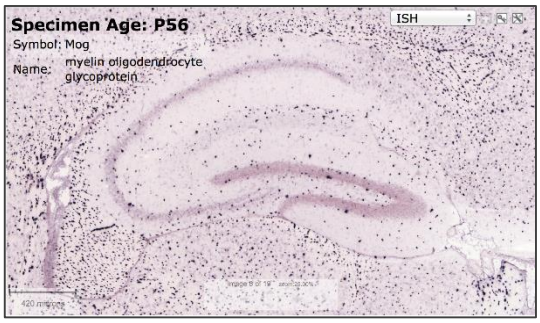
L) *Mog*

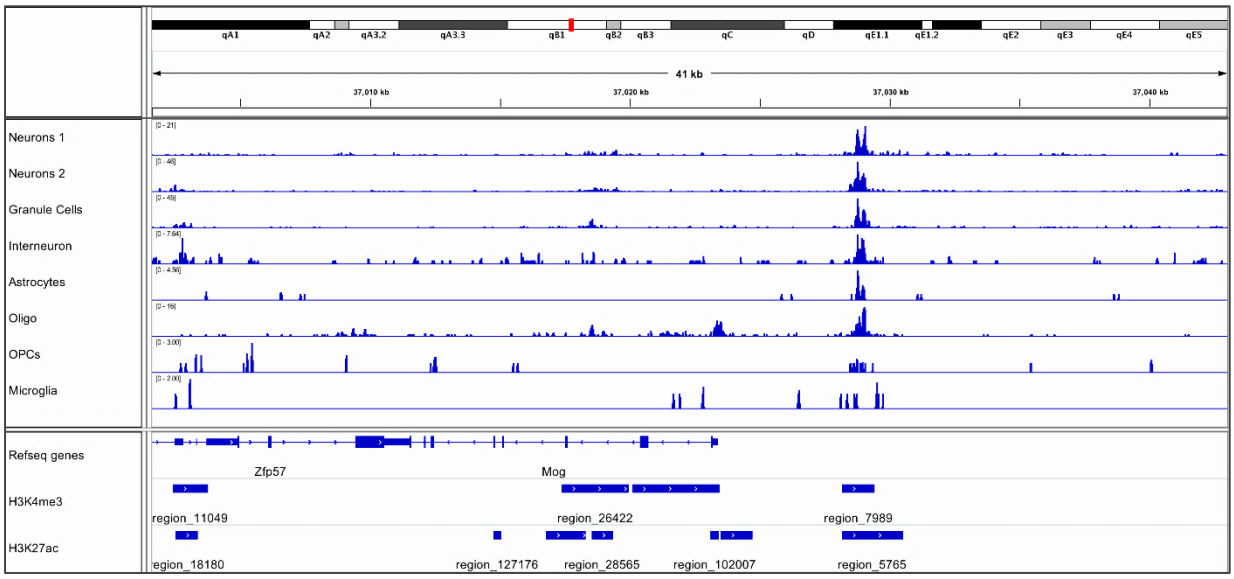
(Oligodendrocytes)

PMID: 2649509
J Neuroimmunol. 1989 May;22(3):169-76.



- AST
- GRN
- INT
- MRG
- NR1
- NR2
- OLI
- OPC

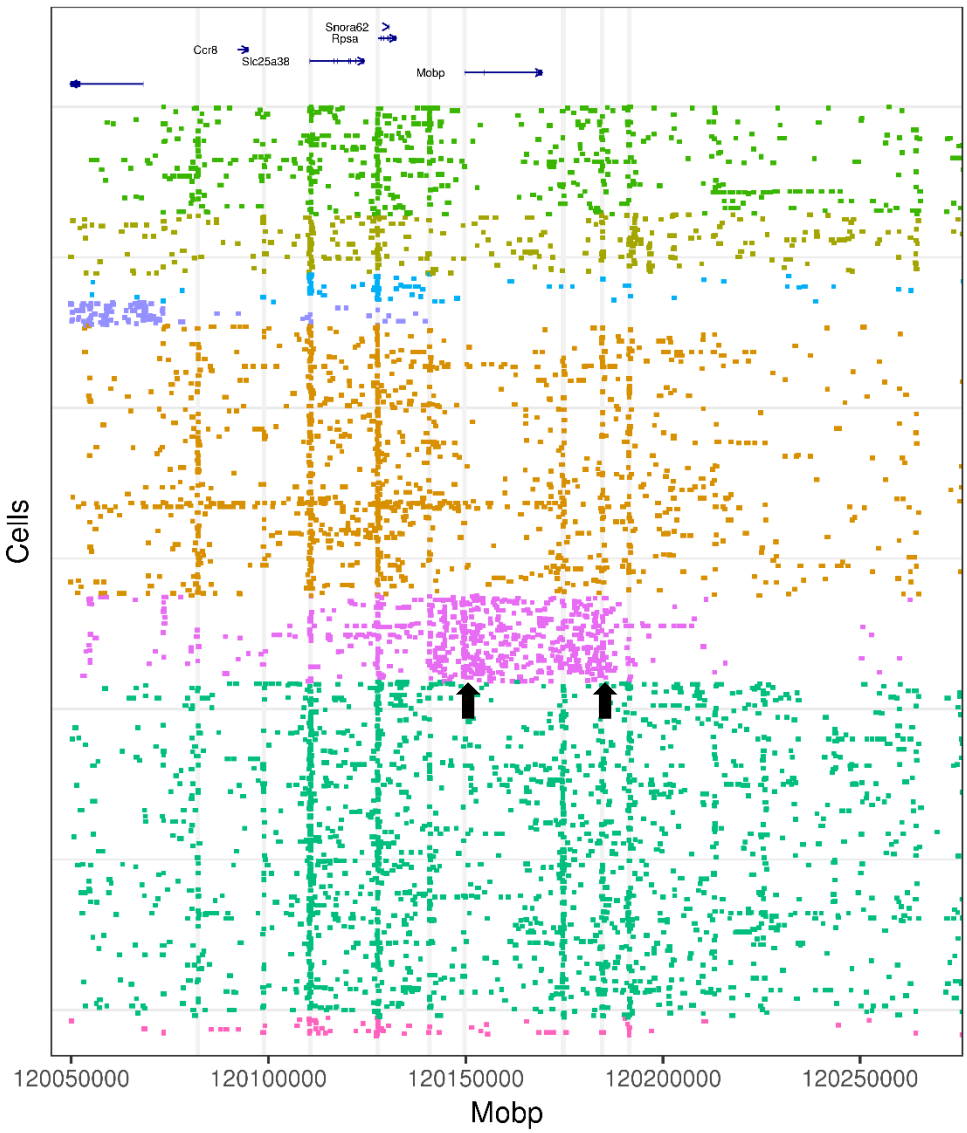




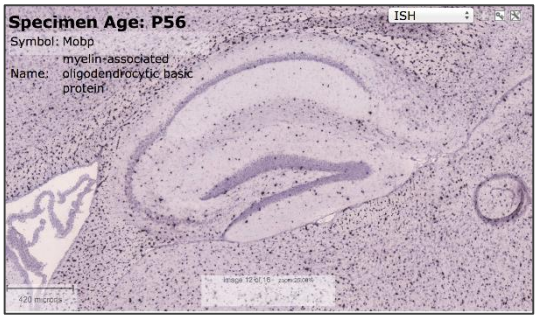
M) *Mobp*

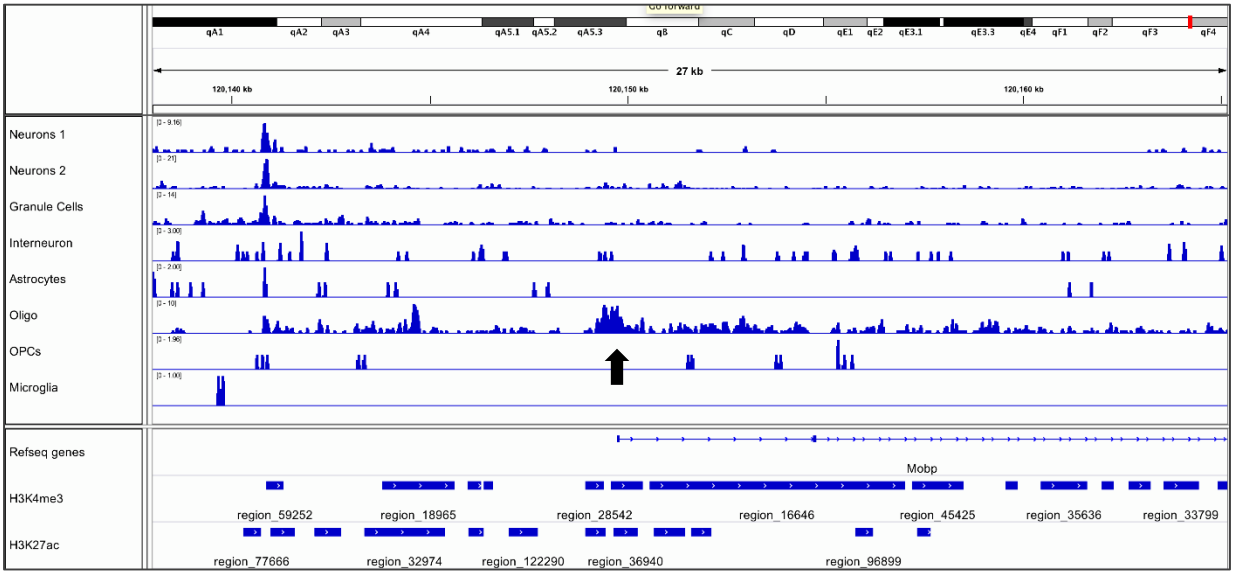
(Oligodendrocytes)

PMID: 17065439
J Neurosci. 2006 Oct 25;26(43):10967-83.



- AST
- GRN
- INT
- MRG
- NR1
- NR2
- OLI
- OPC

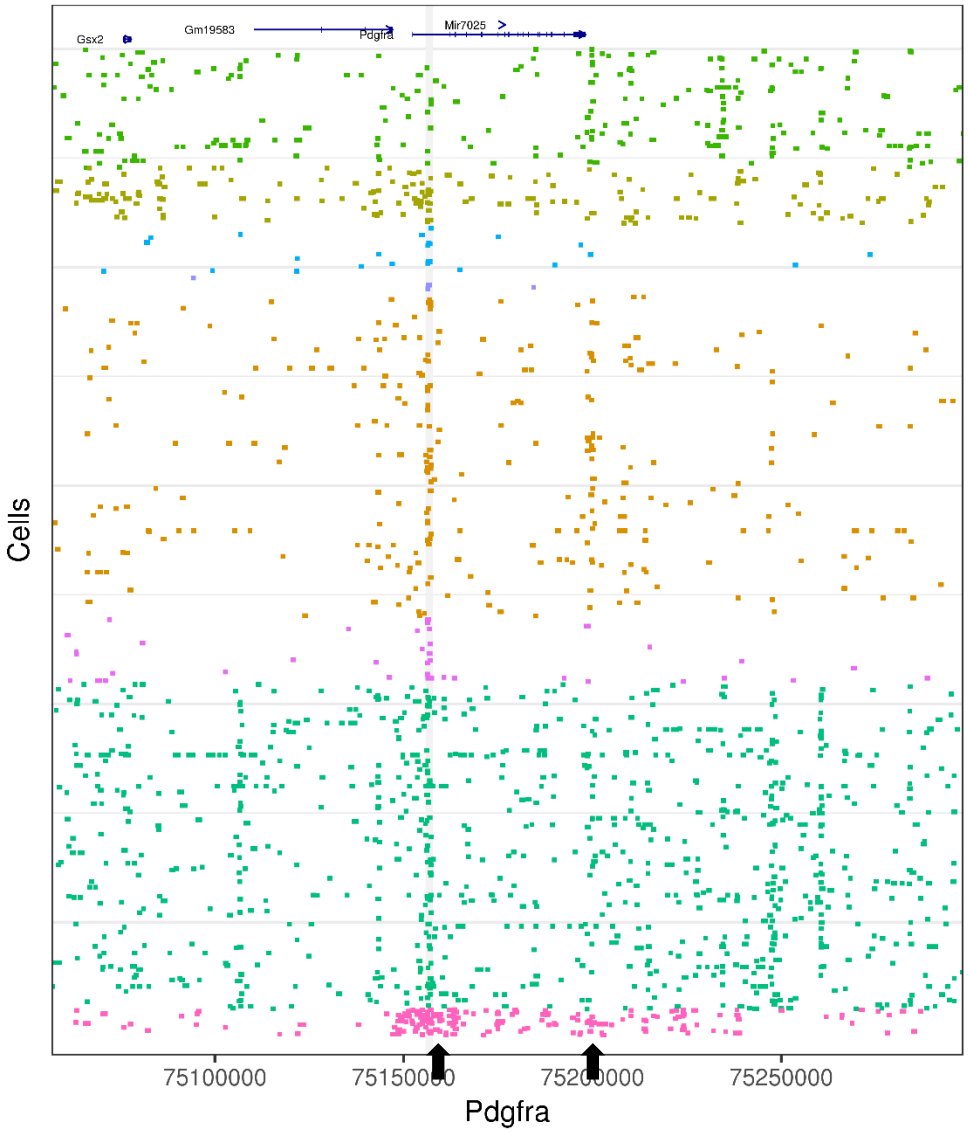




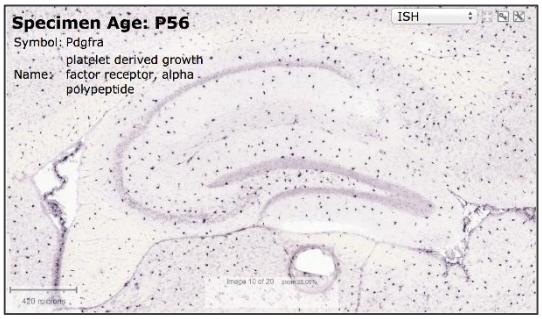
N) *Pdgfra*

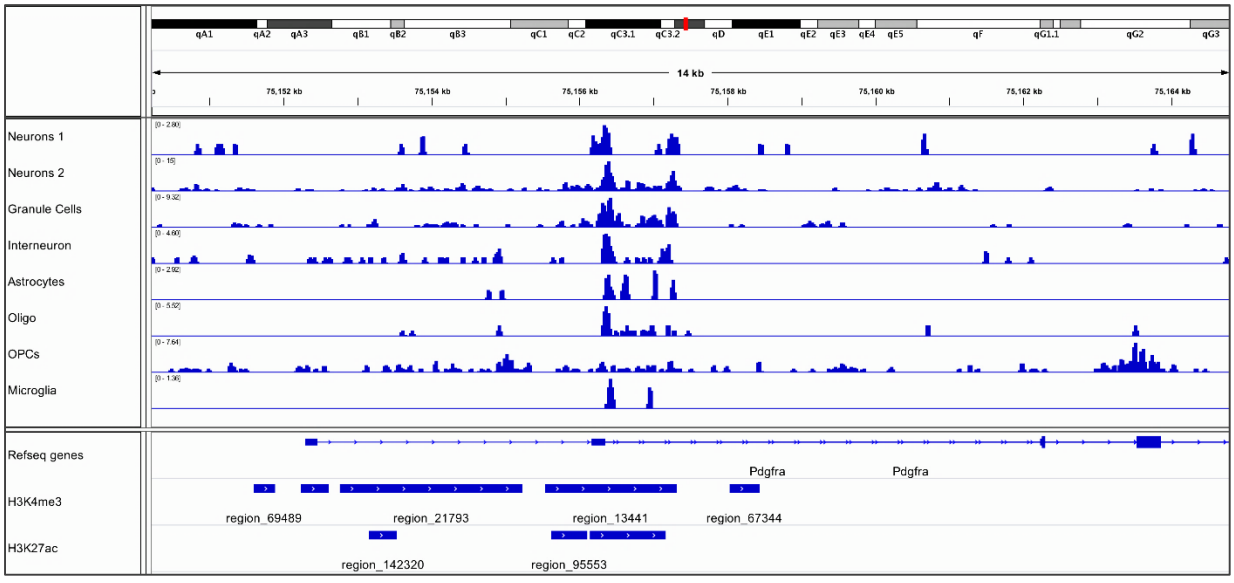
(Oligodendrocyte Progenitor Cells)

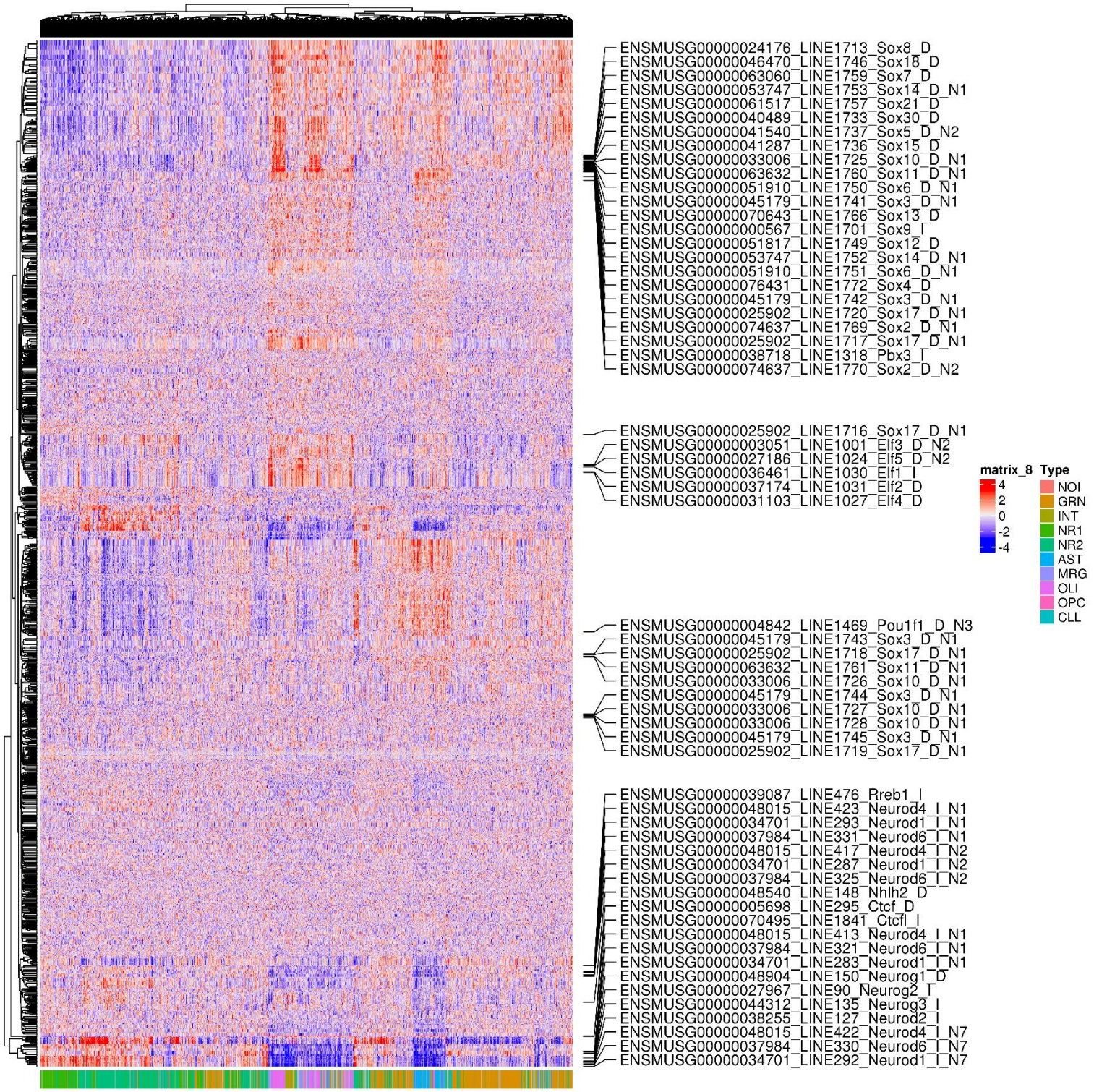
PMID: 18849983
Nat Neurosci. 2008 Dec;11(12):1392-401.
doi: 10.1038/nn.2220



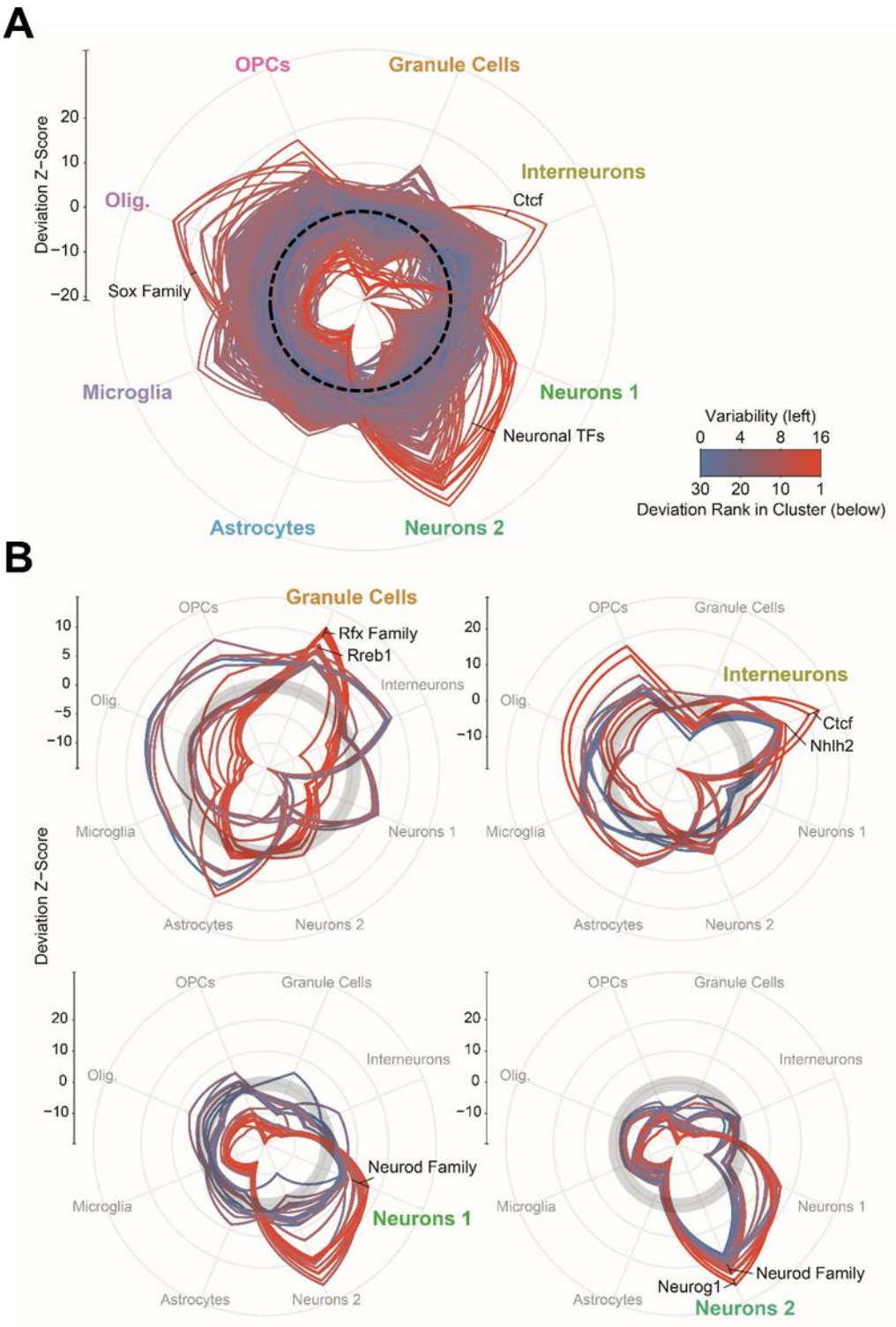
- AST
- GRN
- INT
- MRG
- NR1
- NR2
- OLI
- OPC ⬅



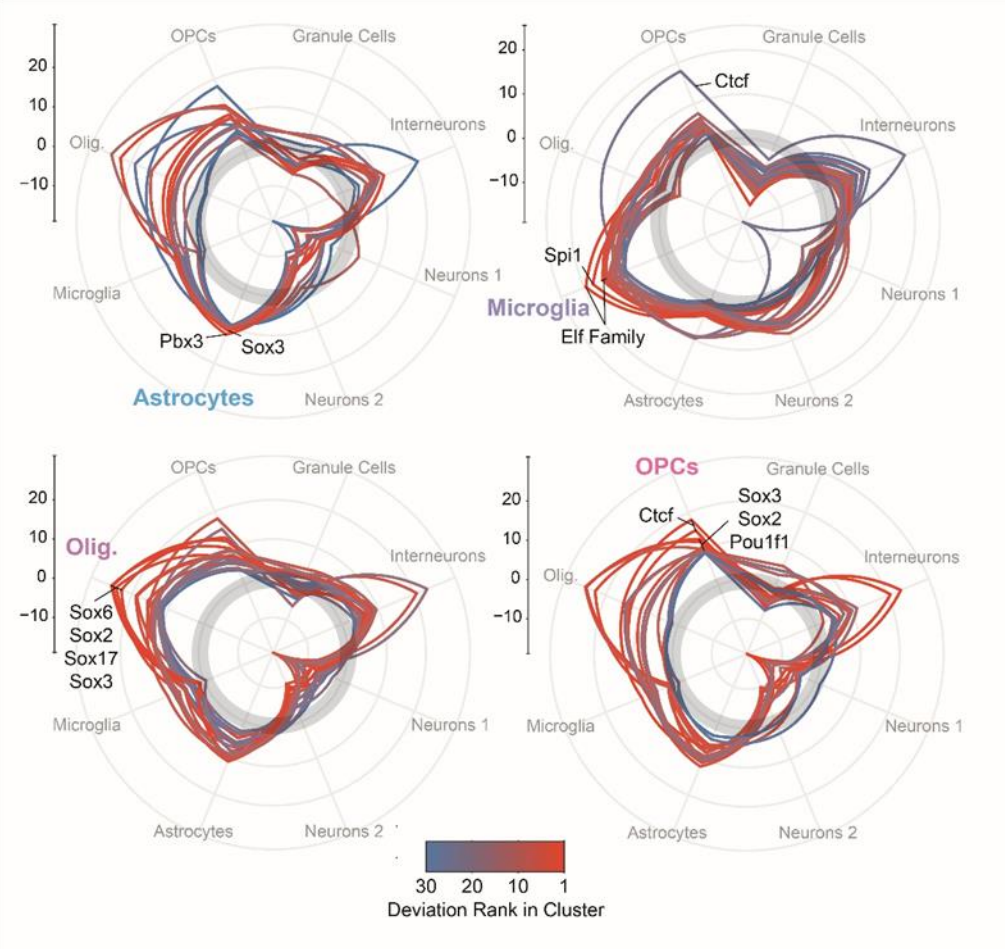




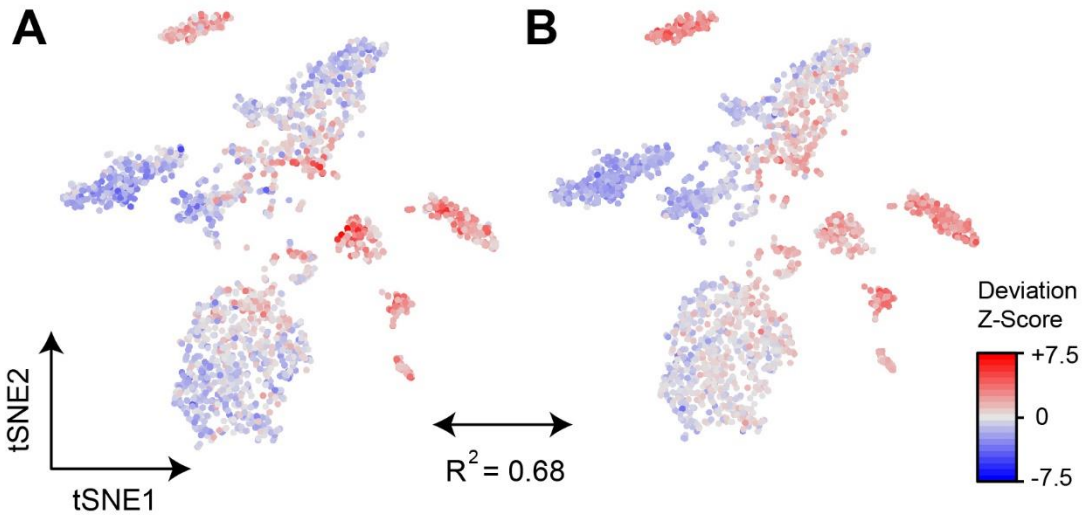
Supplemental Figure 4: ChromVAR motif deviations heatmap
Bottom axis corresponds to cell type, highlighted motifs correspond to important cell type specific factors.


A
OPCs
Granule Cells
Interneurons
Olig.
Microglia
Neurons 1
Astrocytes
Neurons 2
Deviation Z−Score
20
10
0
−10
−20
Ctcf
Sox Family
Neuronal TFs
Variability (left)
0 4 8 16
30 20 10 1
Deviation Rank in Cluster (below)
B
Granule Cells
Rfx Family
Rreb1
Interneurons
Ctcf
Nhlh2
Neurod Family
Neurons 1
Neurog1
Neurons 2
Deviation Z−Score

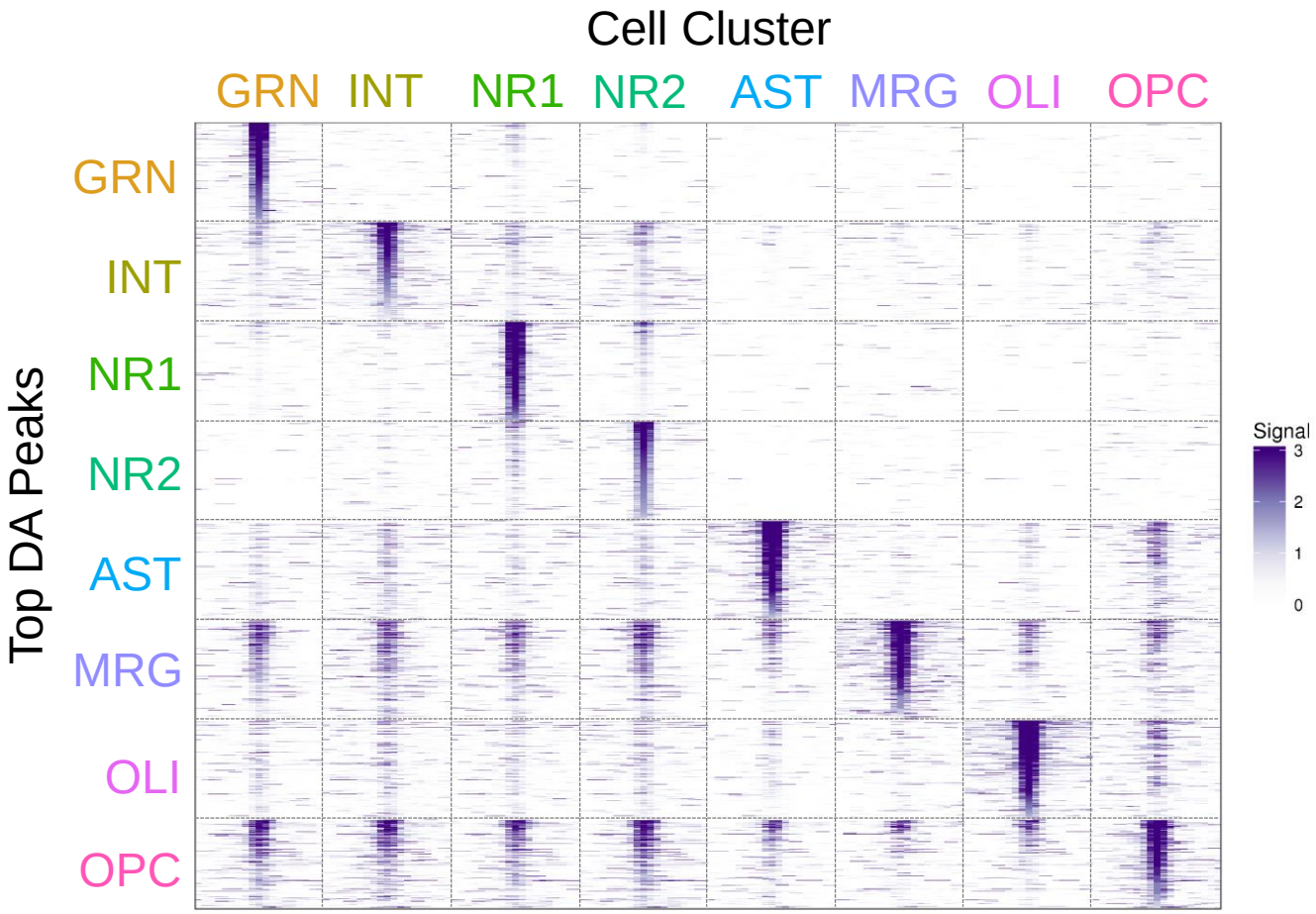
B (cont)



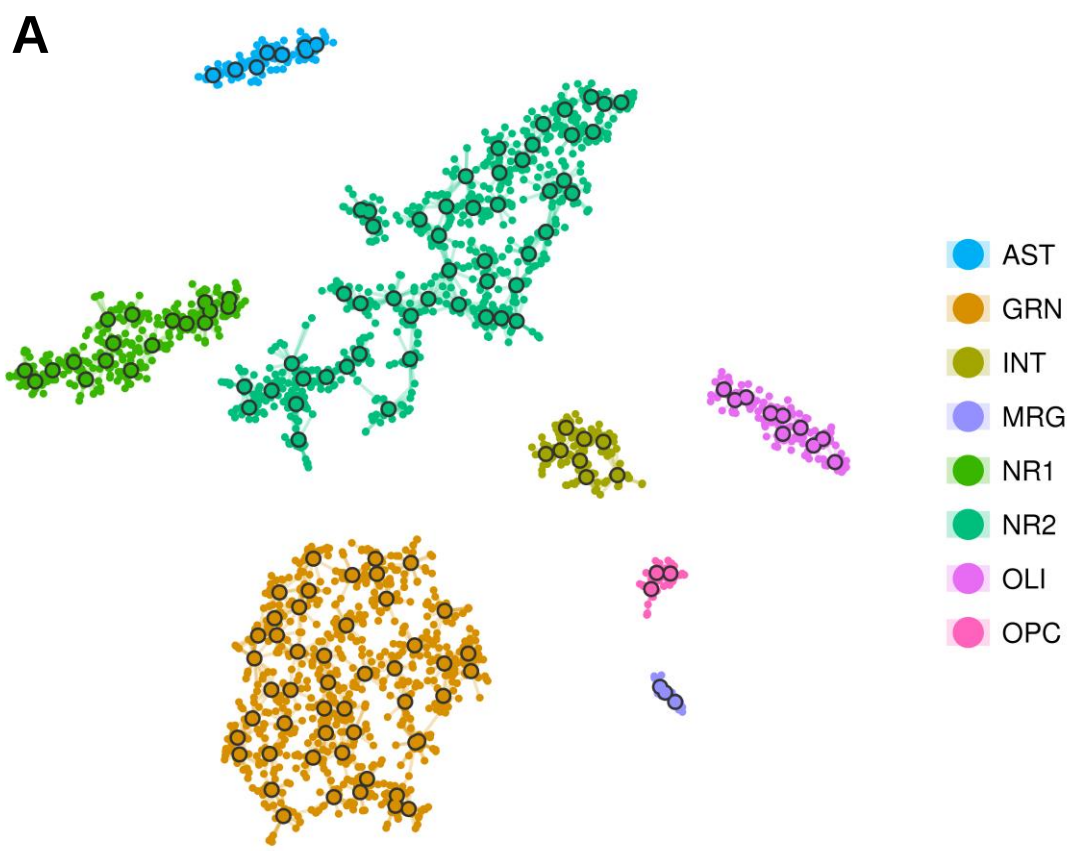
Supplemental Figure 5: Global motif profiles for cell types
A) *ChromVAR* mean deviation scores (y-axis) for each motif for each cell type. Color indicates the variability score for the motif as reported by *ChromVAR*. The black dashed line is at 0, i.e. the mean. B) Deviation scores (y-axis) from *ChromVAR* for the top 30 motifs with the highest deviation z-scores as shown in A. Color indicates the ranking within the cluster. Gray shaded line is at 0, i.e. the mean. Polygon plots were utilized to confer similar shapes of top motif accessibility between clusters.



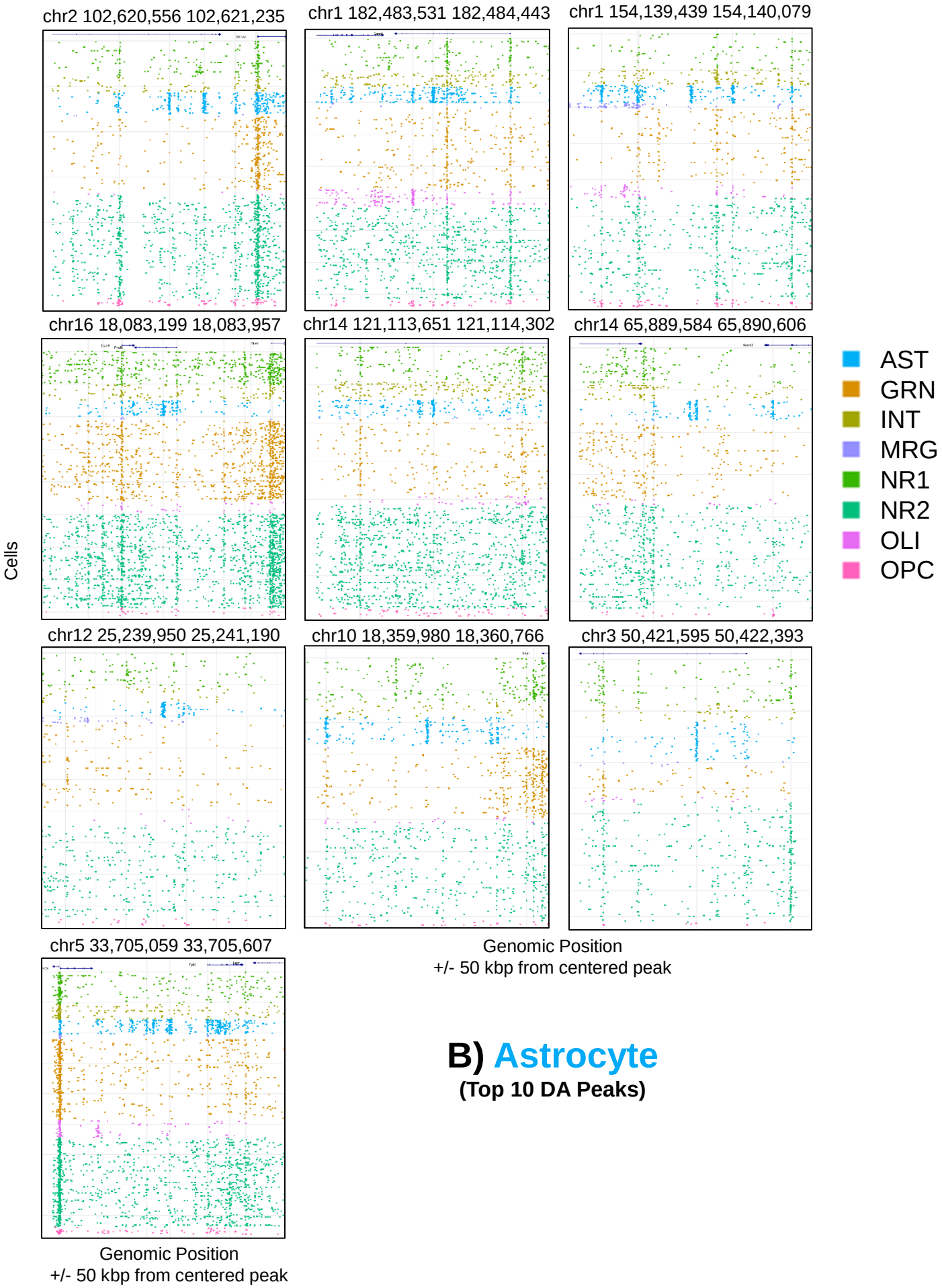
Supplemental Figure 6: Comparison of CTCF motif and ChIP-seq deviations
A) *ChromVAR* analysis of global accessibility using motif locations in the genome. B) *ChromVAR* deviation scores for CTCF ChIP-seq (Sams et al. 2016) peaks. Deviation sets show the same broad trends and show high correlation (Pearson $R^2 = 0.68$).



Supplemental Figure 7: DA Peak ATAC-seq Signal
ATAC-seq signal for the top 1,000 differentially accessible peaks for each cell type are shown as in Fig. 2 but sorted according to peaks with the top signal versus significance of differential accessibility.

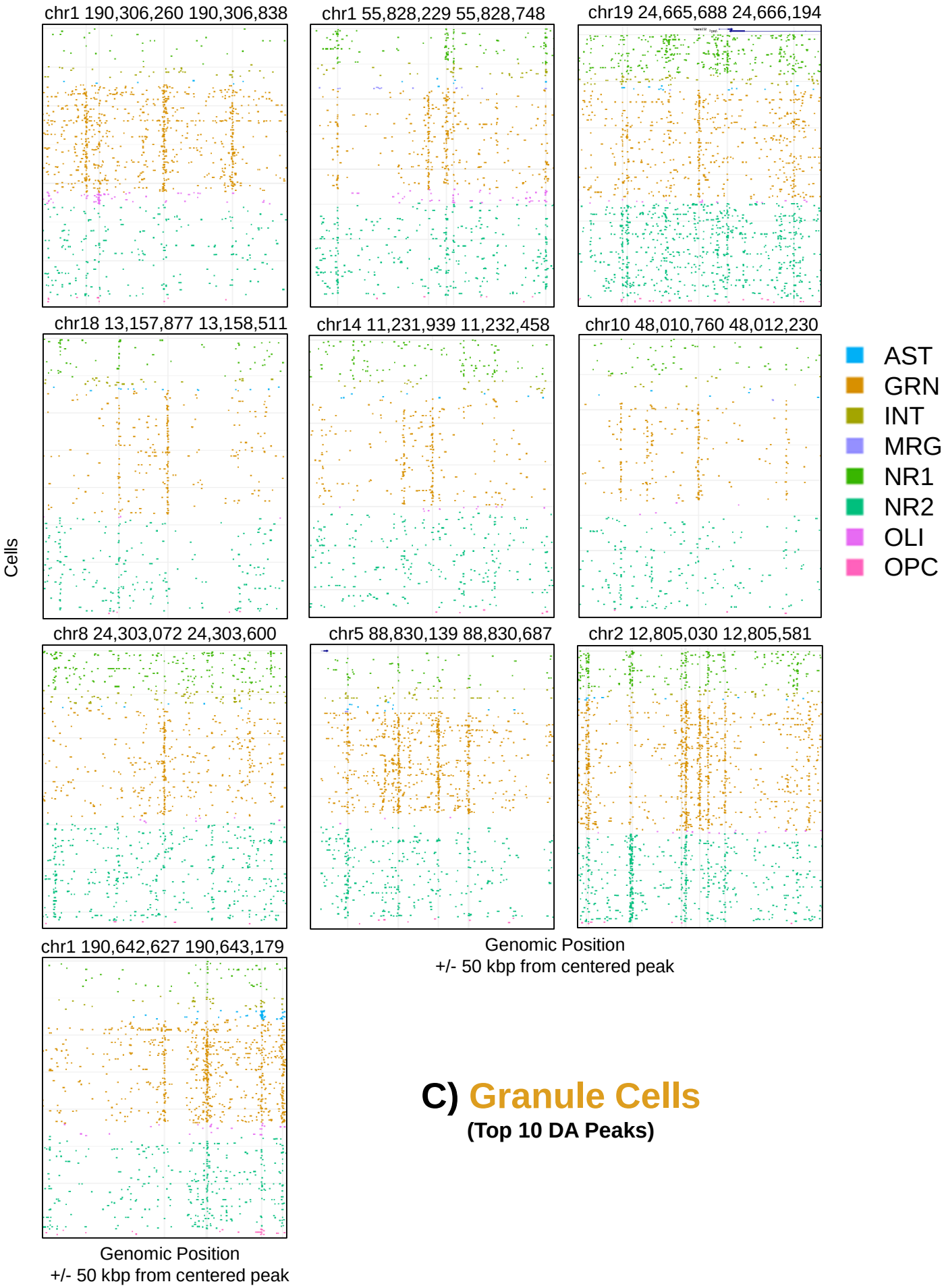


Supplemental Figure 8: Differentially Accessible Peaks
A) Scitools aggregate-cells plot showing which cells were grouped into subclusters that serve as replicates into the DA analysis. B-I) The top 10 differentially accessible (Supplemental Table 2, see Methods) peaks corresponding to each cluster are plotted along with the sci-ATAC-seq reads present within the region +/- 50,000 basepairs of the identified DA peak (panel title, centered).



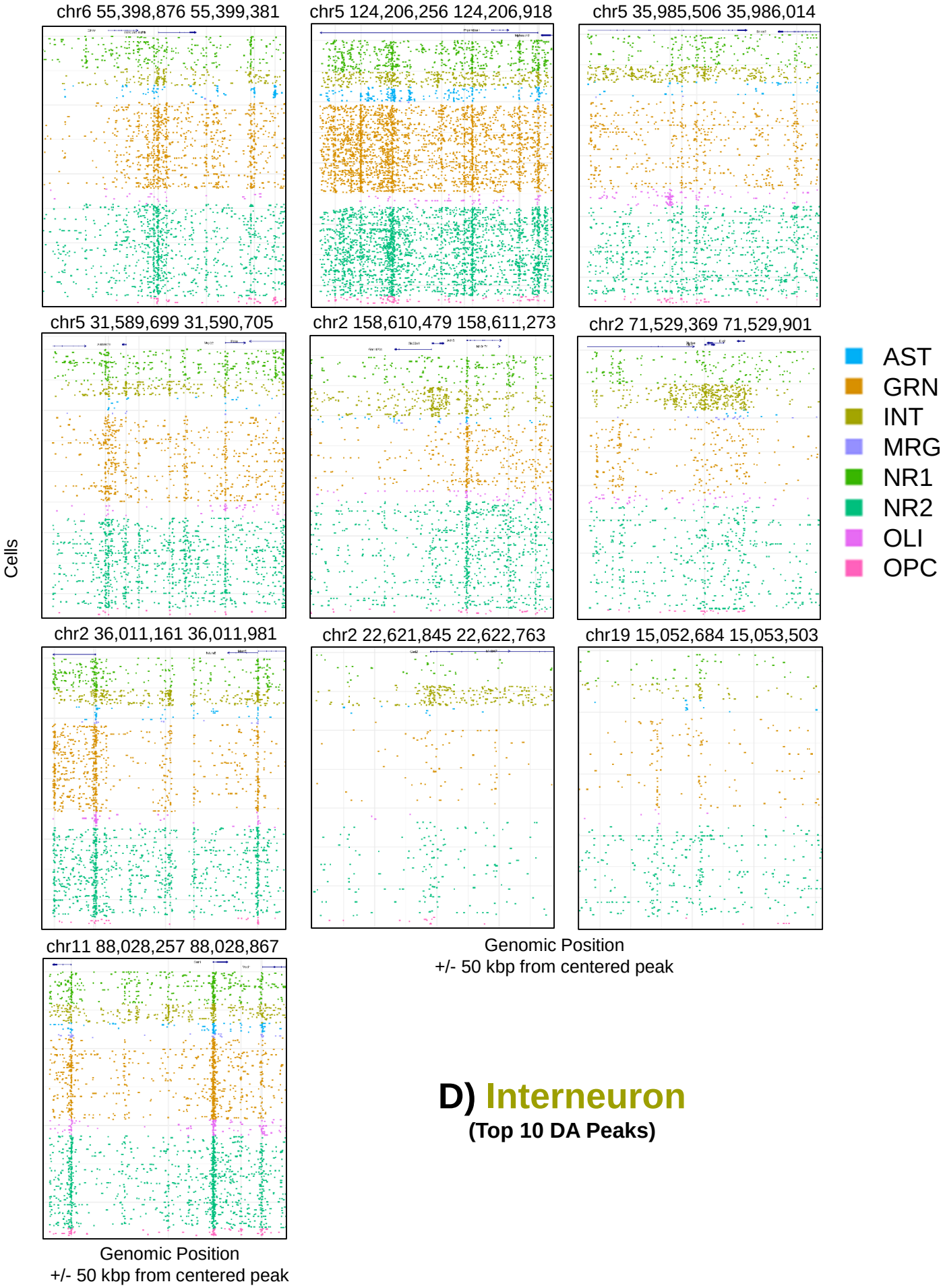
B) Astrocyte

(Top 10 DA Peaks)



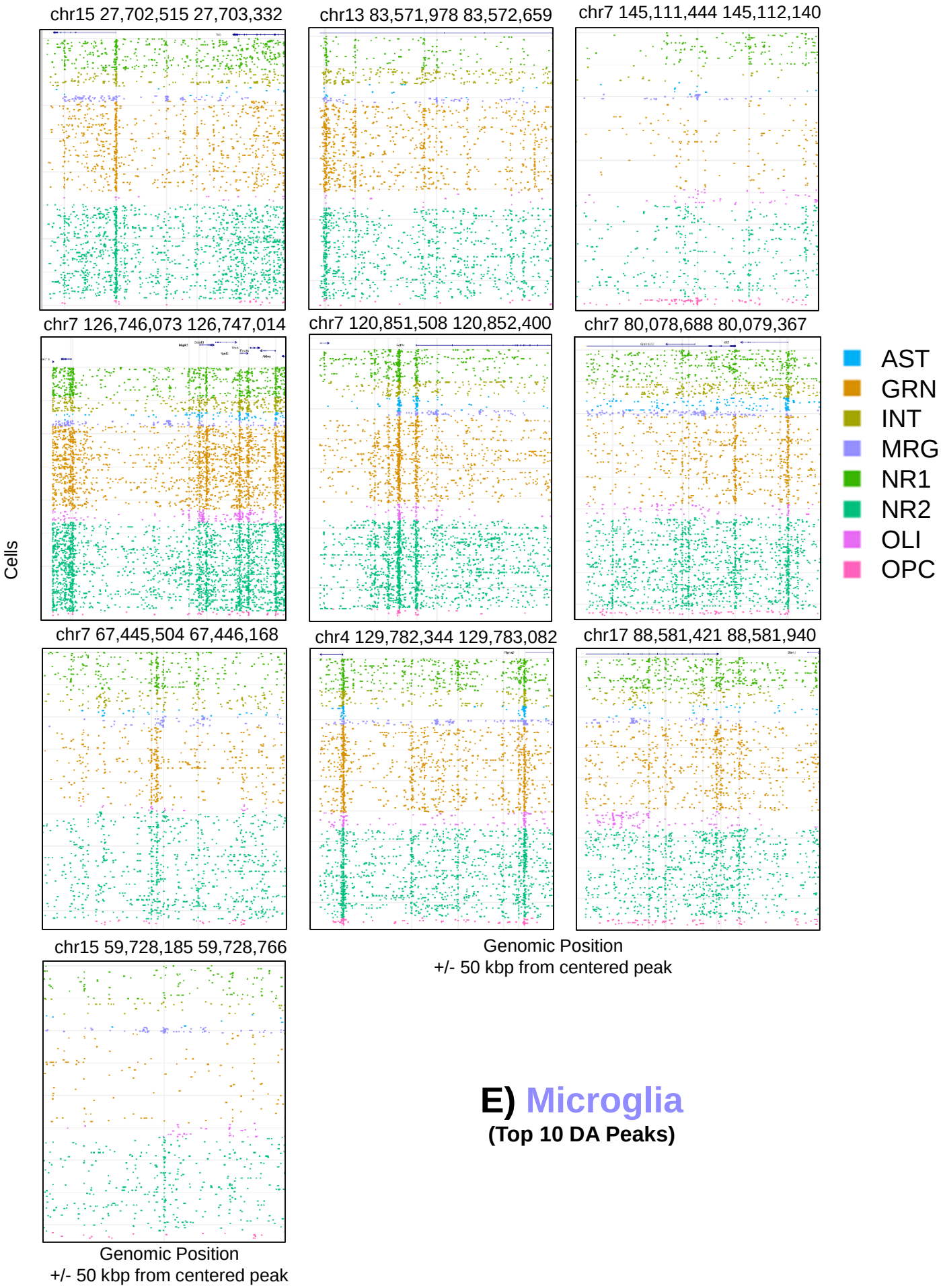
C) Granule Cells

(Top 10 DA Peaks)



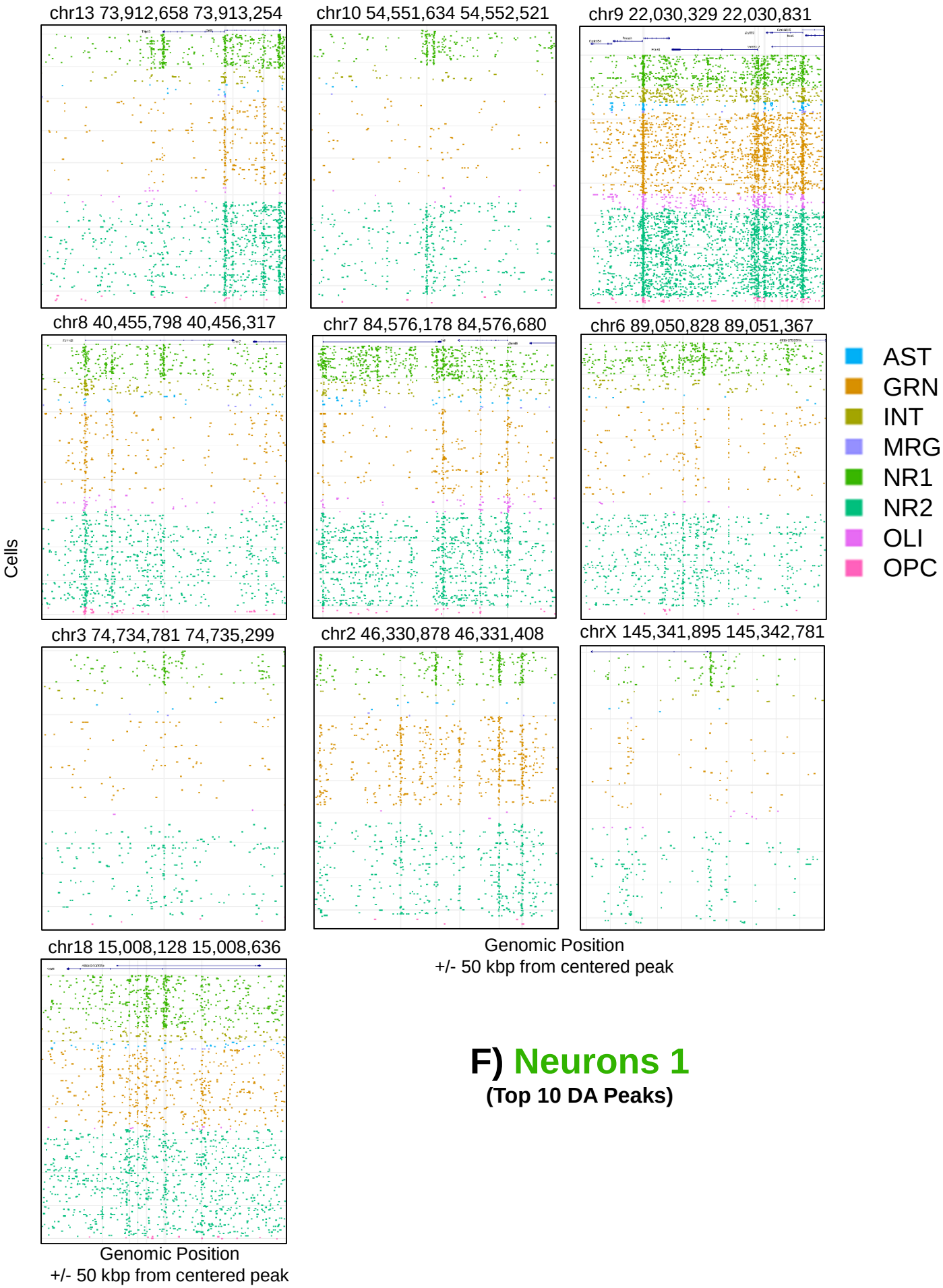
D) Interneuron

(Top 10 DA Peaks)



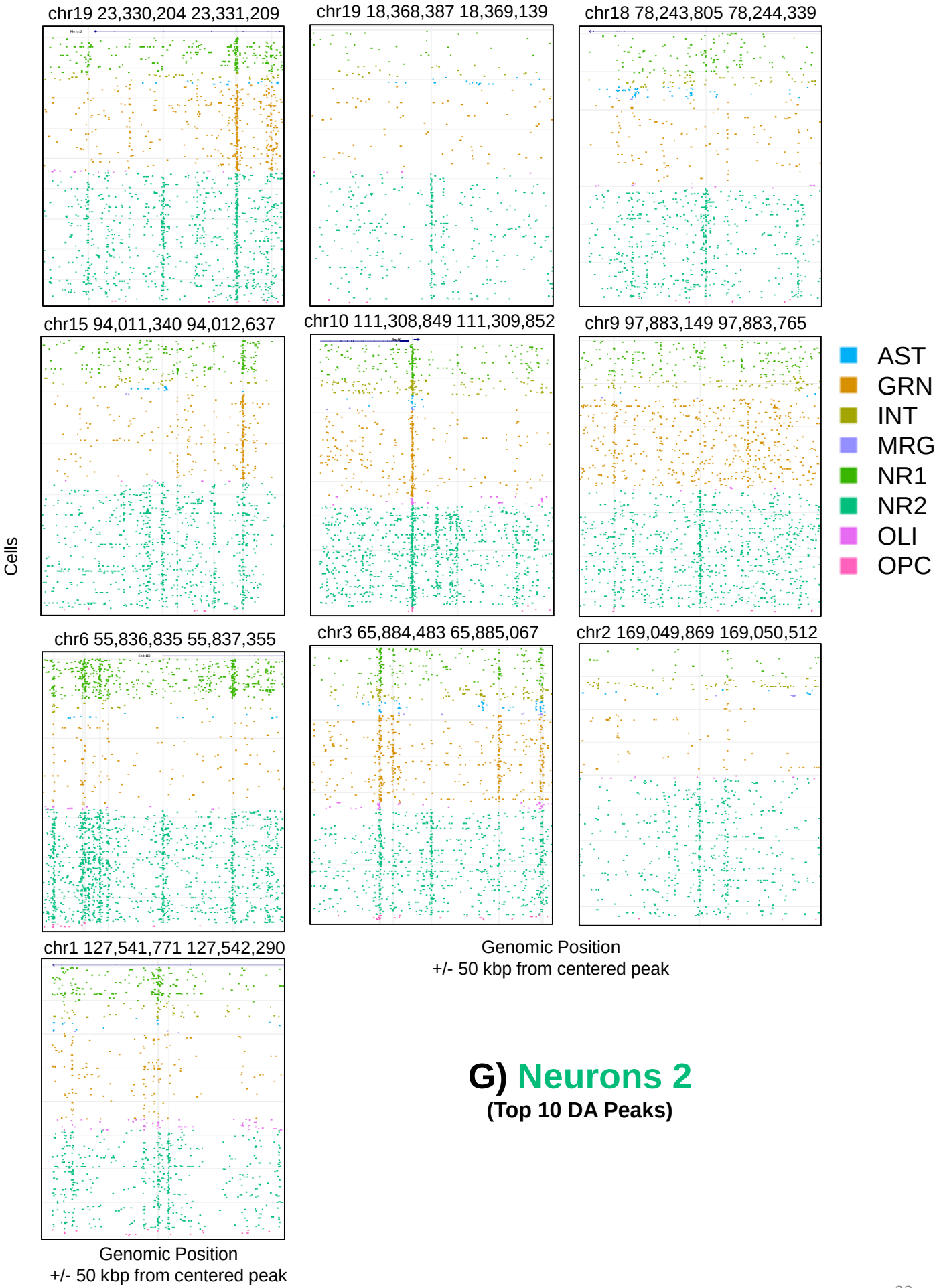
E) Microglia

(Top 10 DA Peaks)



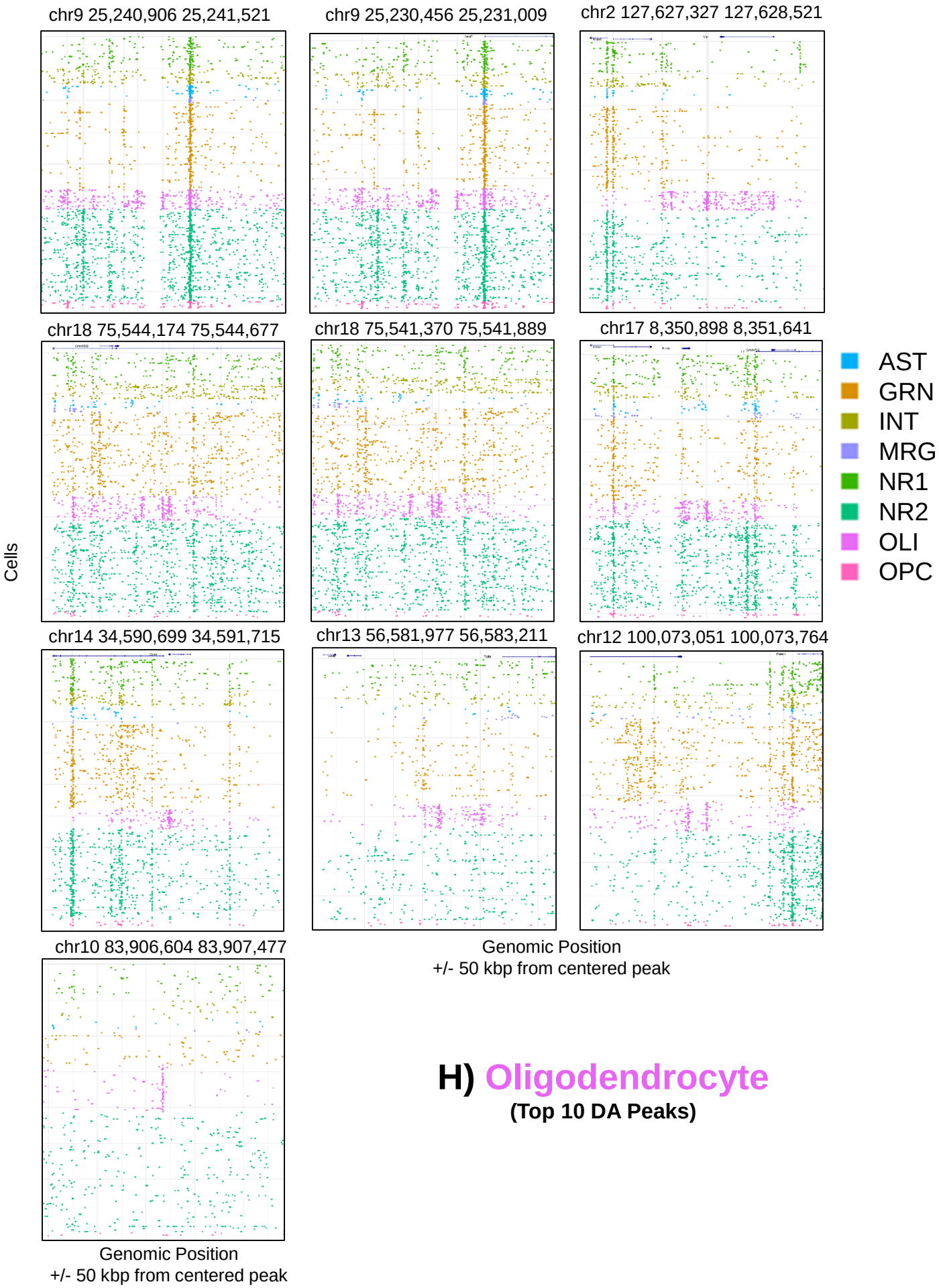
F) Neurons 1

(Top 10 DA Peaks)



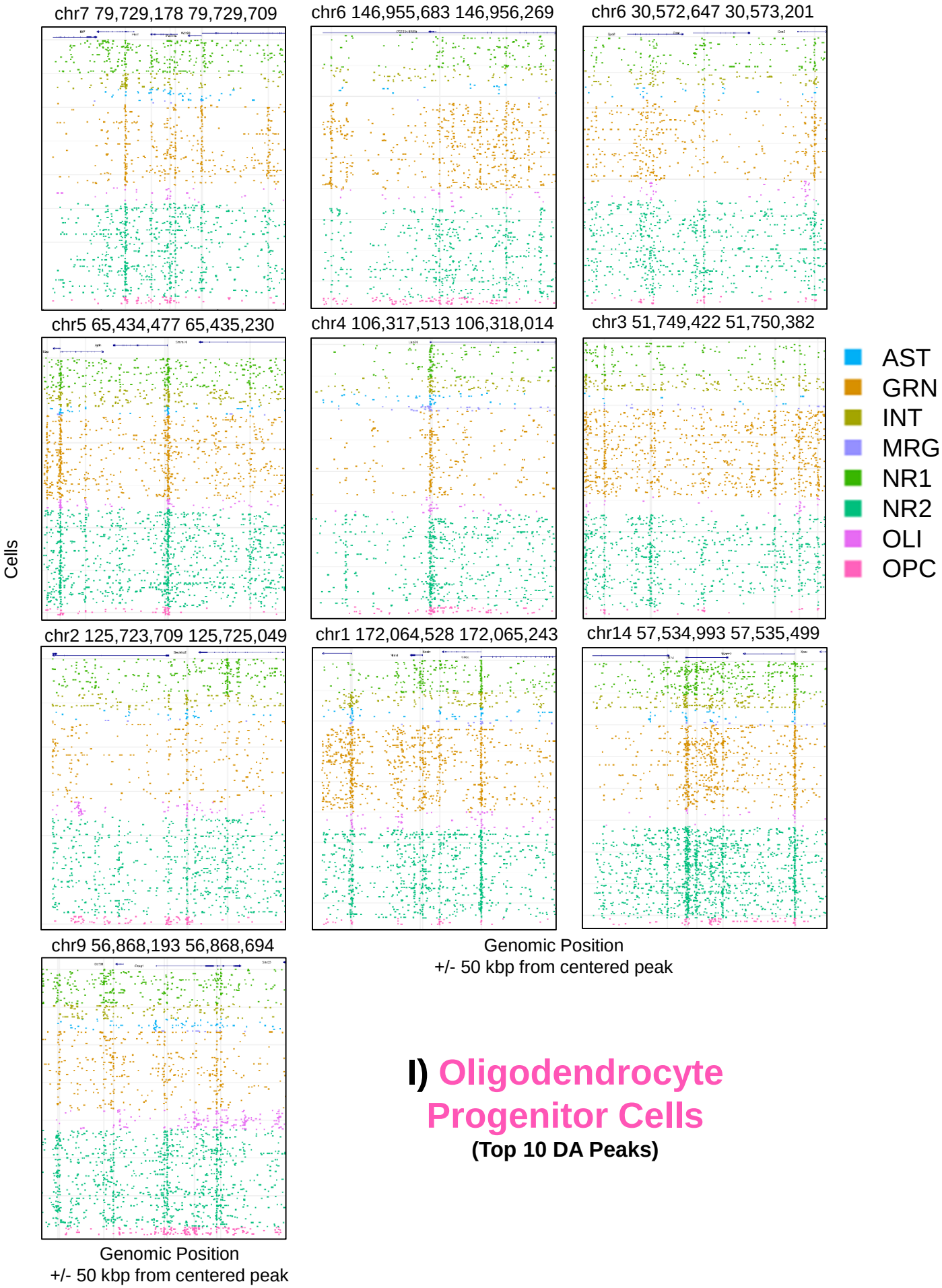
G) Neurons 2

(Top 10 DA Peaks)



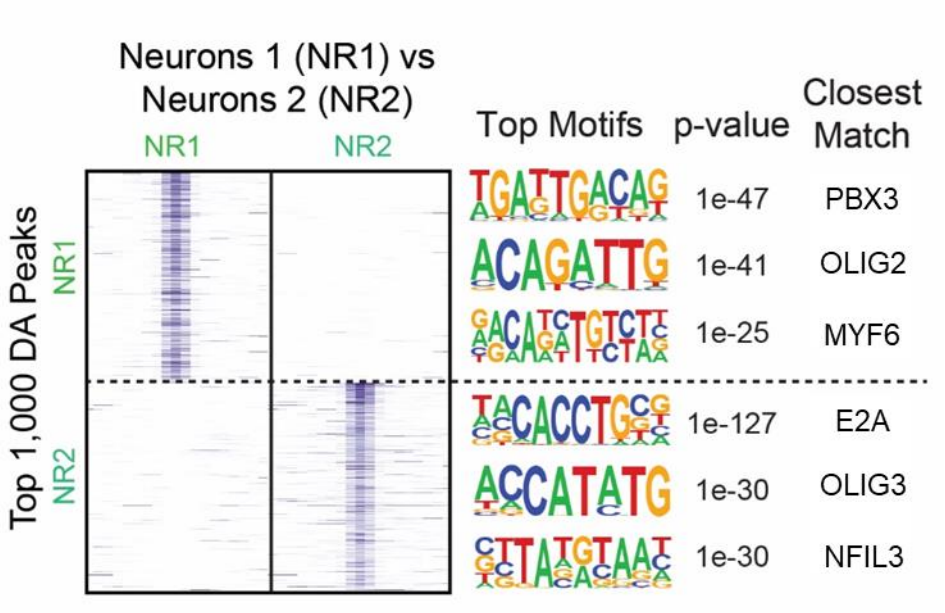
H) Oligodendrocyte

(Top 10 DA Peaks)



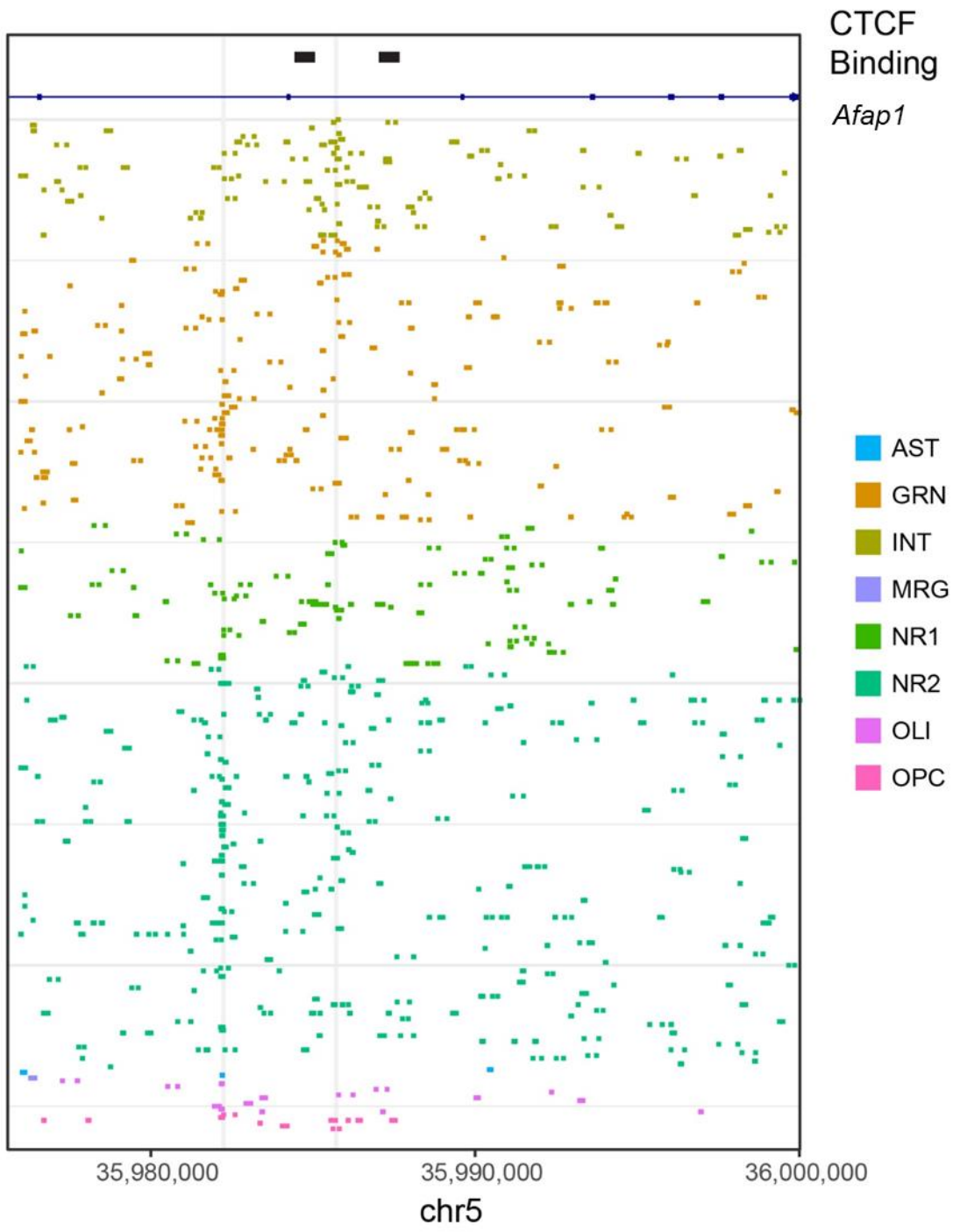
I) Oligodendrocyte Progenitor Cells

(Top 10 DA Peaks)

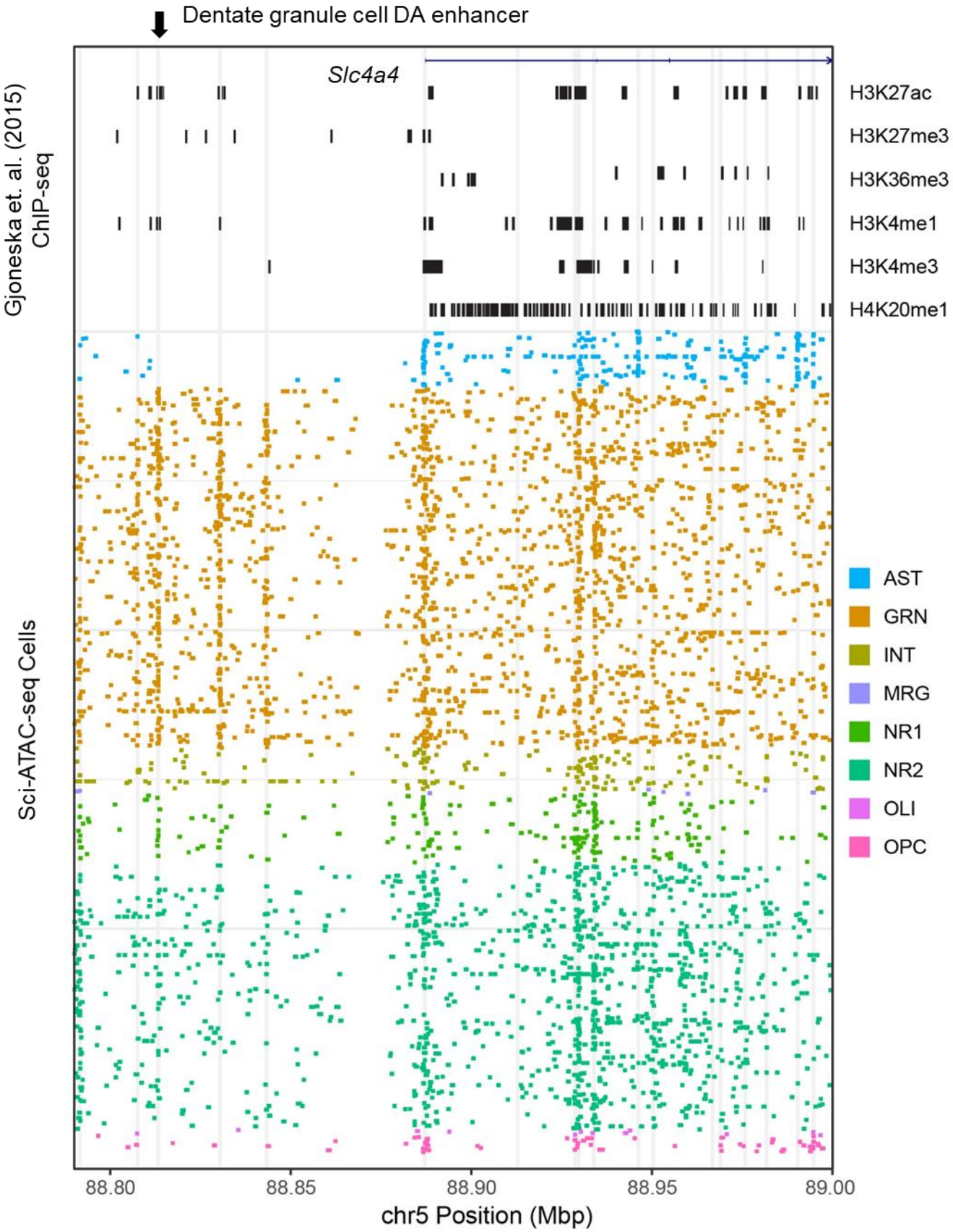


Supplemental Figure 9: Top DA peaks between NR1 and NR2 excitatory neuron cell types

ATAC-seq signal for the top 1,000 DA peaks is plotted for each cell type along with the top three motifs associated with each peak set and their corresponding p-value and closest matching motif.

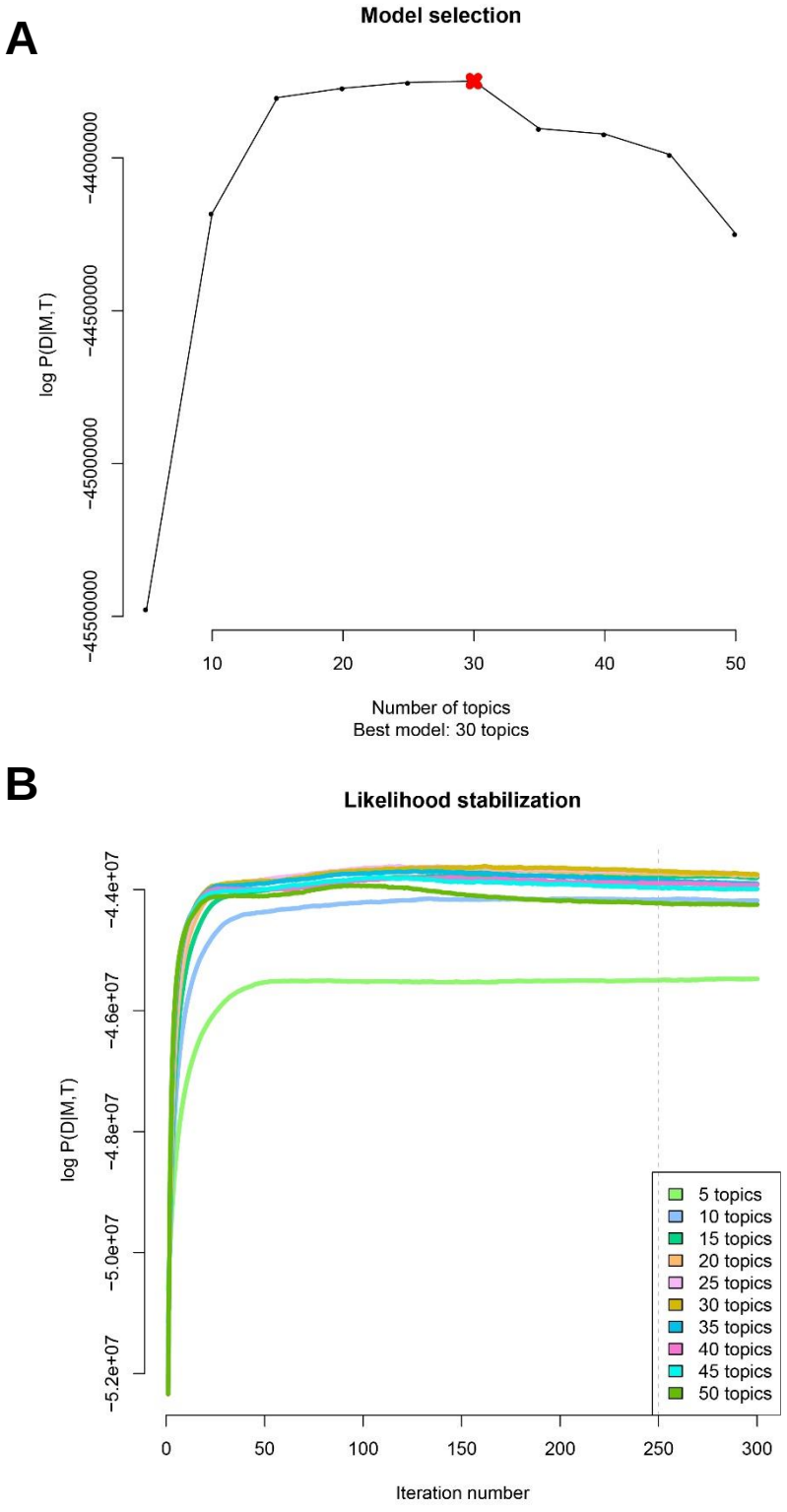


Supplemental Figure 10: *Afap1* locus
Read plot generated by scitools showing a peak highly enriched in interneurons that is flanked by CTCF ChIP-seq peaks.

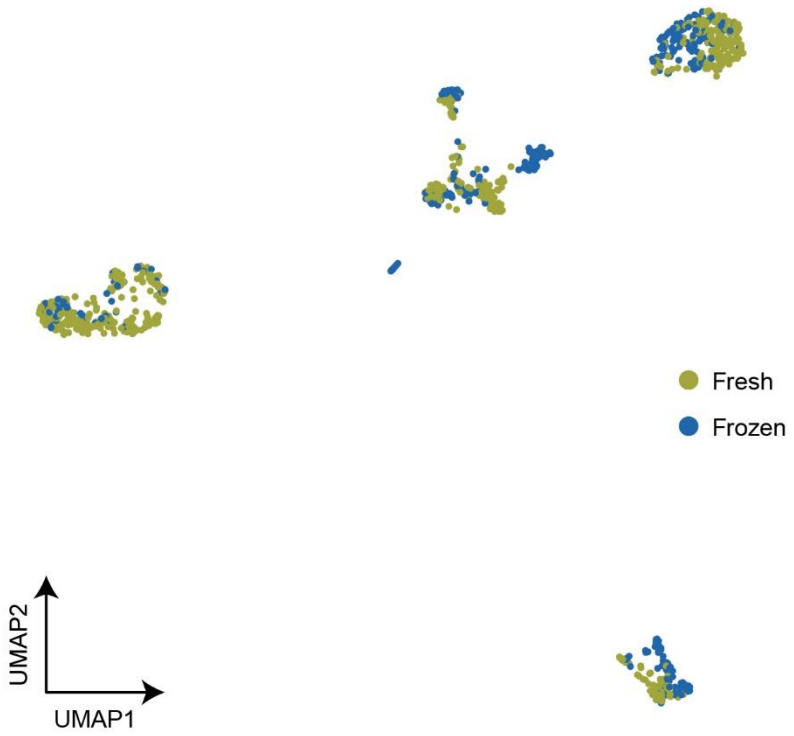


Supplemental Figure 11: *Slc4a4* locus

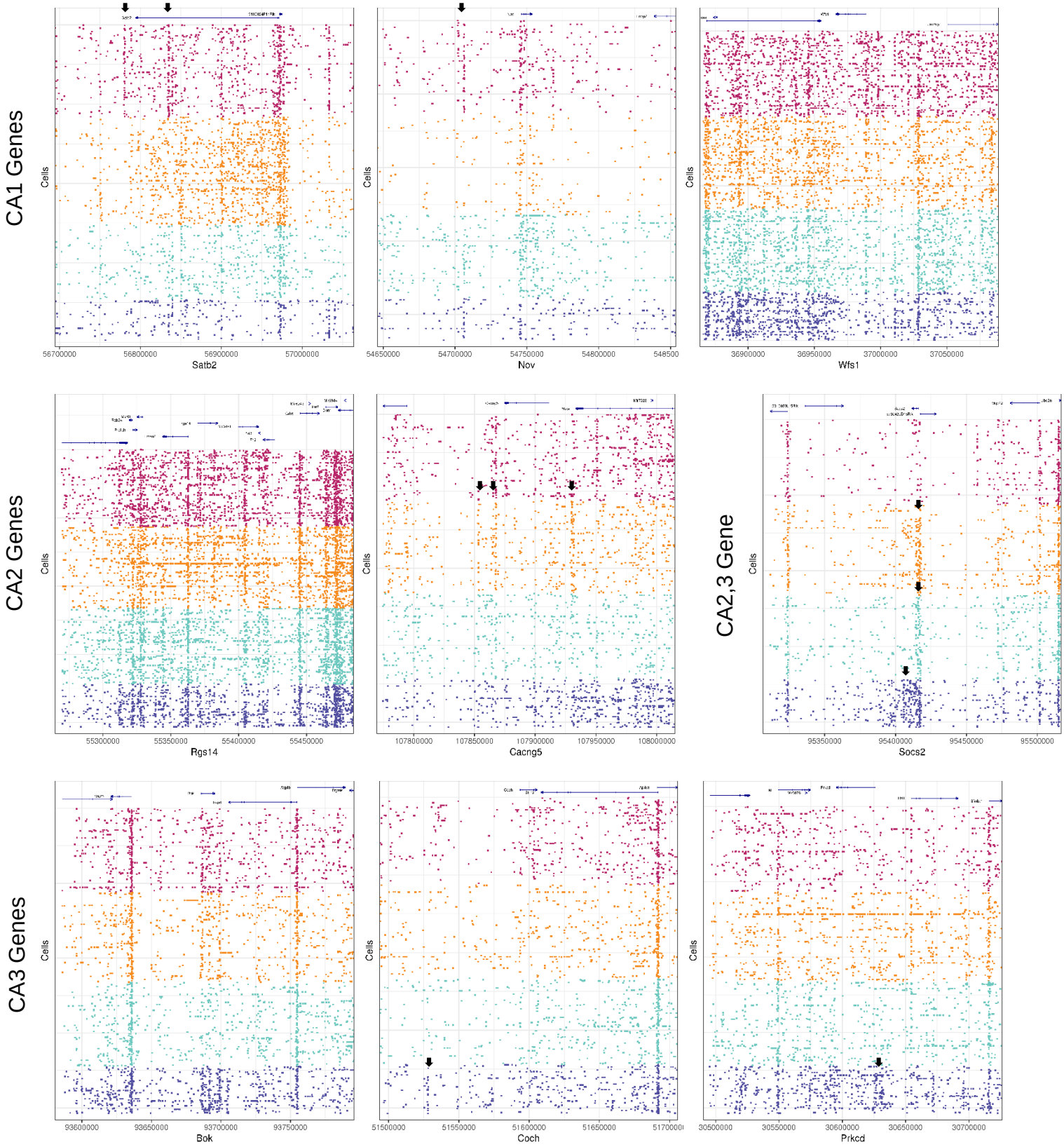
Read plot generated by scitools showing the putative cell-type-specific enhancer of *Slc4a4* that is differentially accessible in the dentate granule cell population. ChIP-seq peaks from Gjoneska et. al. (2015) are shown below the gene track in black.



Supplemental Figure 12: CisTopic Topic Selection for NR 1 & 2 clusters
(A) Multiple topics are run to optimize topic separation. 30 was determined to be the optimum number for subsequent analysis. (B) Stabilization of the model over iterations. All models stabilized by approximately 50 iterations, though 300 were run in all cases.

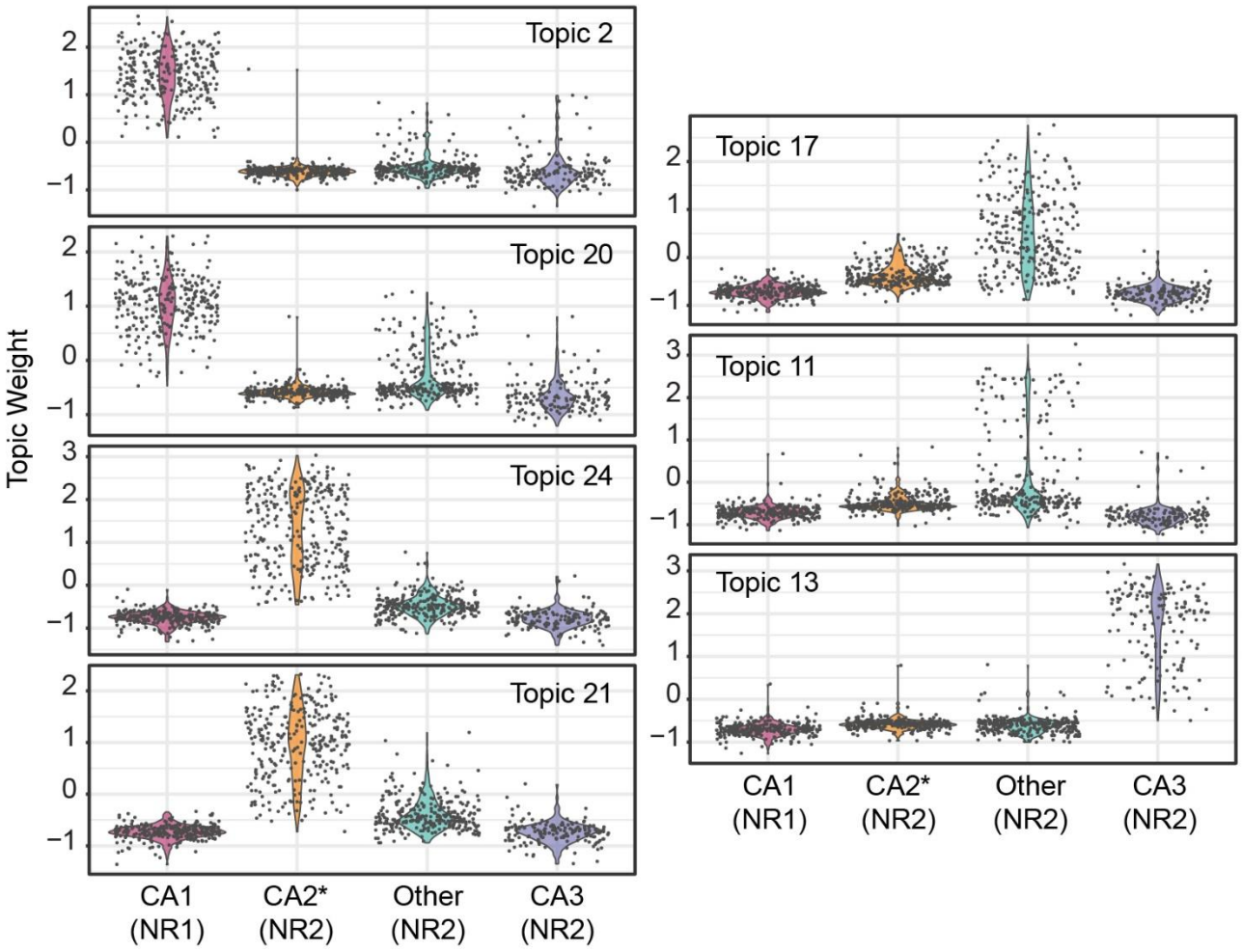


Supplemental Figure 13: Fresh vs. Frozen comparison for pyramidal neuron subclustering
UMAP coordinates as in Figure 3, with cell colors corresponding to fresh vs. frozen preparations.

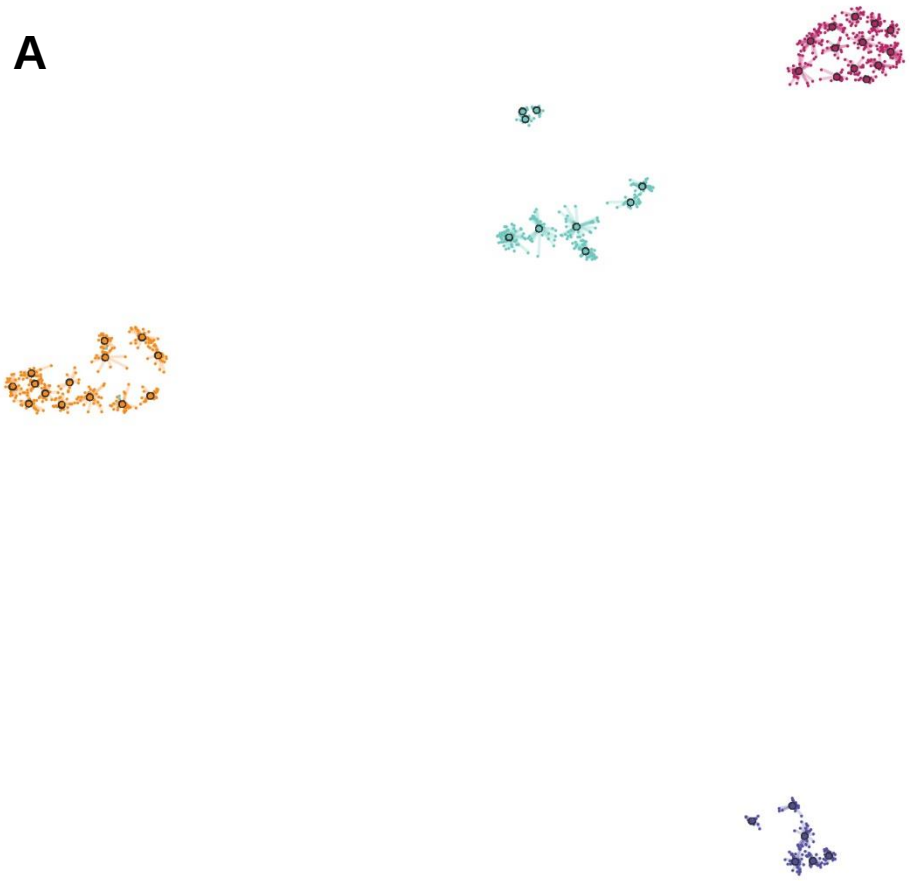


Supplemental Figure 14: Read plots for CA1, CA2, and CA3 genes

Read plots for marker genes associated with CA1, CA2, and CA3. Arrows indicate enrichment of signal present in the corresponding cell type cluster. Arrows denote peaks enriched in the corresponding cell population.

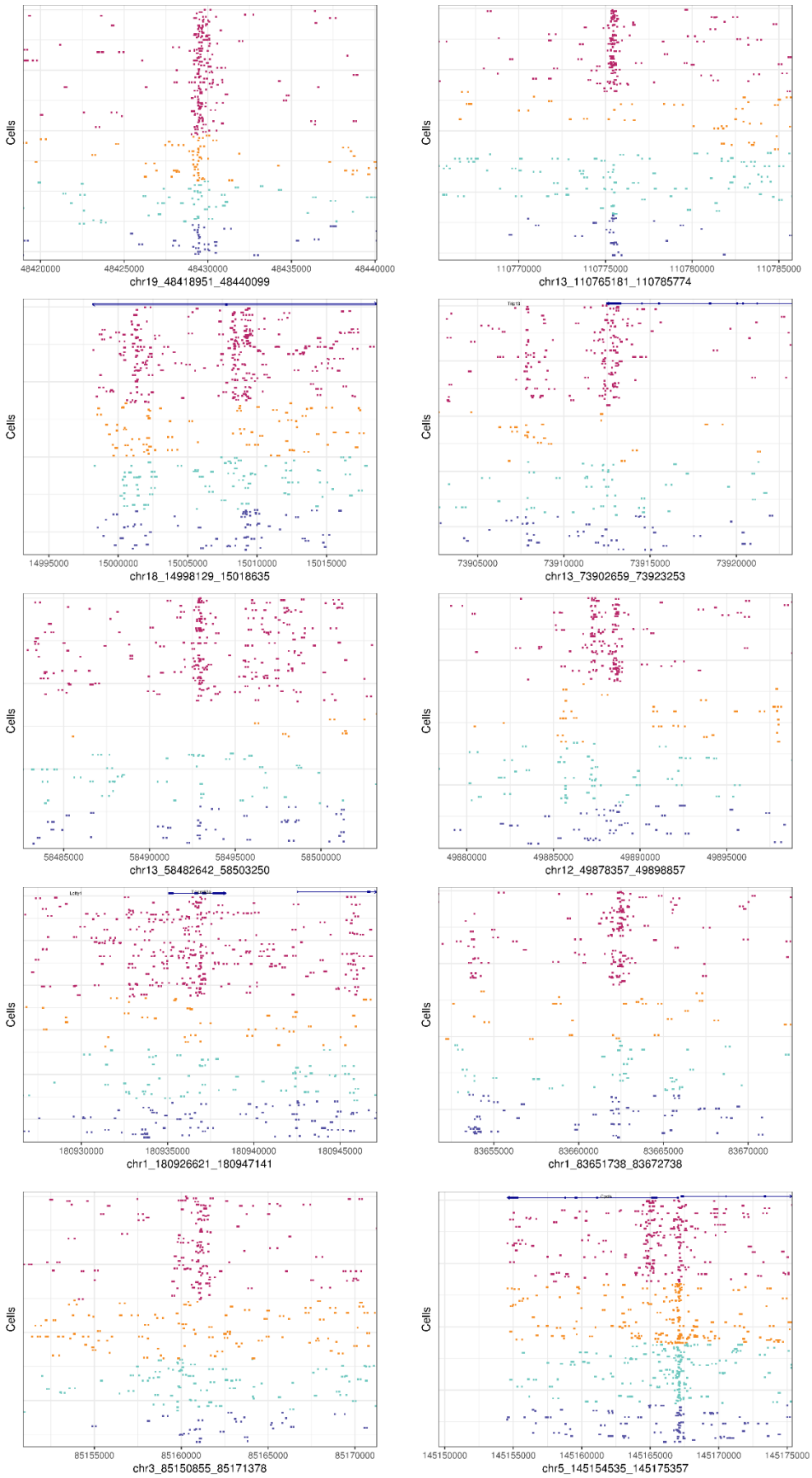


Supplemental Figure 15: Topic weights of cells in each pyramidal neuron cluster
Select cisTopic topics that exhibited cluster specificity are shown. The y-axis represents the weight of each topic for each cell by cluster identity (x-axis). Asterisk denotes putative assignment.



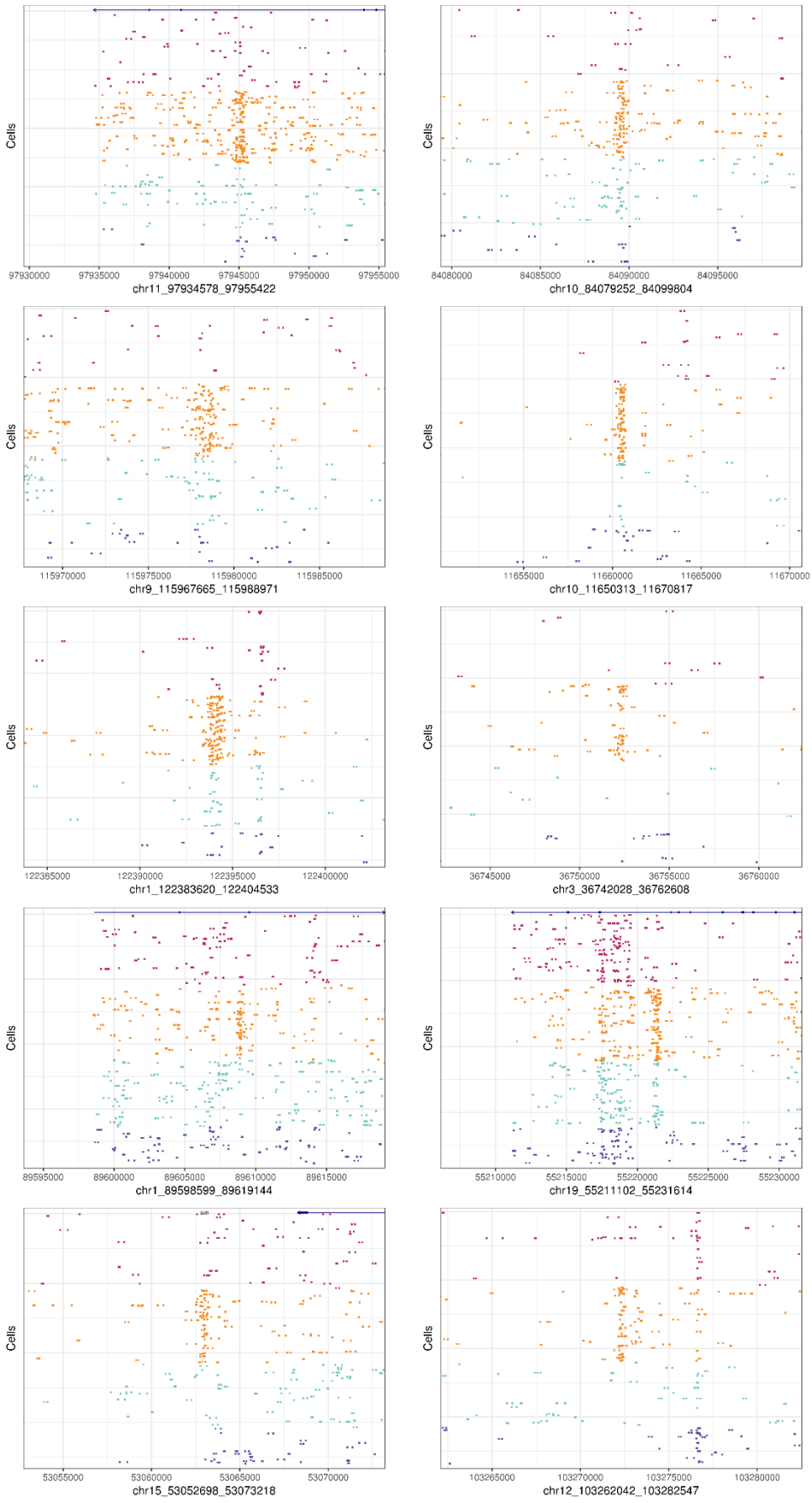
Supplemental Figure 16A: scitools aggregate-cells

B



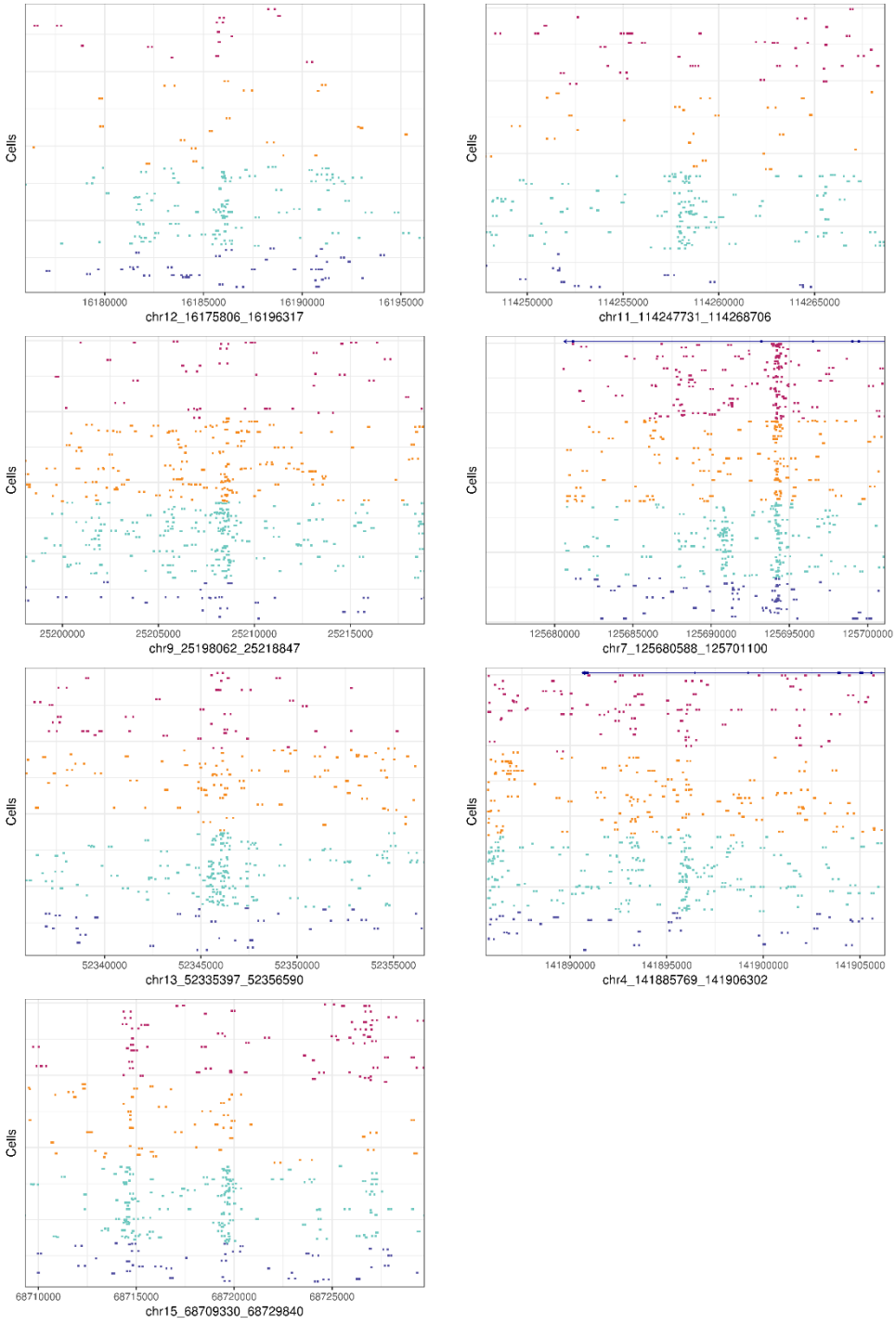
Supplemental Figure 16B: Top 10 DA loci for CA1 cluster

C



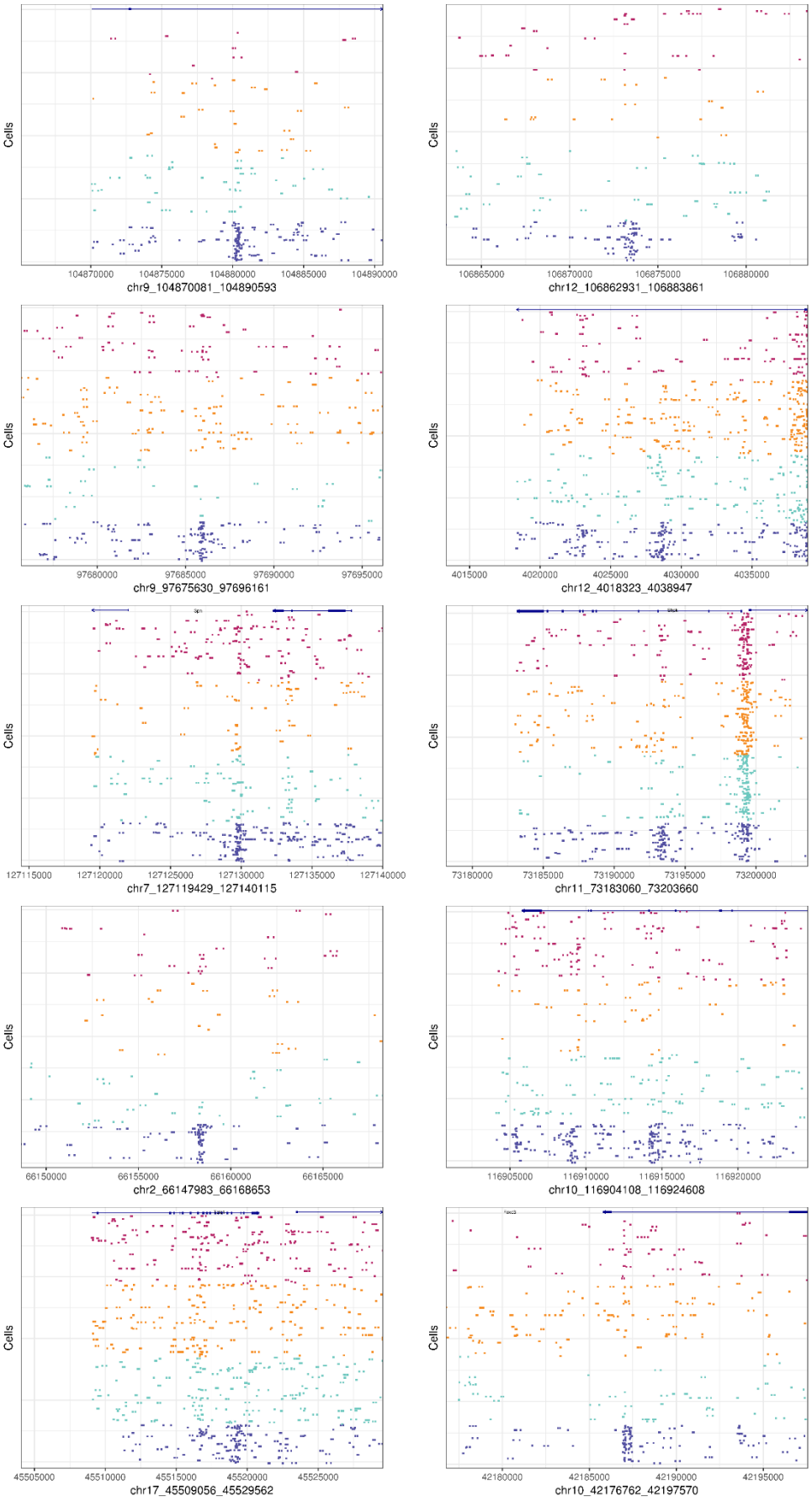
Supplemental Figure 16C: Top 10 DA loci for CA2 cluster

D

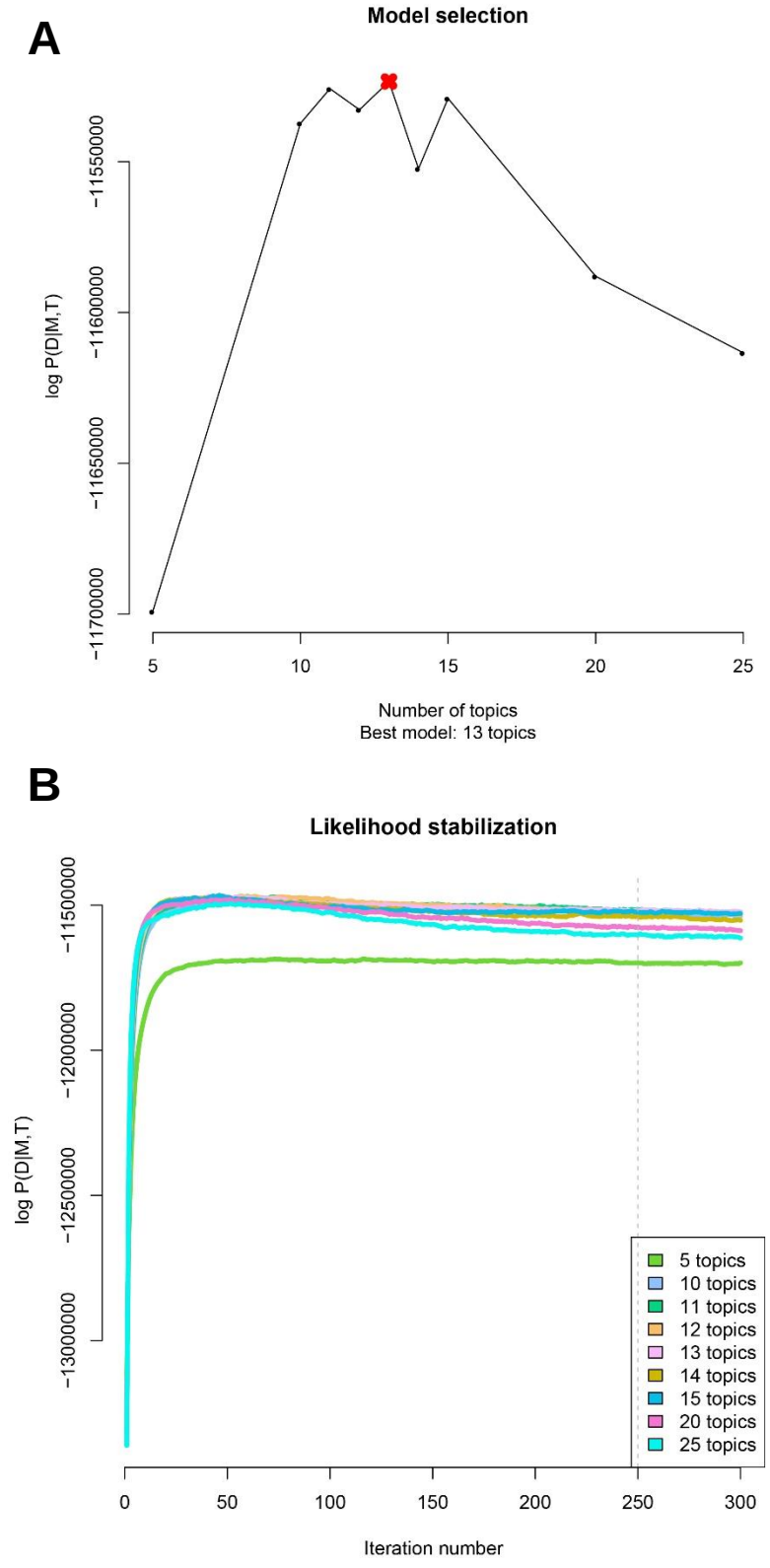


Supplemental Figure 16D: All 7 significant DA loci for Other/MC cluster

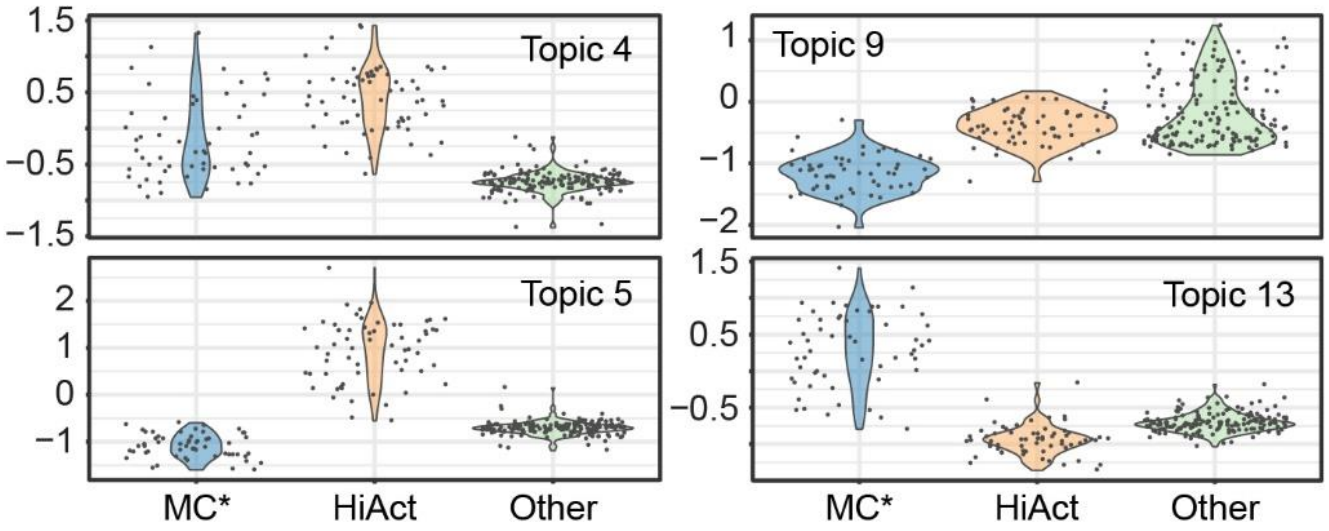
E



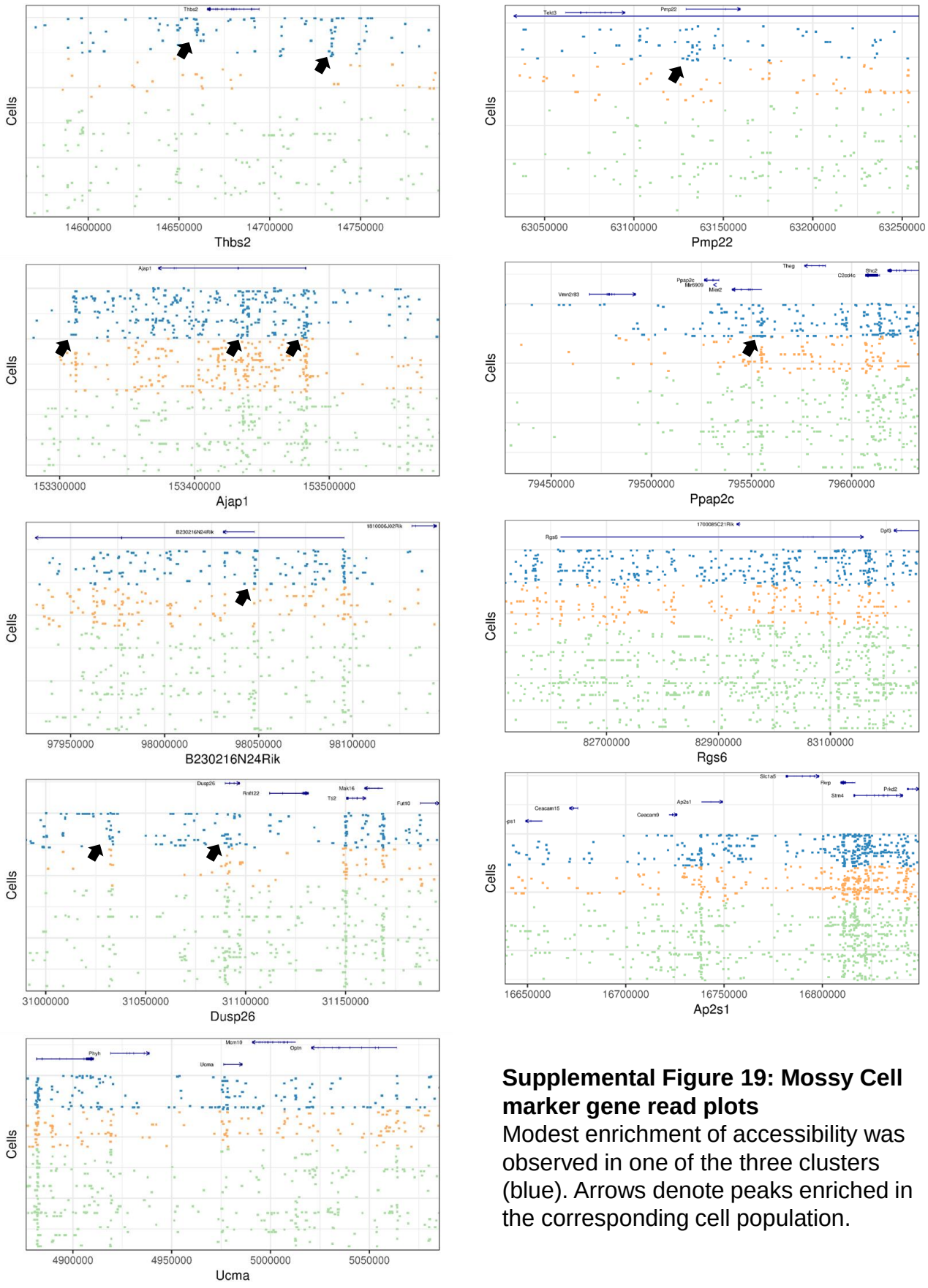
Supplemental Figure 16E: Top 10 DA loci for CA3 cluster



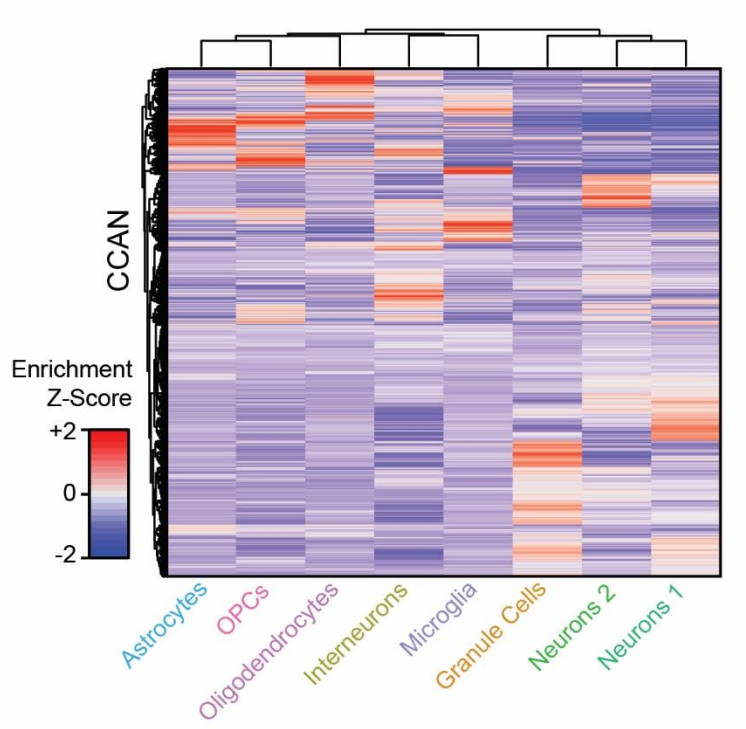
Supplemental Figure 17: CisTopic Topic Selection for the Other subclustering
(A) Multiple topics are run to optimize topic separation. 12 was determined to be the optimum number for subsequent analysis. (B) Stabilization of the model over iterations. All models stabilized by approximately 20 iterations, though 300 were run in all cases.



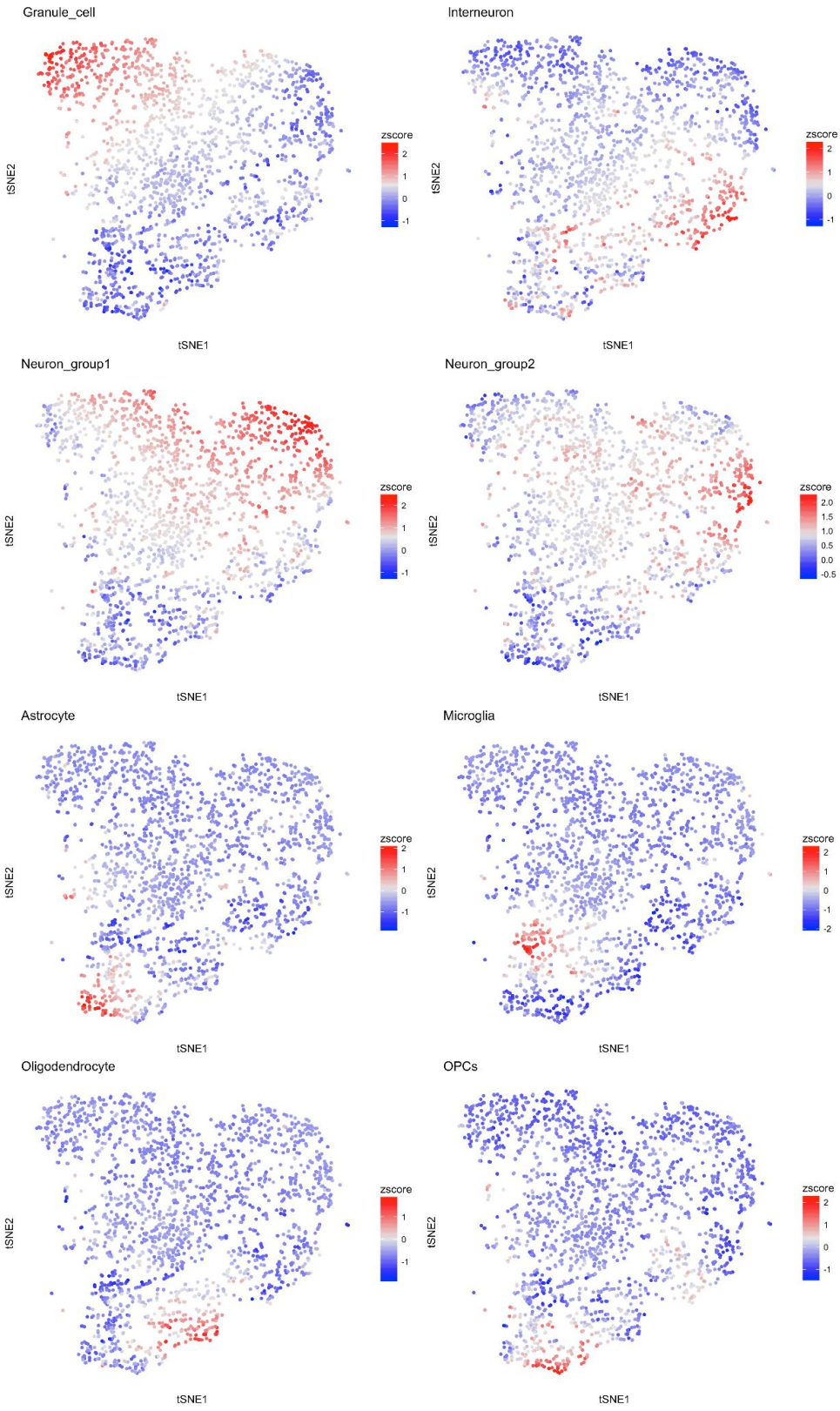
Supplemental Figure 18: Topic weights of cells in each pyramidal neuron subcluster Select cisTopic topics that exhibited cluster specificity are shown. The y-axis represents the weight of each topic for each cell by cluster identity (x-axis). Asterisk denotes putative assignment.



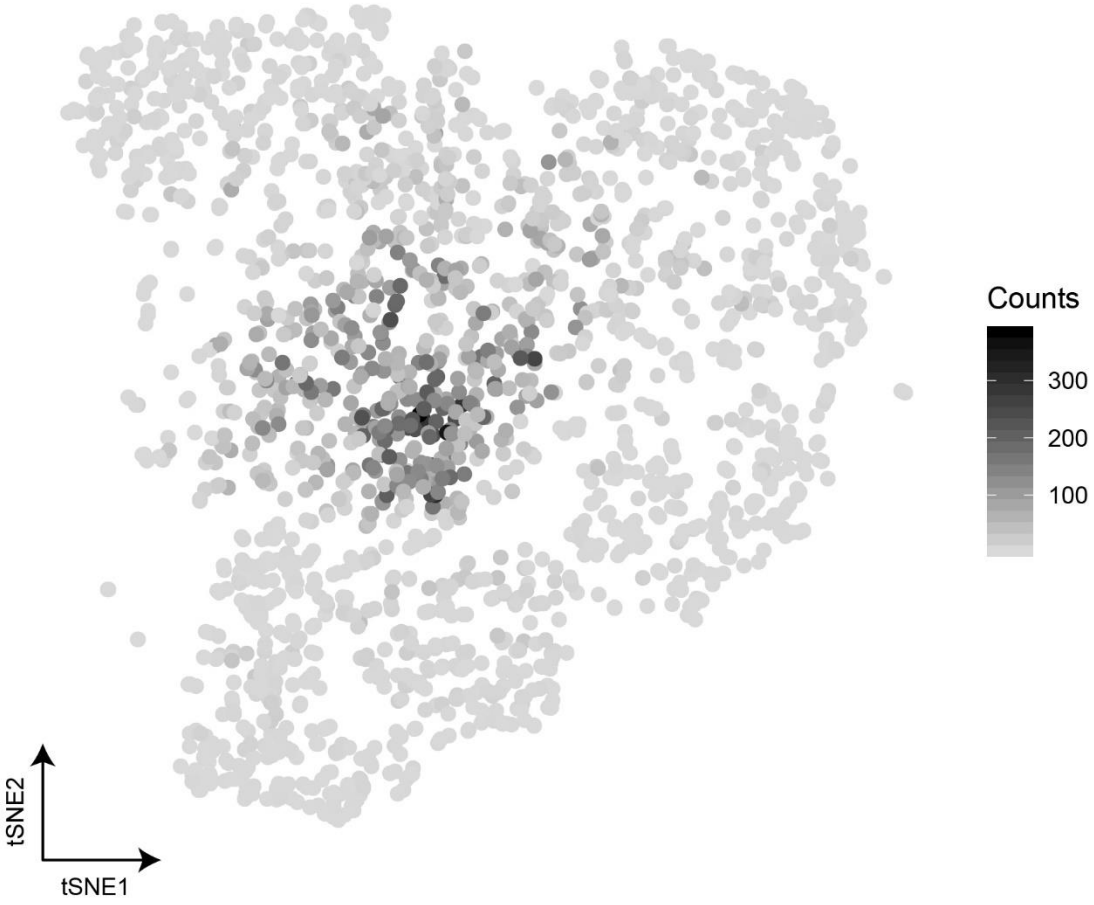
Supplemental Figure 19: Mossy Cell marker gene read plots
Modest enrichment of accessibility was observed in one of the three clusters (blue). Arrows denote peaks enriched in the corresponding cell population.



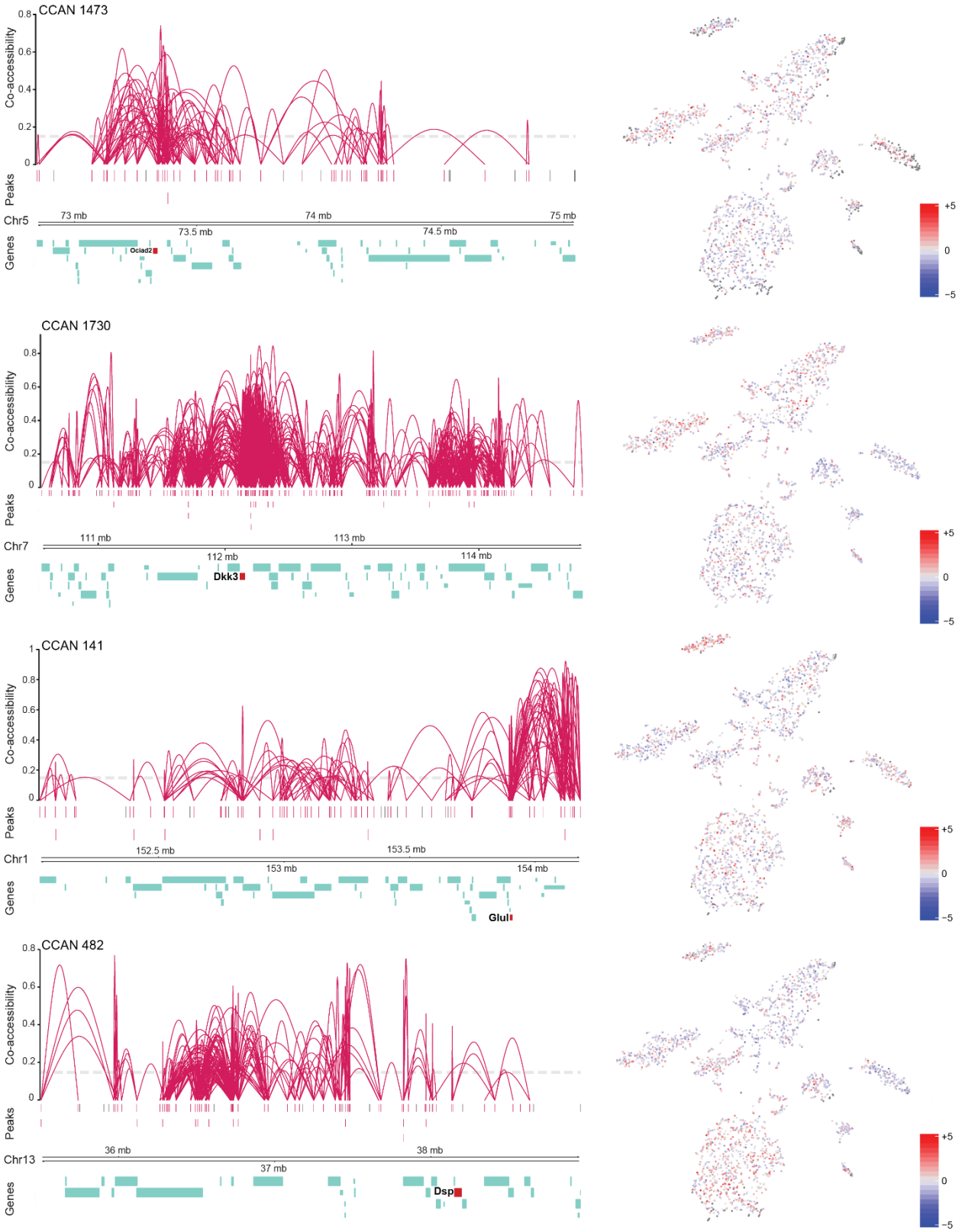
Supplemental Figure 20: Hierarchical clustering of CCANs by cell type enrichment
Hierarchical clustering of enrichment z-scores for peaks contained with in each CCAN with respect to cell type.



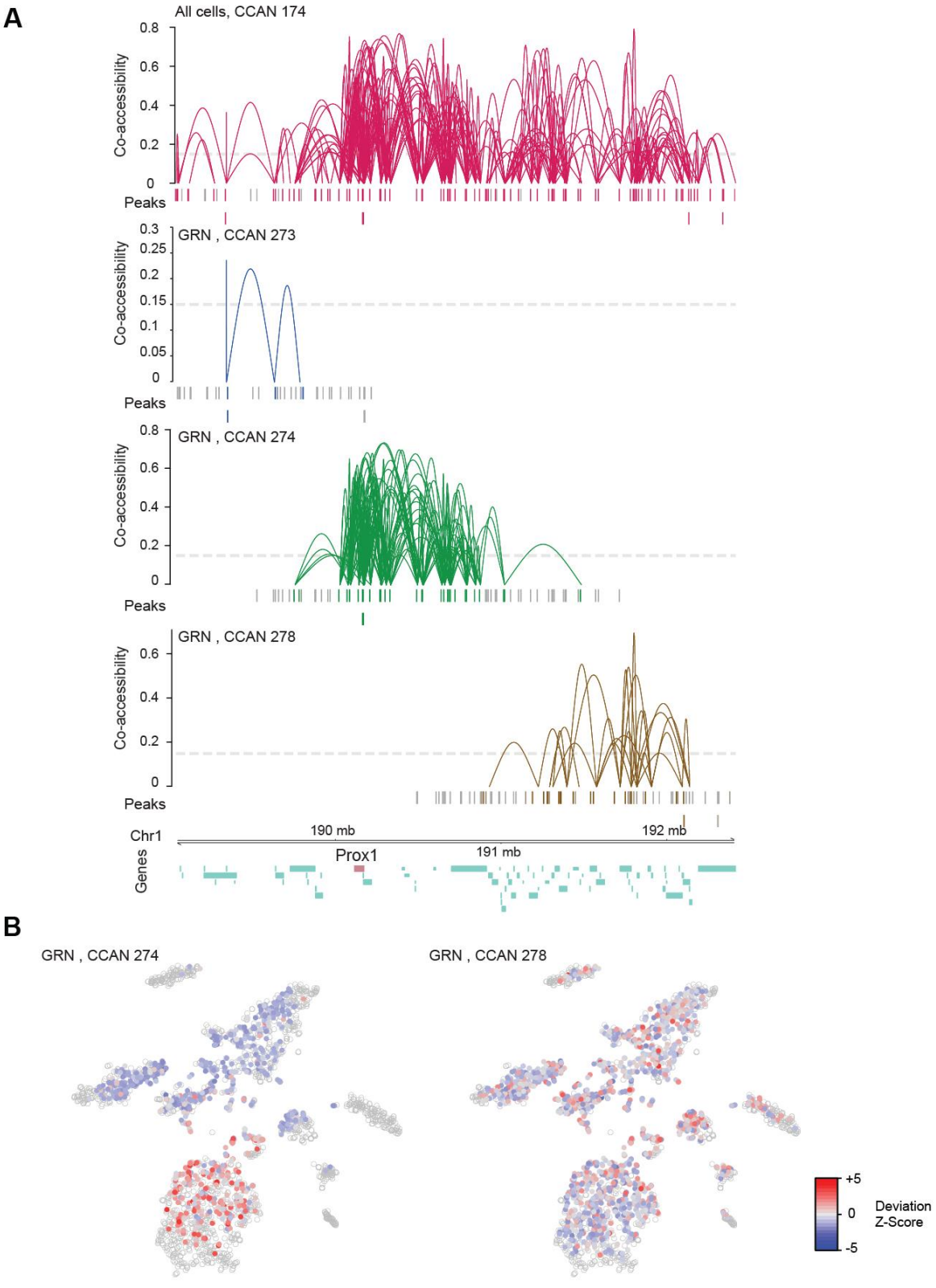
Supplemental Figure 21: CCAN cell type enrichments on tSNE projections
Enrichment for CCAN's within each cell type is plotted as in Fig. 3, but for each of the eight major clusters.



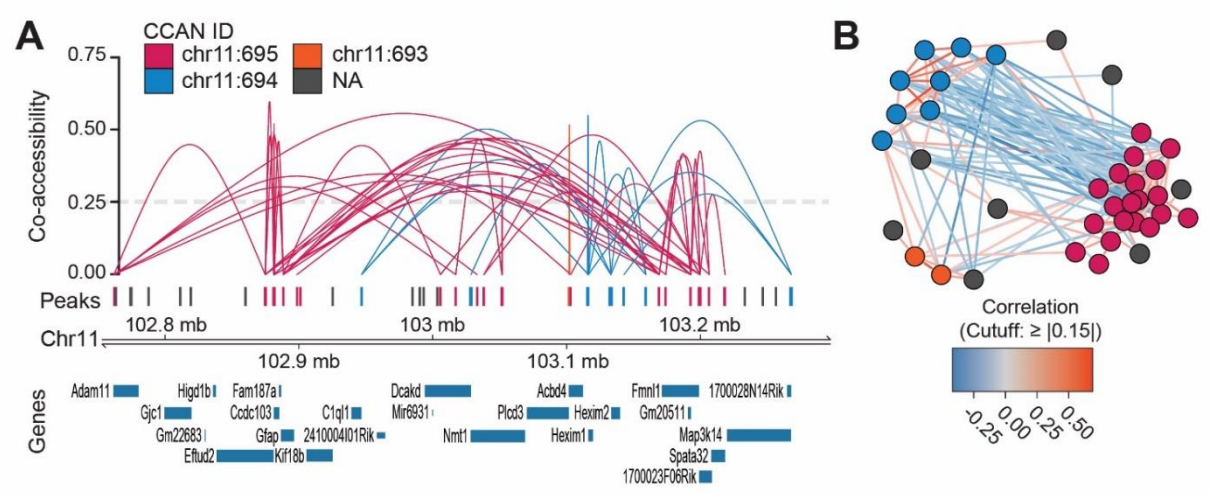
Supplemental Figure 22: CCAN peak counts
tSNE of CCANs by cell type specificity (as in Figure 3), but shaded by peak membership count. CCANs with greater numbers group towards the center and are less-cell type specific overall. This is likely due to modular CCANs where certain portions of CCANs are cell-type-specific that link up with other, more universally established CCANs, as in the Prox1 example.



Supplemental Figure 23: Additional cell-type-specific CCANs
Left: plots of the CCANs near cell-type-specific marker genes. Right: Accessibility enrichment across cells.



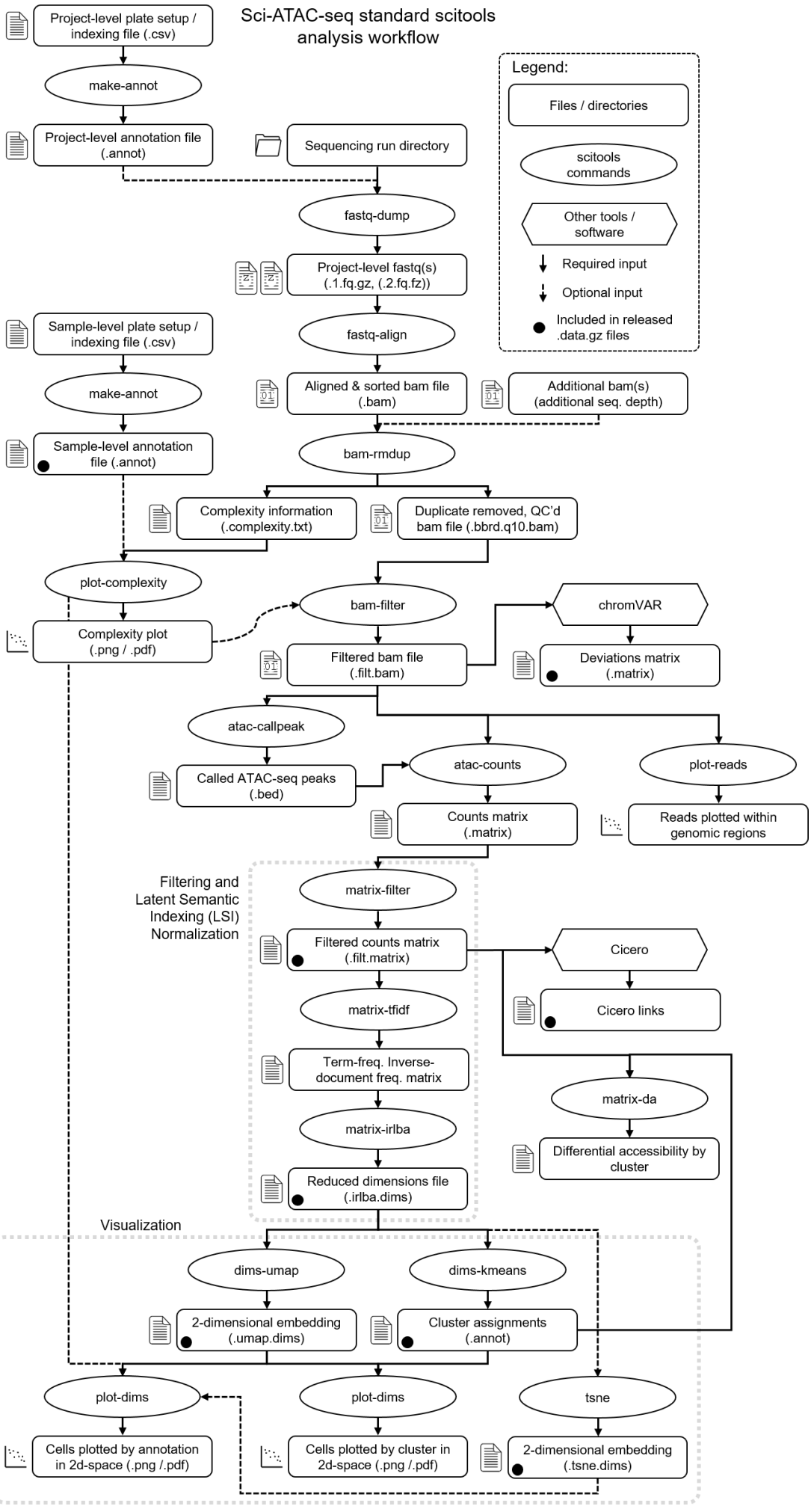
Supplemental Figure 24: Dissection of CCAN 174 (*Prox1*)
A) CCAN 174 as determined by Cicero performed on all cell types (top). The CCAN is large and likely contains links that are not cell-type specific. When Cicero is carried out only on the dentate granule cell population, the larger CCAN is split into three distinct networks. B) The cell-type specificity for the granule-specific CCAN (274) centered on *Prox1* is more accessible than the adjacent CCAN that does not exhibit cell type specificity (278).



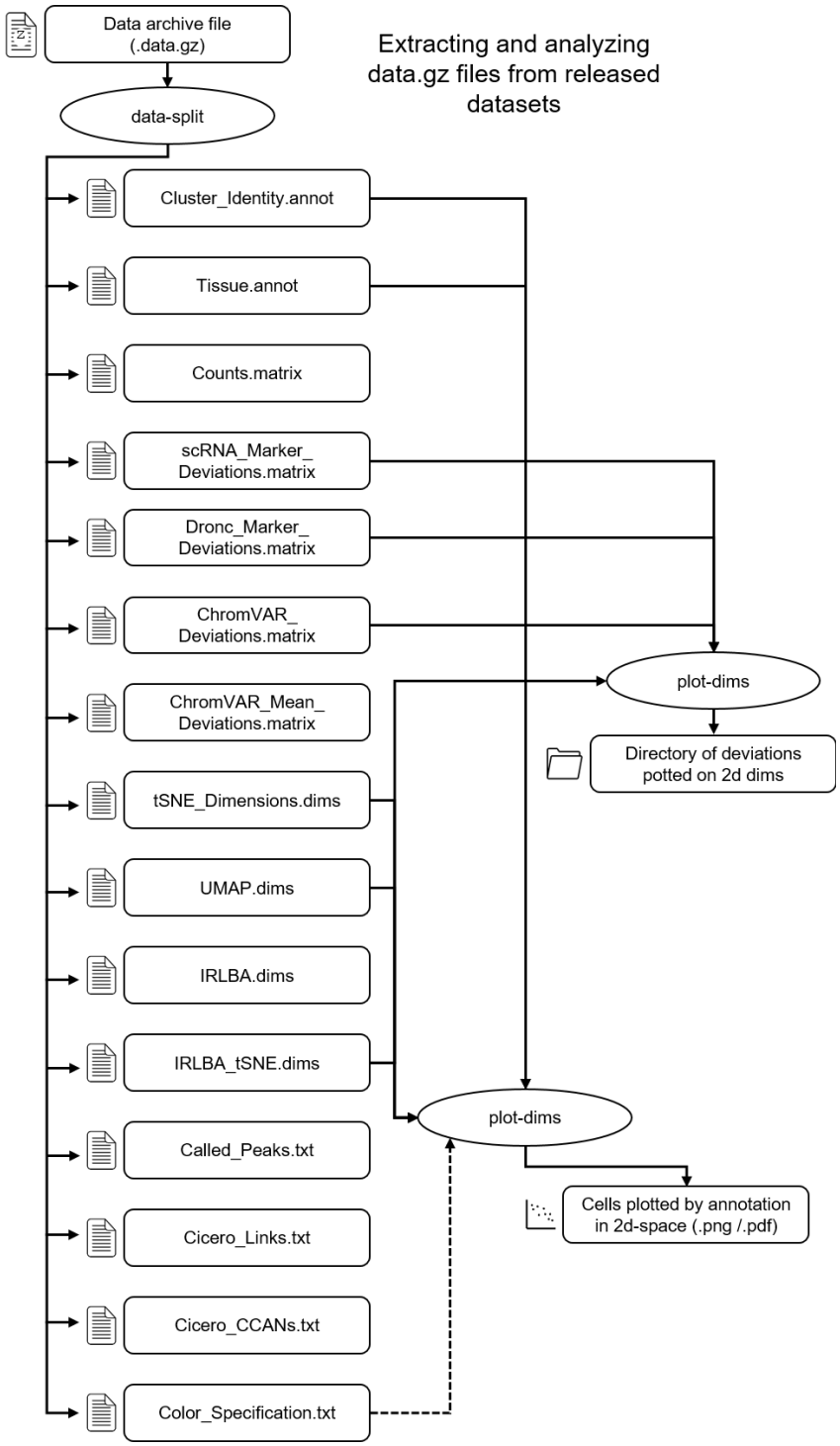
Supplemental Figure 25: Mutually exclusive CCAN

A) Example of a mutually-exclusive set of CCANs that includes the astrocyte marker gene Gfap. CCAN 695 (pink) and 694 (blue) overlap one another in genome positions; however, they are comprised of mutually-exclusive peak sets, suggesting two alternative chromatin conformations. B) The mutually exclusive CCAN peak sets have negative correlations with peaks from one another's network.

A


Sci-ATAC-seq standard scitools analysis workflow
Legend:
Files / directories
scitools commands
Other tools / software
Required input
Optional input
Included in released .data.gz files
Project-level plate setup / indexing file (.csv)
make-annot
Project-level annotation file (.annot)
Sequencing run directory
fastq-dump
Project-level fastq(s) (.1.fq.gz, (.2.fq.fz))
Sample-level plate setup / indexing file (.csv)
make-annot
Sample-level annotation file (.annot)
fastq-align
Aligned & sorted bam file (.bam)
Additional bam(s) (additional seq. depth)
bam-rmdup
Complexity information (.complexity.txt)
Duplicate removed, QC'd bam file (.bbrd.q10.bam)
plot-complexity
Complexity plot (.png / .pdf)
bam-filter
Filtered bam file (.filt.bam)
chromVAR
Deviations matrix (.matrix)
atac-callpeak
Called ATAC-seq peaks (.bed)
atac-counts
Counts matrix (.matrix)
plot-reads
Reads plotted within genomic regions
Filtering and Latent Semantic Indexing (LSI) Normalization
matrix-filter
Filtered counts matrix (.filt.matrix)
matrix-tfidf
Term-freq. Inverse-document freq. matrix
matrix-irlba
Reduced dimensions file (.irlba.dims)
Cicero
Cicero links
matrix-da
Differential accessibility by cluster
Visualization
dims-umap
2-dimensional embedding (.umap.dims)
dims-kmeans
Cluster assignments (.annot)
plot-dims
Cells plotted by annotation in 2d-space (.png /.pdf)
plot-dims
Cells plotted by cluster in 2d-space (.png /.pdf)
tsne
2-dimensional embedding (.tsne.dims)

B



Supplemental Figure 26: Scitools analysis and data access workflow
A) Detailed breakdown of all analyses performed by scitools in this study, and general best-practice guidelines for processing sci-ATAC-seq data using scitools. B) scitools data-split command to extract data files associated with this study from the GEO archive.

Supplemental Note 1: Data & Software Description

We have released our data in a format compatible with the scitools software suite, which can be found at https://github.com/adeylab/scitools. For more information on scitools commands and usage, refer to user manual provided at the GitHub site.

To extract associated data, run the following command:

```
$ scitools data-split [dataset].data.gz
```

This will produce a set of files for each of the data sets with self-descriptive titles:

File	InVitro	InVivo	Comb.	Description
Cluster_Identity				Cluster assignments
Tissue				Tissue assignments
Counts				Counts matrix (unfiltered)
scRNA_Marker_Deviations				Linarsson lab marker gene set deviations
Dronc_Marker_Deviations				Regev Lab Dronc-seq gene set deviaitons
ChromVAR_Deviations				Motif deviations
ChromVAR_Mean_Deviations				Mean motif deviations by cluster
tSNE_Dimensions				LSI-tSNE 2D dimensions
UMAP				LSI-UMAP 2D dimensions
IRLBA				IRLBA 50 dimensions
IRLBA_tSNE				IRLBA-tSNE 2D dimensions
Called_Peaks				Called peaks
Cicero_Links				Cicero peak correlations (all)
Cicero_CCANs				CCANs from Cicero links
Color_Specification				Color coding for plotting

Scitools commands can then be run to reproduce figures in the manuscript. Below are several examples:

1) Plotting tSNE projections of cells:

```
$ scitools plot-dims -A [dataset].Cell_Type.annot [dataset].tSNE.dims
```

2) Plotting the chromVAR deviation scores onto tSNE dims:

```
$ scitools plot-dims -M [dataset].ChromVAR_DevZ.matrix
[dataset].tSNE.dims
```

The counts matrix is also produced by the data-split command which can be used to run other scitools functions, for example:

1) Performing latent semantic indexing:
   a. Filter matrix:

```
$ scitools matrix-filter -C 1000 -R 50 [dataset].Counts.matrix
```

   b. Tfidf transform:

```
$ scitools tfidf [dataset].Counts.matrix
```

   c. Perform LSI:

```
$ scitools lsi [dataset].tfidf
```

d. Perform tSNE:

```
$ scitools tsne [dataset].tfidf.LSI.matrix
```

2) Alternative dimensionality reduction strategies:

a. On the tfidf matrix, perform irlba:

```
$ scitools irlba [dataset].tfidf
```

b. Perform tSNE:

```
$ scitools tsne [dataset].tfidf.irlba.matrix
```

To list additional commands that can be run on the datasets, run the following:

```
$ scitools list
```

Additional Supplementary Data files contain bed files for topics associated with pyramidal neuron subclustering and homer motif calling results on the respective topics along with annotation files for subclustering, UMAP dimensions, and cisTopic weight matrices.

Supplementary Data – CisTopic_Subclustering_1.rar

Supplementary Data – CisTopic_Subclustering_2.rar

Supplementary Protocol – sci-ATAC-seq

Reagent Preparation:

Preparing NIB (50mL):

- Note: For most preparations, 15mL is fine. NIB seems to stay for ~1 month.

Final Concentration	For Construction
10 mM Tris, pH 7.5	500 uL of 1M Tris, pH7.5
10 mM NaCl	100 uL of 5M NaCl
3mM MgCl2	150 uL of 1M MgCl2
0.1 % Igepal	500 uL of 10% Igepal
0.1 % Tween	**500 uL of 10% Tween**
ddH20	to 50 mL
Protease Inhibitor (Only if dealing with tough tissue)	2 tablets

Tissue preparation:

Mechanical dissociation in NIB of Flash-frozen brain tissue

- For each 1.5mL tube containing flash-frozen tissue:
  - Resuspend in 100 uL ice cold NIB.
  - Pipette up and down to break up cells. (~10-20 times)
  - Incubate on ice for 10 minutes.
  - Add another 100 uL ice cold NIB
  - Pipette up and down to break up cells (this time until all flaculents dissociate)
  - Run total volume through 100um filter.
  - Add 3uL DAPI (@5mg/mL)
  - Incubate for 3 min then load sample into flow sorter.
  - Store on ice until ready.
  - Sort.

Transposase should be prepared according to Picelli *et. al.* Genome Research 2014 and diluted to 17.9 uM

Anneal indexed oligos to ME':

For each i5 barcoded oligo:

a. 12.5 uL 100 uM i5 barcoded oligo
b. 12.5 uL 100 uM MEr oligo
c. 53.125 uL 2x Tn5 dialysis/dilution buffer

For each i7 barcoded oligo:

a. 8.5 uL 100 uM i7 barcoded oligo
b. 8.5 uL 100 uM MEr oligo
c. 36.125 2x Tn5 dialysis/dilution buffer

Anneal: 95C, 5 min, then -5C/2 min to 20C
result: ~78 uL of 16 uM i7 or i5/MEr

Plate indexed oligos:

III. In 96 well plate:
a. Add 5 uL of i5/MEr (16 uM) to respective wells
b. Add 5 uL of i7/MEr(16 uM) to respective wells
result: 10 uL at 8uM/well

Load uniquely indexed tn5s:

Tn5 + salt:
Tn5 Stock @ 17.9 uM.
For every: 12 uL of Tn5 stock, add 1.5 uL 5M NaCl
Then, add 12 uL of salt-adjusted Tn5 to each oligo well
result: 8 uM Tn5 @ 500 nM Nacl
Incubate to assemble Tn5/oligo mixture for 1 hour @ 25C

CRITICAL: Freshly annealed oligos must be used. Annealed oligos that have been subjected to a freeze-thaw cycle will not form viable complexes. Once compelxed, transposomes can be stored in the freezer and can be used for up to 6 months without an observable decline in library preparation quality.

For first sort preparation:

- Prepare Plates:
  - Per well use 5uL NIB and 5uL TD Buffer (2X) from Illumina.
  - Sort 2K-5K nuclei into each well using FACS protocol (below).
    - Note, numbers may vary by prep, but ensure that numbers *within* a prep are constant.
      - I.e. all wells receive the same number of cells.

First Sorting Protocol (Sony SH800):

- Before turning on the Sony SH800 Cell sorter:
  - Check black metal Sheath fluid container.
    - If liquid is low, fill with sheath fluid (in jug dispenser)
    - If jug dispenser is low, dilute out 50X Sheath fluid stock (in boxes near jug dispenser) with DI water
    - Ensure sheath fluid container is firmly closed before continuing
  - Turn on air (in back room, behind door)
    - Check pressure rises to 80 psi.
- Start Sony sorter and wait until it gets to standby.
  - Open Cell sorter software
  - Scan in QR code using the webcam for sorter chip. For most sciATAC experiments, chip should be 100um size.
- Calibration and System Check
  - Blue calibration beads are located in liquid dropers in the fridge
    - These are light sensitive
    - A working stock is generally available contained in a sorting tube in the same box
    - Choose 405 and 488 for DAPI sorting
  - Continue through the calibration and system check wizard.
  - Wait for fluid check, to complete (~30 min or so)
  - If check or calibration fails, follow protocol in the manual
- Set Up Experiment
  - Name experiment and turn on FSC, BSC, FL1 (set to DAPI in drop-down menu)
    - Make sure height and area are selected for all of them
  - Check instrument tank levels:
    - If DI water tank is low:
      - Open side door of the Sony (left side when facing its front)
      - Unclip the large plastic bottle from the lines (this is the DI water bottle)
      - Refill with DI water from dispenser jug.
      - More DI water is available in lab next to sorting room (use Millipore Purified water)
        - Sorting room key also works on that lab door if locked
    - If Waste tank is too high:
      - Slide out waste tank from bottom shelf under the sorter.
        - The waste tank has a red line leading into it.
      - Unscrew the cap and move to the back room.
      - Add bleach to the waste liquid and pour down sink
  - Perform water de-bubble

Cytometer > Settings > Maintenence > DI filter-debubble

Check champer with samples and lasers is wiped down and clean

If any salt residue is visible, wipe down chamber and magnetic plates with DI water

Ensure everything is dried by kimwipe (water droplets will messup the charge of droplets)

Set Cytometer chambers to 5C

Cytometer > Collection > Sample 5C

Cytometer > Collection > Collection Chamber 5C

Set Up Gates (y by x)

BSC-A vs. FSC-A (All Events): Dot plot

- Use a oblong gate (GATE A) to include almost all events, point here is the exclude bottom left high density of debris/false events

DAPI-H vs. DAPI-A (GATE A): Dot plot

- Should see a strong diagonal bias in this plot
- Use a polygon gate (GATE B) to select events along this diagonal.
- If you switch to a density plot, you should be able to see a population forming at ~10^4.5 along the diagonal (these are the singleton nuclei)

BSC-A vs. DAPI-A (GATE B): Dot plot

- Should see two or more sweeping patterns along the dot plot.
  - The first swoop should be wider, this is debris
  - The second swoop should be tighter and more defined, this is nuclei
  - Any subsequent swoops are doublets, triplets etc. these should be less populated and smaller
- Use a polygon gate (GATE C) to select for the nuclei population (second swoop)
  - SORT ON GATE C

Events vs. DAPI-A: (GATE C) Histogram

- This plot should show a strong and tight peak, with the nuclei
- You may see a degenerative smaller outward peaks forming which are the doublets, triplets etc.
  - If this is the case, make sure your GATE C is tight around the single nuclei swoop

Add sample to sample chamber

Assign tube based on biological source

Test the rate and the sample to set gates.

Sample pressure is by default 4, this can be increased to 6 to increase the rate of events.

- Do not exceed 7, the sample line might pop off due to pressure

Set to sort on 2-5K nuclei/well (dependent on experimental design)

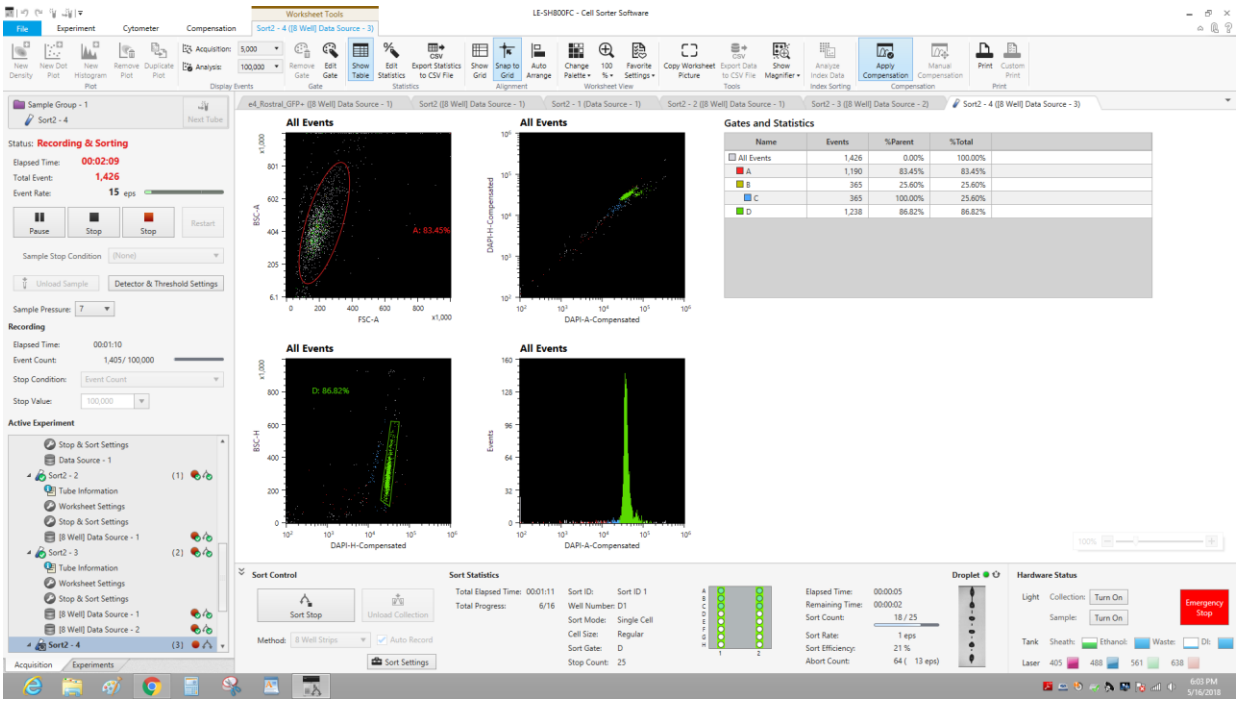
Select sort gate (GATE C)

Load collection tubes and perform sorting.

Sort using the "Normal" setting, not "single-cell" (this step doesn't need as much precision)

Example sorting set up.

Note: some of the axes and gates are different than described in the above protocol, but looks should be similar.



- Following sorting
  - Perform a DI rinse.
  - Put sorting in standby mode and proceed to tagmentation.
    - Standby mode relieves the pressure on the machine.
      - Cytometer > Options > Advanced Options > Standby

Tagmentation:

Spin down plate immediately after sorting finishes.

Store covered and on ice until the full plate is complete.

Note: to cover strips as they finish sorting, I tend to tear up aluminum plate covers into single-strip width.

- Add 1uL 8uM uniquely indexed transposome to each well
- Seal plate and incubate at 55C for 15 minutes with gentle shaking (on eppendorf thermomixer, ~300 rpm).
- Plate plate on ice immediately to stop reaction. Keep samples on ice to prevent over-transposistion and nuclei lysis.
- Pool all wells, while maintaining everything on ice.
- Add 2uL/per mL pooled sample of DAPI (5mg/mL) and bring to sorter for second sort.

Second Sort Plate Setup:

- Add 0.25 µL 20 mg/mL BSA, 0.5 µL 1% SDS, 7.75 µL nuclease-free water to each well
- Add 2.5 uL of 10 uM i5 Indexed PCR Primer and 2.5 uL of 10 uM i7 Indexed PCR Primer prior to sort.

Second Sort Protocol:

- Thaw RT-PCR reagents on ice before second sort
- Sort **X** nuclei per well (X is dependent on number of wells tagmented in first sort, as a linear trend)
    - **1 plate = 22 nuclei/well**
    - **1.5 plates = 33 nuclei/well**
    - **2 plates = 44 nuclei/well etc...**

- Using same gates as first sort:
    - Sort **X** nuclei per well with modified sort settings:
        - "Single cell" rather than "Normal"
        - This leads to a higher abort count (less efficient sorting) but is far more precise

  Keep sorted samples on ice.

- After sorting:
    - Perform a Shut down rinse on the machine using 15 mL volumes
        - Cytometer > Shut down Rinse
        - Perform both the bleach cleaning and the DI rinse
            a. Select option to store DI water in machine upon shut down
        - Follow the instructions on the shutdown wizard
- Once machine is turned off:
    - Turn off the air in the back room.
    - Release pressure using the toggle on the air line
    - Release pressure from the machine by pulling on the top ring of the sheath fluid container.
    - Ensure lights are off and door is locked.

Transposase Denaturation and PCR Setup:

- Spin down plate
- Cover plate and hold on ice until sorting is complete.
- Denature Tn5 in a Thermocycler by incubating at 55C for 15 minutes
- To each well add: 7.5 µL NPM, 4 µL nuclease-free water, 0.5 µL 100X SYBR Green

PCR:

Perform PCR on a real-time thermocycler

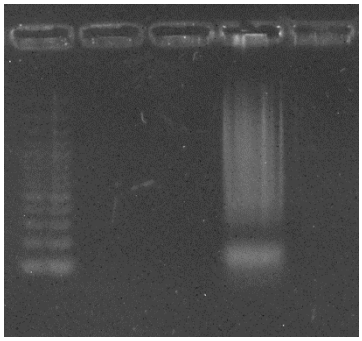
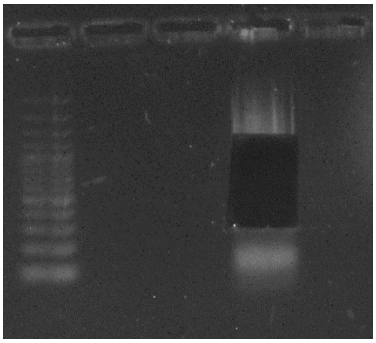
Temperature (C)	**Time**	
72	5 min	
98	30 sec	
98	10 s	
63	30 s	
72	1 min	x30 Cycles
	PLATE READ	
72	10 s	

Pull plates once a majority of wells are mid-exponential phase
Following PCR, plates can be stored in Post-PCR Fridge at 4C until cleanup.

- Library Clean up and Quantification:
  - Well Pooling:

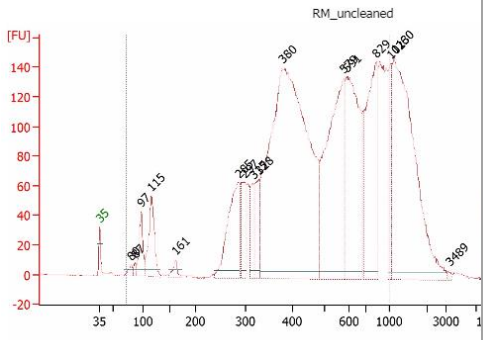
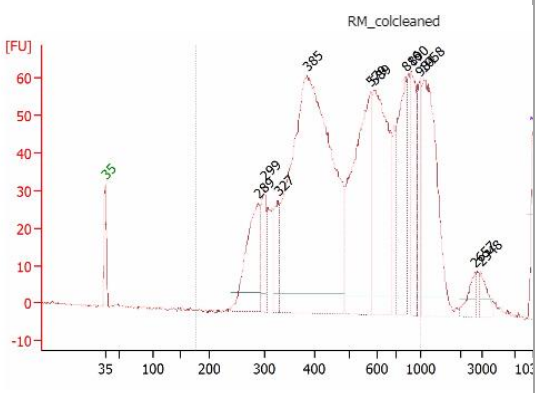
    Pool 10uL for each well.
    Ran full pool volume through Qiaquick PCR column following manufacturer's protocol.
    Eluted in 32 uL 10mM Tris HCl pH 8
    Quantified 2 uL in 2:200 dsDNA HS Qubit assay
    - This is to ensure that the library amount will be visible on a gel.
  - Gel Extraction Clean-up (Optional – perform if adaptor band is present):

    Loaded 10uL of Library pool (+2uL 6X Loading Dye) in 1% Agarose made with SeqPlaque low-melt agarose.
    Ran sample for 45min at 110V
    Ran 100bp Plus Ready-to-Load Generuler Ladder (5uL)
    Example Gel Cut: Note the adapter dimer at ~110bp which was excluded in the cut.

Pre-cut	**Post-cut**
	

    - Gel extracted from ~250-1.3Kb and performed Qiaquick Gel Extraction protocol following manufacturer's instructions.
    - Eluted in 25 uL 10mM Tris-HCl, pH 8
    - Quantified with 2uL:200uL total volume Qubit Broad Range
    - Use Qubit reading to dilute to ~4ng/uL and run 1uL on HS Bioanalyzer chip
    - Run 1uL of sample at 4ng/uL library dilution on Bioanalyzer High-Sensitivity DNA chip (following manufacturer's protocol)
      - Quantify library from the range of 100-1000bp
      - Dilute this down to 1nM concentration for sequencing.

- *Example DNA HS Bioanalyzer Traces: Note the 97 and 115 bp peaks of adapters in the Pre-Cut Sample.*

Pre-cut	Post-cut
	

Custom Nextseq500 Chemistry Protocol

Thawing Reagents

The Reagent Cartidge:

Ensure to take the correct kit for capacity (Mid/High) and number of cycles (150/300). Make sure it is a v2 kit.

Remove from and thaw at 4°C overnight (at least 18 hours) or in a RT diH2O bath for 1 hour.

Once thawed, cartridge can be stored at 4°C for a day, if need be.

The Flow Cell:

Remove from Pre-PCR O'Roak Lab fridge and allow to reach room temperature, while still covered in packaging (~30min)

i. Note: flowcell is light-sensitive and should remain in packaging until sequencer is being loaded.

- **The HT1 Buffer:**
  i. These boxes are not cycle/capacity specific for kits, so any will work.
  ii. Open box and thaw HT1 Buffer in diH2O until fully liquid.
    1. Following full thaw, store on ice until ready to use.
- **1N NaOH Stock Solution:**
  i. Thaw at RT

- **Custom Sequencing Primers:**

Primer Identity	Oligo Name
Read 1 Primer	AA221_CPT_R1_Primer
Read 2 Primer	AA222_CPT_R2_Primer
Index 1 Primer	AA223_CPT_I1_Primer
Index 2 Primer	AA224_CPT_I2_Primer

Thaw at RT

Dilutions

- Make 1mL of fresh 0.2N NaOH from stock.
  a. 200uL 1N NaOH + 800uL ddH2O
- Dilute libraries based on Bioanalyzer quantification from 200-1000bp and generate final pool.
  a. 20uL at 1nM concentration.
  b. If any libraries are being spiked in, ensure volumes are correct for final pool
    i. e.g. 5% spike in means 1uL of the final 20uL 1nM Library Pool should be the spike-in library
  a. Combine 20uL 1nM Library Pool + 20uL 0.2N NaOH
  b. Vortex and spin down.
  c. Incubate at RT 5 mins
  d. Add 200uL 200mM Tris-HCl pH 7.5
  e. Vortex and spin down.
  f. Add 940 uL HT1
    i. This dilutes to 20pM
  g. Place 20pM library pool on ice until ready to proceede.
- Dilute library to loading concentration.
  a. For sciATAC standard assays, we load at 1.1pM
  b. 82.5uL 20pM Library Pool + 1417.5 uL HT1
    i. Total volume is now 1.5mL at 1.1 pM
    ii. Vortex to mix and store on ice.

- Dilute primers to loading concentration.

Final Diluted Reagent	Primer Identity	Oligo Name	Stock Concentration	Loading Concentration	Volume of Stock Oligo	Volume of HT1 Dilutant	Total Volume
Custom Read 1 Primer Dilution	Read 1 Primer	AA221_CPT_R1_Primer	100uM	0.6uM	9uL	1491uL	1500uL
Custom Read 2 Primer Dilution	Read 2 Primer	AA222_CPT_R2_Primer	100uM	0.6uM	9uL	1491uL	1500uL
Custom Index Primer Dilution	Index 1 Primer	AA223_CPT_I1_Primer	100uM	0.6uM (when combined)	18uL	1482uL	1500uL
Custom Index Primer Dilution	Index 2 Primer	AA224_CPT_I2_Primer	100uM	0.6uM (when combined)	18uL	1482uL	1500uL

a. Combine full volumes of Index 1 Primer and Index 2 Primer Dilutions to a total of 3mL Custom Indexing Primer Pool

Loading the Reagent Cartridge

- Ensure the Reagent cartidge is fully thawed.
  a. Without inverting the cartridge, lift up and check underneath. Tilt cartridge back and forth to ensure reagents are flowing as liquid and not frozen.
- Using a sterile pipette tip, puncture the aluminum seals for cartridge positions 7,8,9, and 10.
- Load the specified dilution and volume into each position.

Cartridge Position	Reagent	Total Volume
7	Custom Read 1 Primer Dilution	1500 uL
8	Custom Read 2 Primer Dilution	1500 uL
9	Custom Index Primer Dilution	3000 uL
10	Library Dilution	1500 uL

Loading the Nextseq

- Load up the Illumina Control Software.
- From the Home Screen press the "Sequencing" button and follow the set-up wizard to load the sequencer.
- For sciATAC runs:
  a. Use a custom protocol:
     i. For Mid Capacity 150 kits: "NextSeq CPT 10bp Mid"
     ii. For High Capacity 150 kits: "NextSeq CPT 10bp High"
  b. Select option: "Paired End"
  c. Set number of cycles and custom primer options:

	Read 1	Read 2	Index 1	Index 2
Cycles	50	50	18	18
Custom Primer?	yes	yes	yes	yes