

Counterfactual Confounder Adjustment for Single-cell Differential Expression Analysis

Yongjin Park

Results

Cell type annotation



PDF

PsychENCODE's marker gene selection might be imperfect since some “marker” genes are shared by different cell types. However, it has enough discriminatory power. Using this initial dictionary of marker genes, we can start analyzing 10x data, which consist of these three files:

- Sparse count matrix:

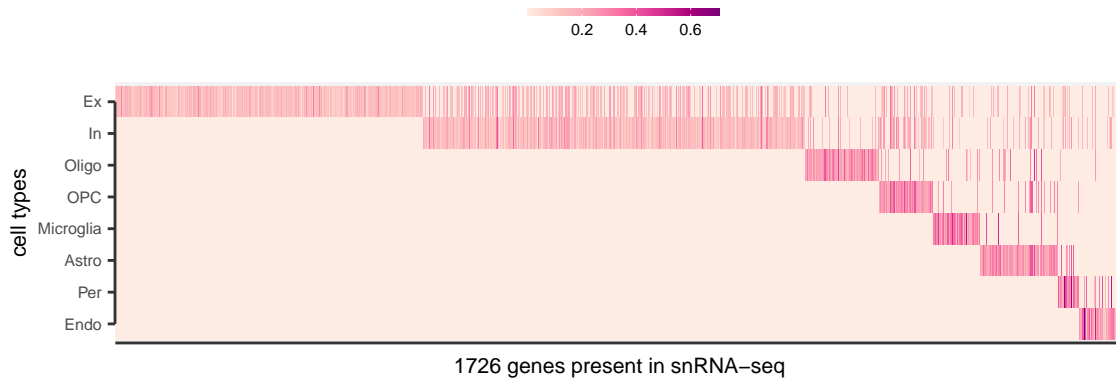
```
## %MatrixMarket matrix coordinate integer general
## 17926 70634 119696200
## 56 1 1
## 139 1 1
## 270 1 1
## 324 1 1
## 375 1 1
## 389 1 1
## 420 1 1
## 557 1 1
```

- Cell barcode list:

```
## AAACGGGAGATCCCGC.1
## AAATGCCTCCAATGGT.1
## AACCATGTCAGTGCAT.1
## AACCATGTCTGTACGA.1
## AACCGCGTCCGCATAA.1
## AACGTTGGTTCAGGCC.1
## AACTGGTGTACCGGCT.1
## AAGACCTAGTTAACGA.1
## AAGGAGCAGCAATCTC.1
## AAGGAGCTCTGCTGCT.1
```

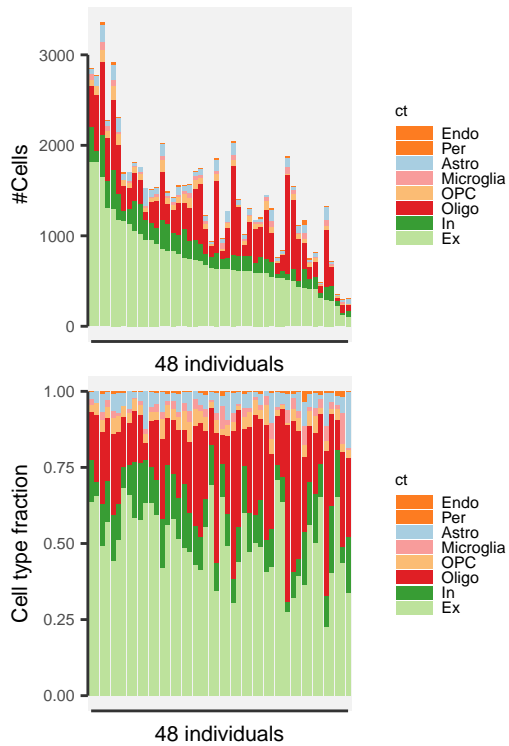
- Gene/feature list (these names should exist in the marker file):

```
## F0538757.2
## SAMD11
## NOC2L
## KLHL17
## PLEKHN1
## PERM1
## HES4
## ISG15
## AGRN
## RNF223
```



PDF

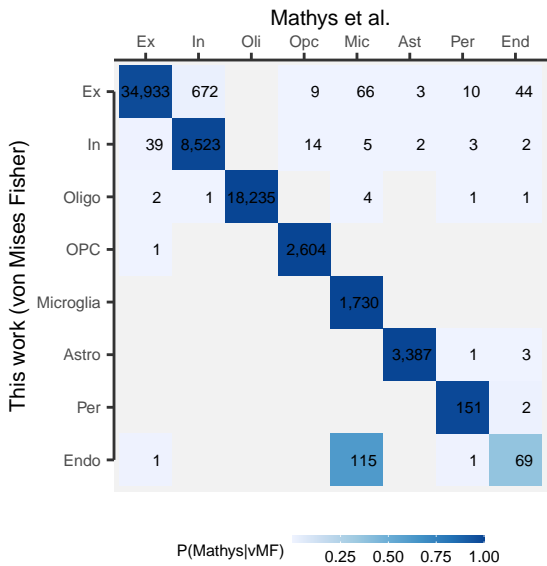
Basic statistics for cell type annotation



PDF

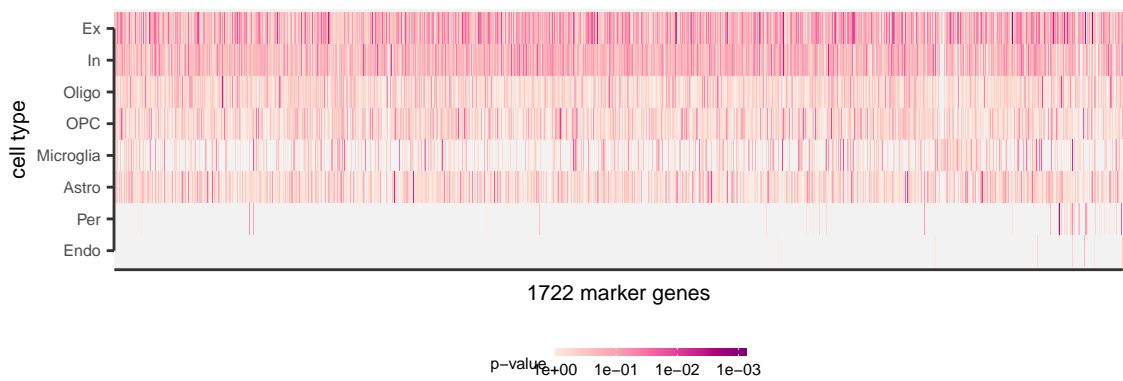
- 1,472 cells per individual: 50.9% ($N=745 \pm 388$) cells for Ex, 12.23% ($N=179 \pm 103$) cells for In, 3.54% ($N= 54 \pm 35$) cells for OPC, 0.33% ($N= 4 \pm 5$) cells for Endo, 25.22% ($N=380 \pm 252$) cells for Oligo, 2.61% ($N= 36 \pm 25$) cells for Microglia, 4.95% ($N= 71 \pm 47$) cells for Astro, 0.34% ($N= 4 \pm 4$) cells for Per on average.

Correlation with the annotations of Mathys *et al.*

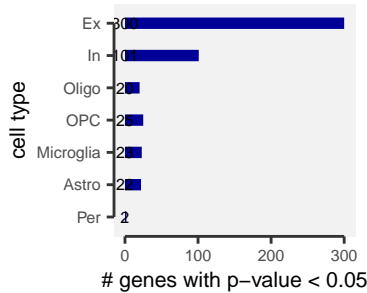


PDF

Marker genes are associated with the AD phenotype



PDF

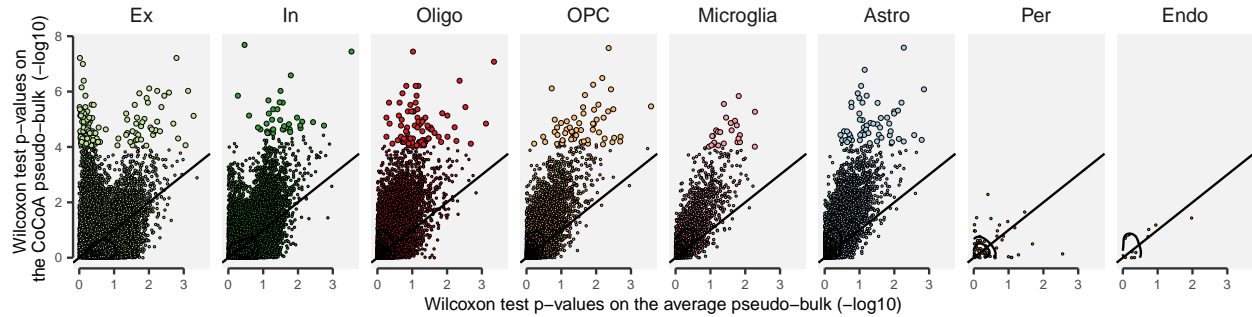


PDF

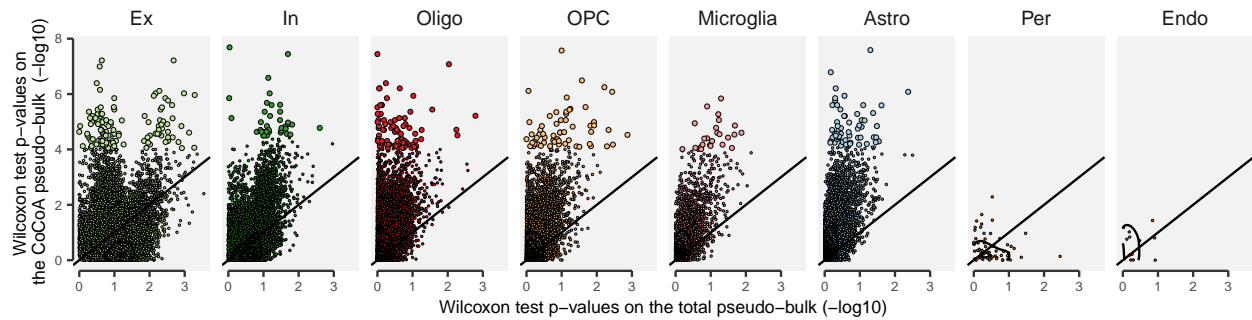
Differential expression analysis

CoCoA identified 377 genes are differentially-regulated between different AD conditions

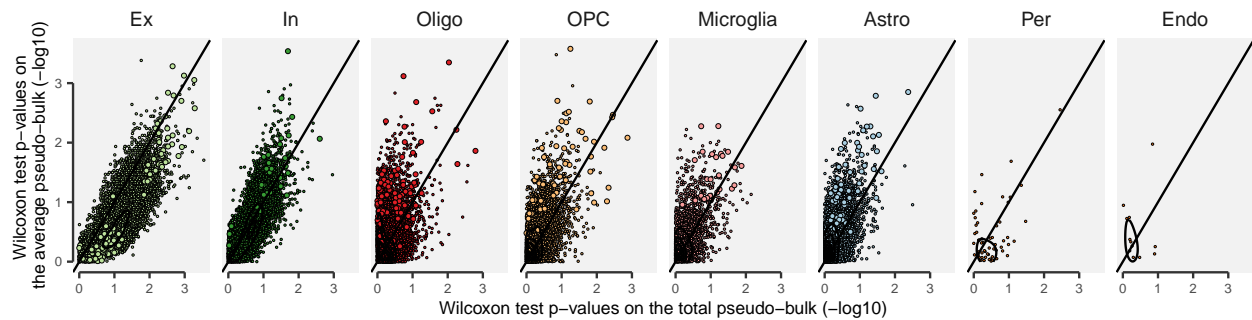
- 377 genes (2.72%) consist of 123 found in Ex, 36 found in In, 86 found in Oligo, 58 found in OPC, 24 found in Microglia, 57 found in Astro



PDF



PDF



PDF

```
## pdf
```

```
## 2
```

p-value distribution

Why do we find more genes with CoCoA?

Take the examples of Microglia genes

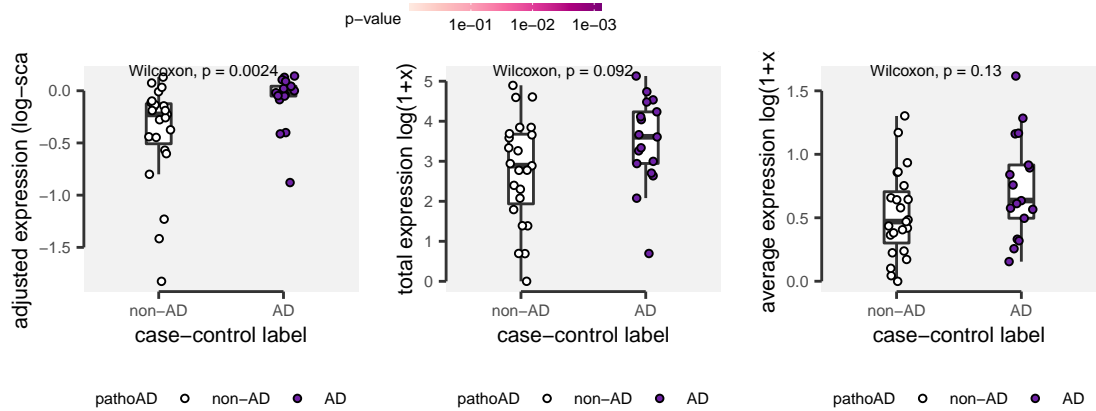
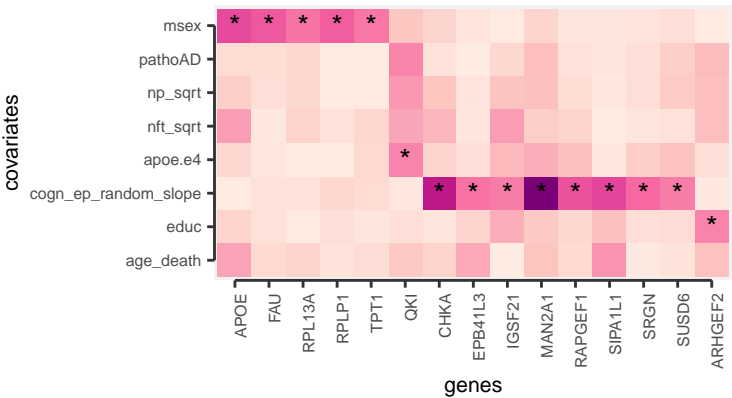
```
## Loading required package: cba
```

```
## Loading required package: proxy
```

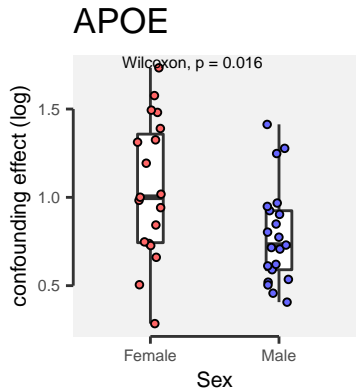
```
##
```

```
## Attaching package: 'proxy'
```

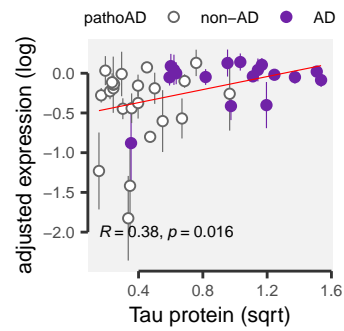
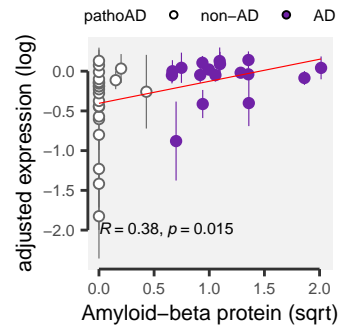
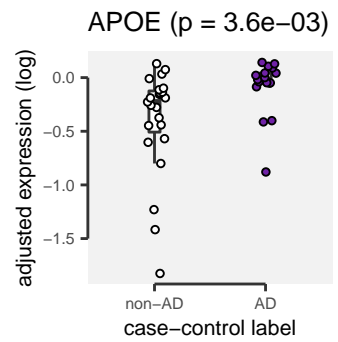
```
## The following objects are masked from 'package:stats':
##
##   as.dist, dist
## The following object is masked from 'package:base':
##
##   as.matrix
```



PDF

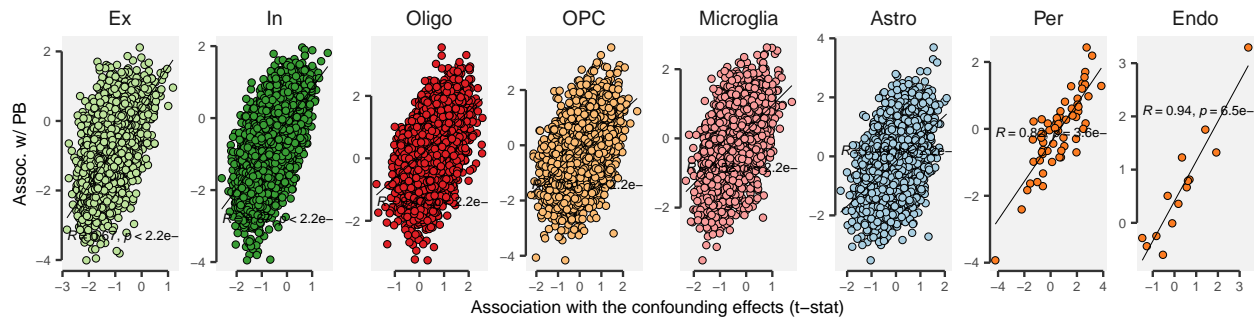


PDF



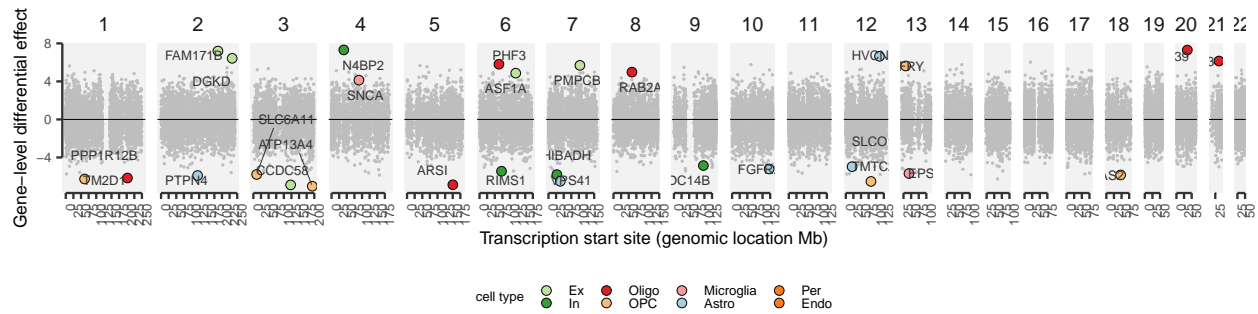
PDF

Confounding effects prevalent...



PDF

A genomic view combining all the cell types

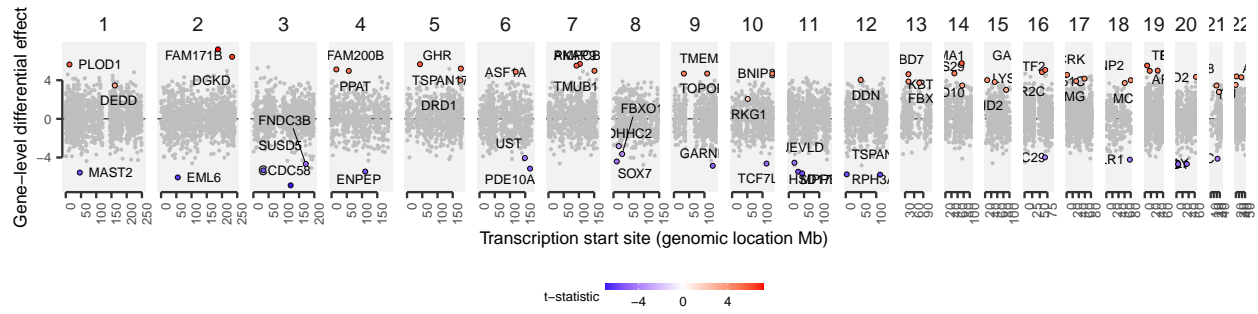


Combined genomic view

- PMPCB @ Ex
- ASF1A @ Ex
- CCDC58 @ Ex
- FAM171B @ Ex
- DGKD @ Ex
- CDC14B @ In
- N4BP2 @ In
- HIBADH @ In
- RIMS1 @ In
- TTC3 @ Oligo
- RBM39 @ Oligo
- RAB2A @ Oligo
- PHF3 @ Oligo
- ARSI @ Oligo
- PPP1R12B @ Oligo
- FRY @ OPC
- PIAS2 @ OPC
- SLC6A11 @ OPC
- TMTC2 @ OPC
- TM2D1 @ OPC
- ATP13A4 @ OPC
- EPSTI1 @ Microglia
- SNCA @ Microglia
- VPS41 @ Astro
- SLCO1C1 @ Astro
- FGFR2 @ Astro
- HVCN1 @ Astro

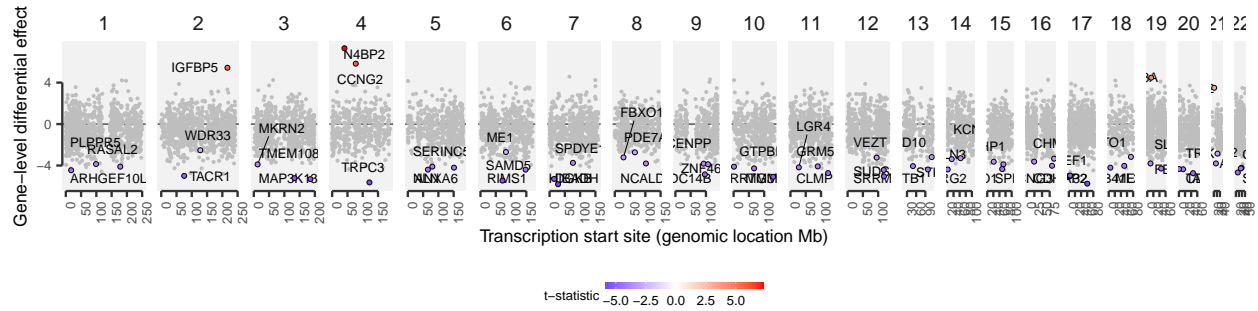
- PTPN4 @ Astro

Genomic views for each cell type



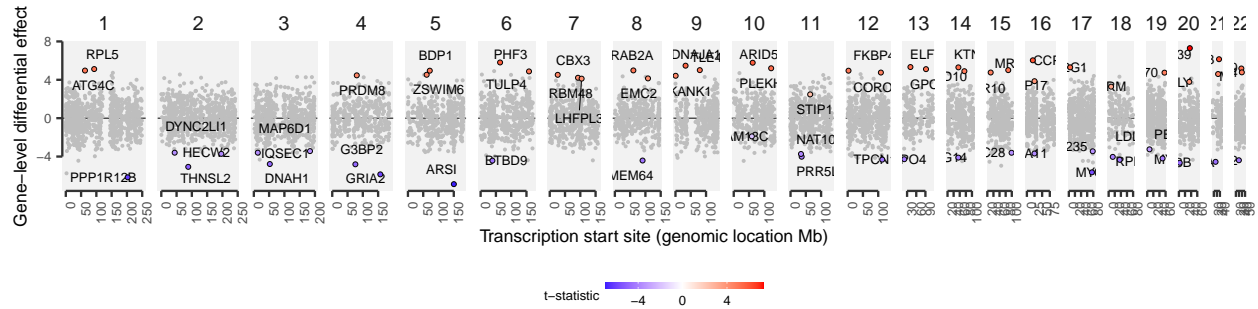
Ex

Ex top genes



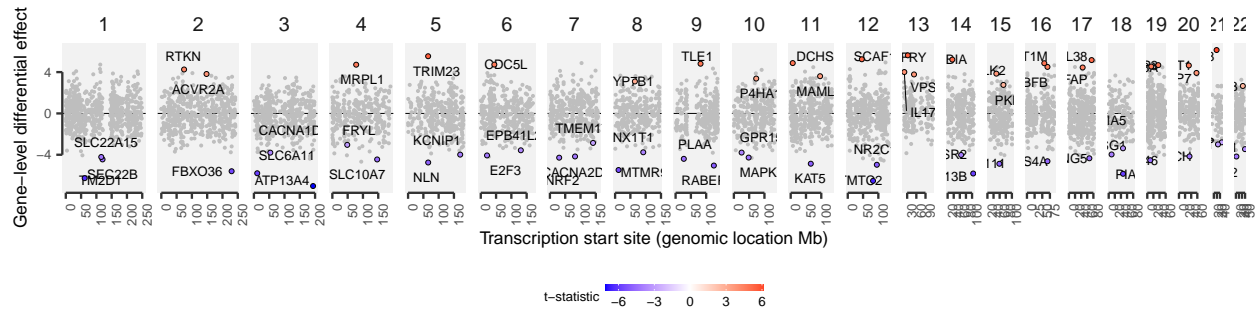
In

In top genes



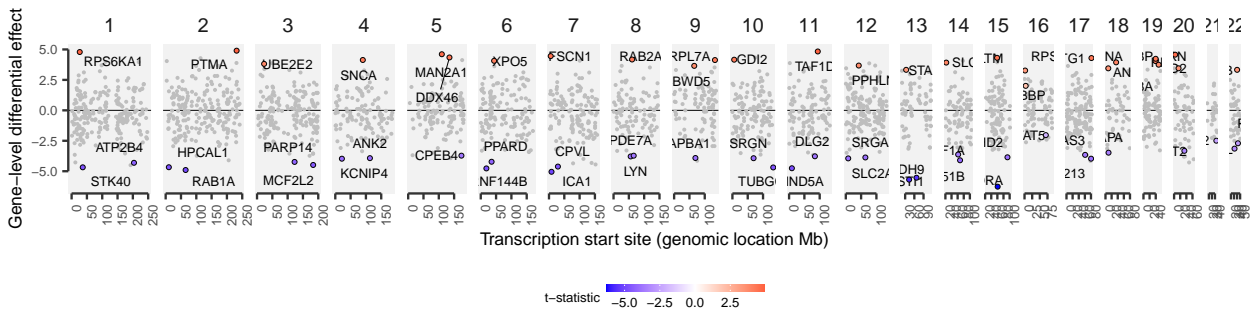
Oligo

Oligo top genes



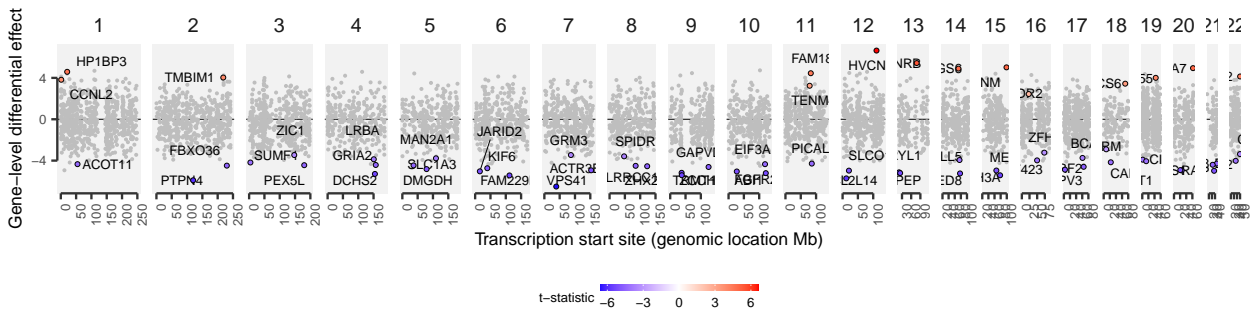
OPC

OPC top genes



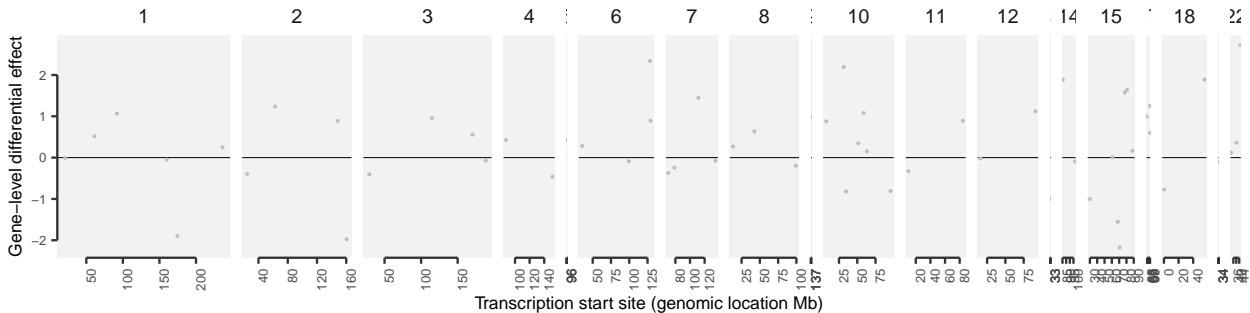
Microglia

Microglia top genes



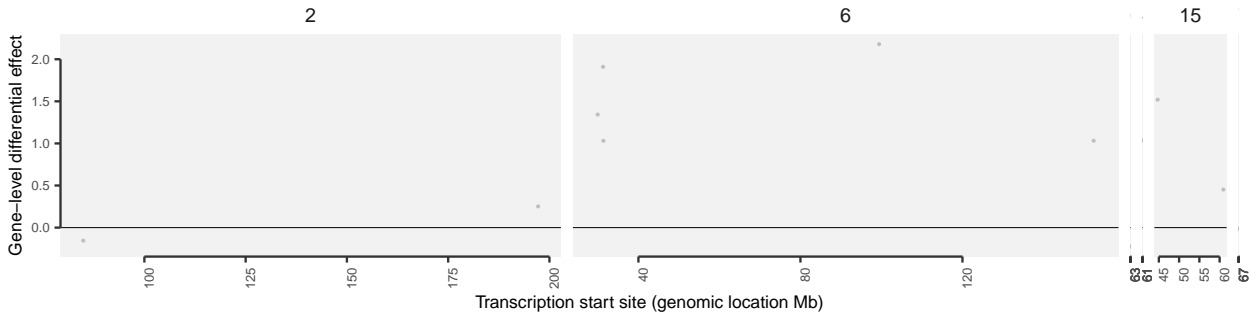
Astro

Astro top genes



Per

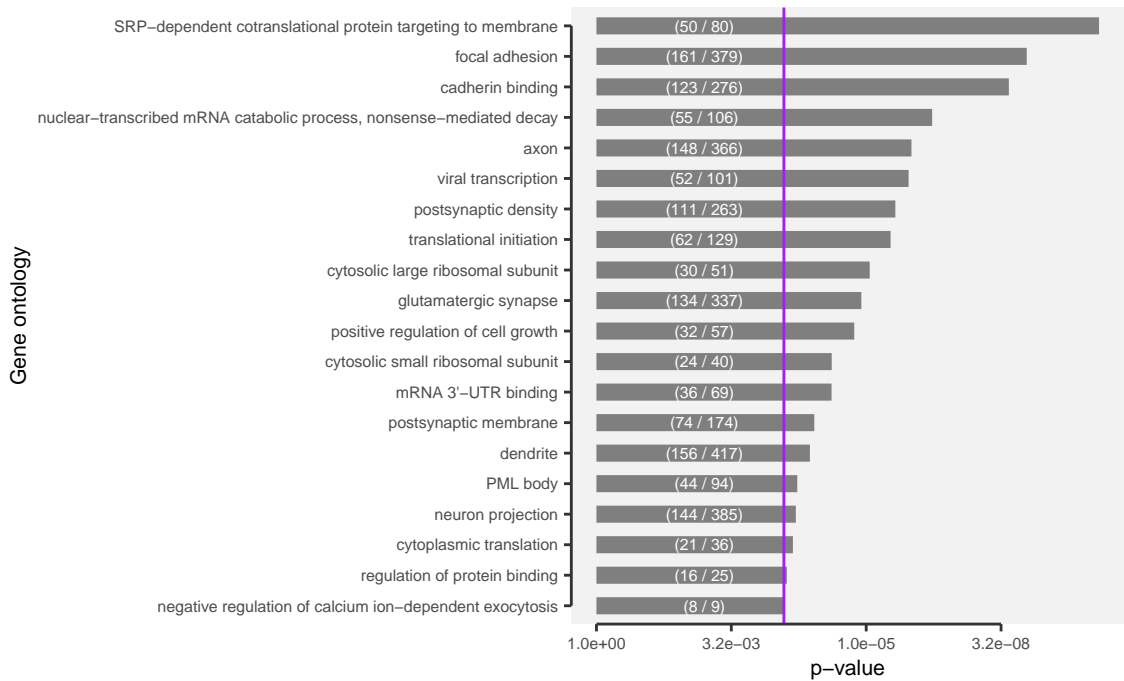
Per top genes



Endo

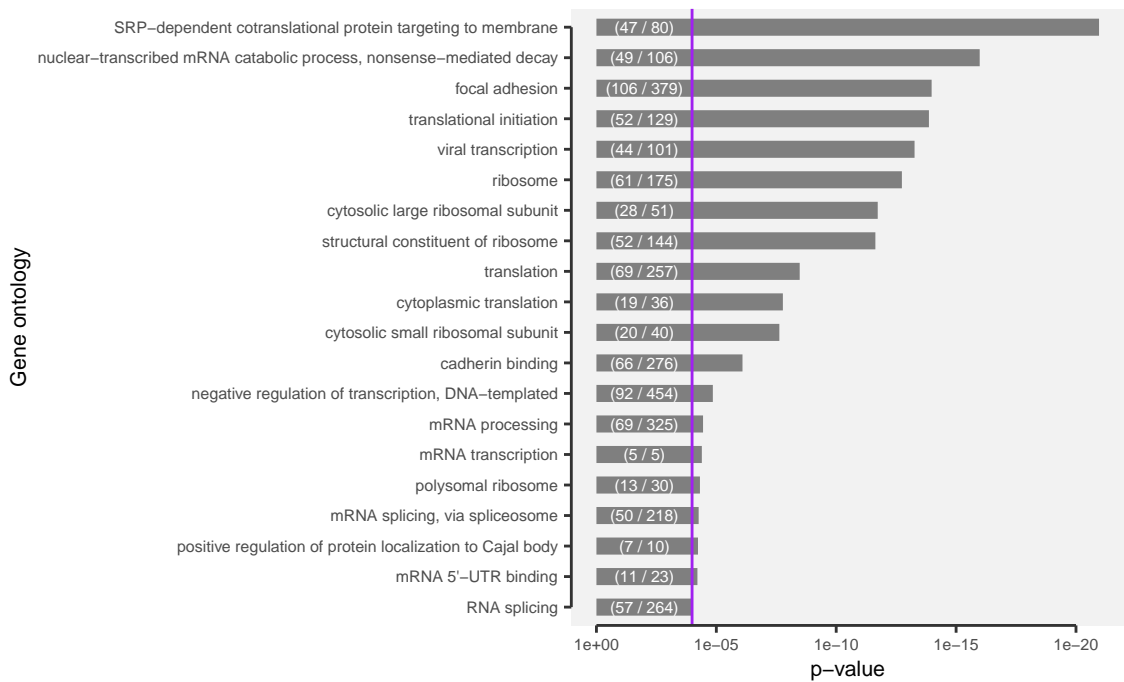
Endo top genes

Gene ontology enrichment



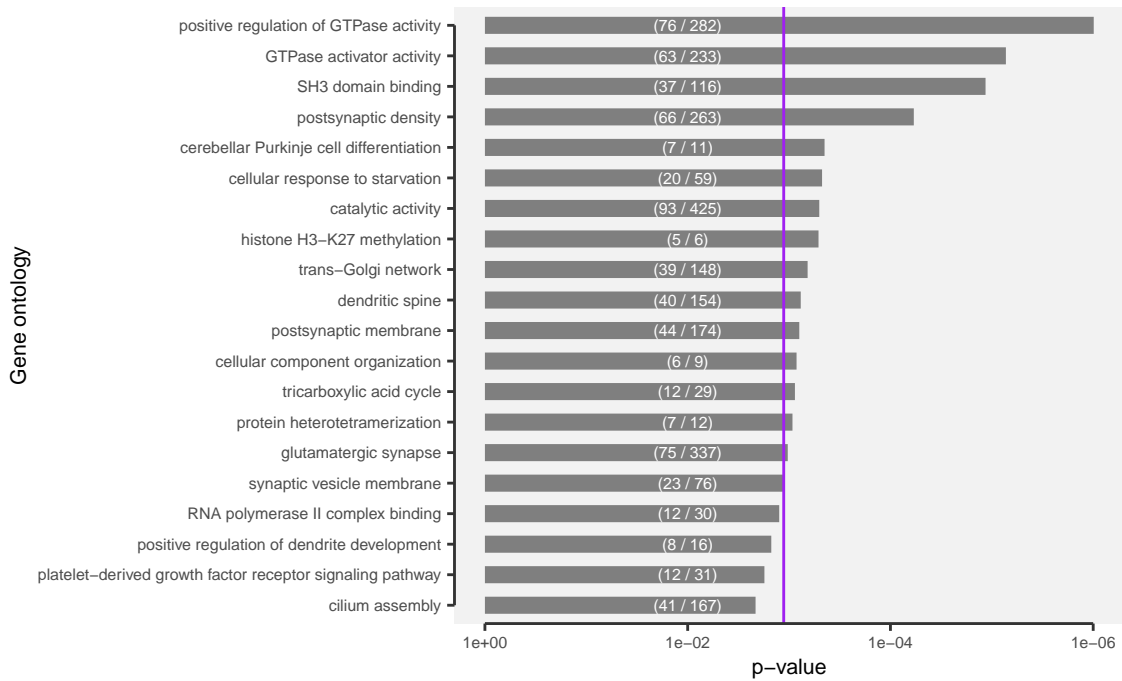
PDF

Up-regulated DEGs (2,092 genes)



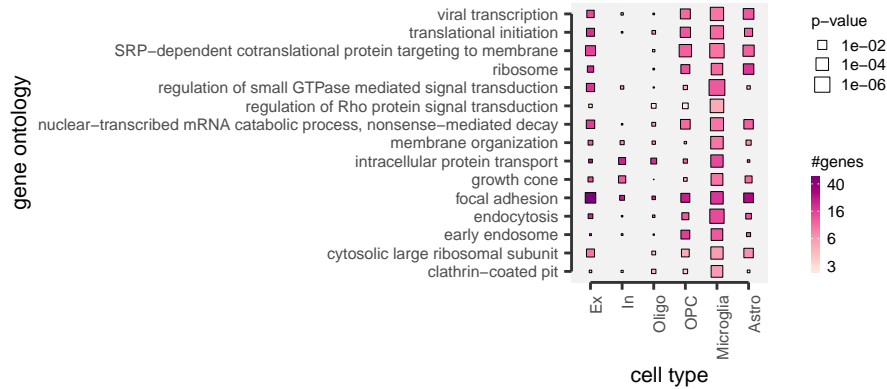
PDF

Down-regulated DEGs (2,317 genes)



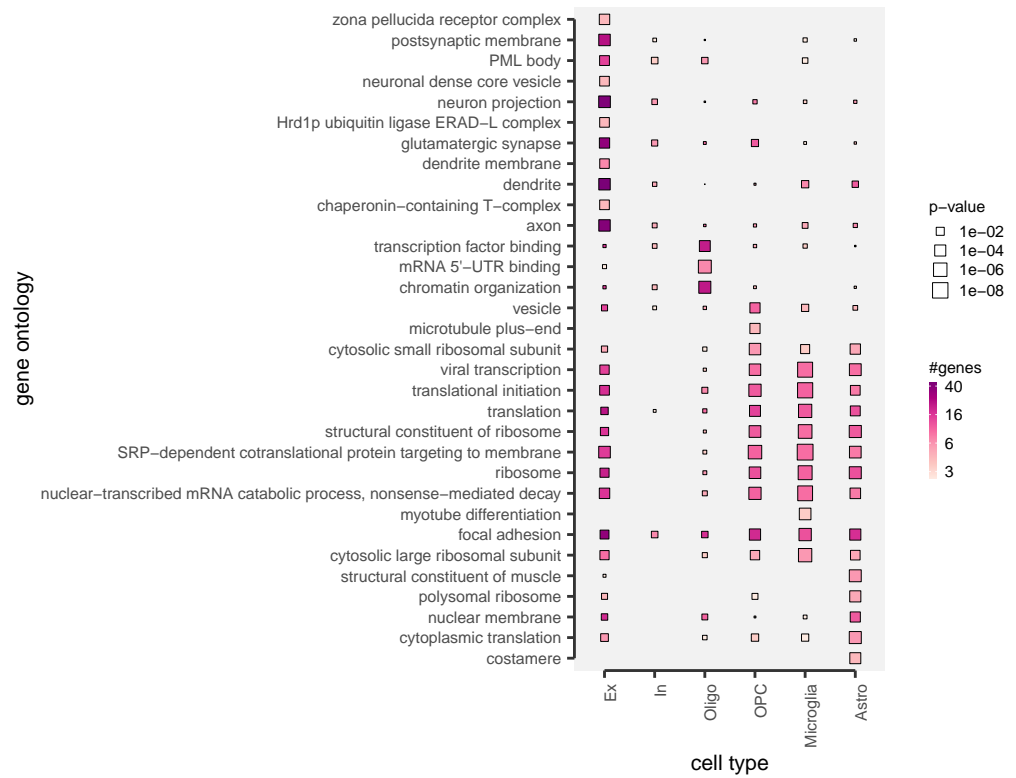
PDF

GO enrichment for each cell type



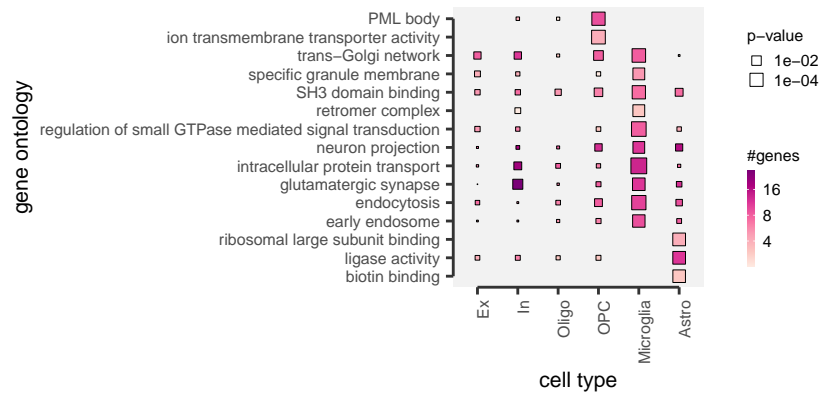
DEGs combined

PDF



DEGs up-regulated

PDF



DEGs down-regulated

PDF