Counterfactual Confounder Adjustment for Single-cell Differential Expression Analysis

Yongjin Park

Results

Cell type annotation

2,648 marker genes

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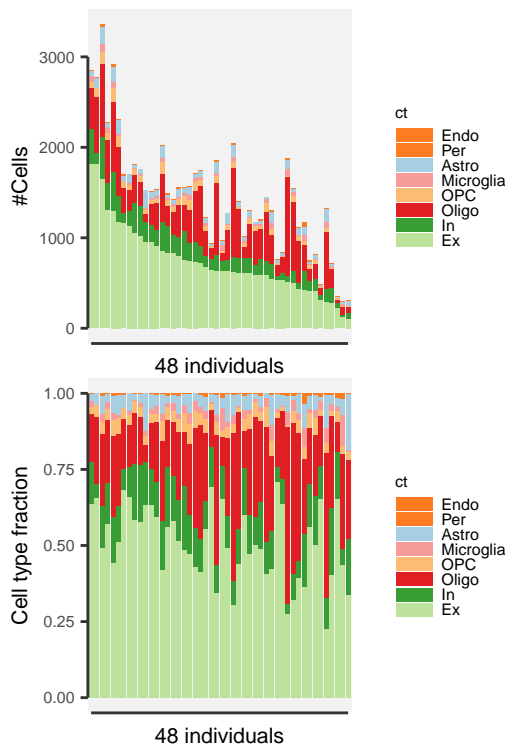
- Gene/feature list (**these names should exist in the marker file**):

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


1726 genes present in snRNA−seq
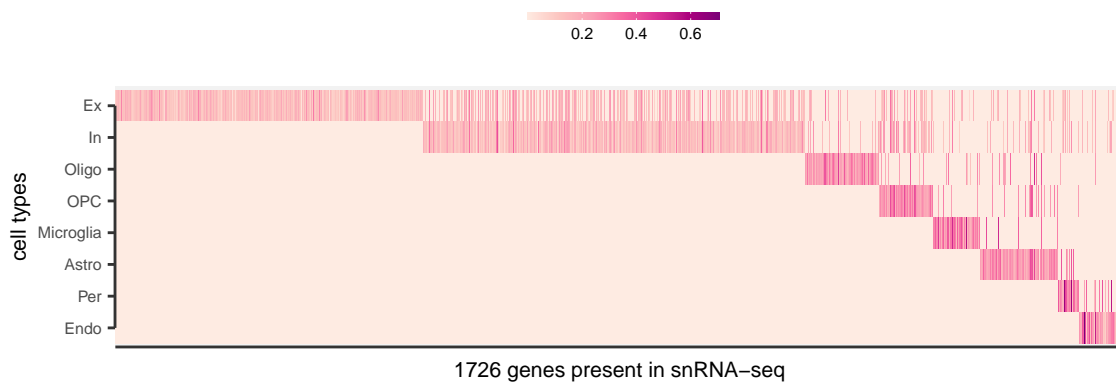
PDF

Basic statistics for cell type annotation

PDF

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

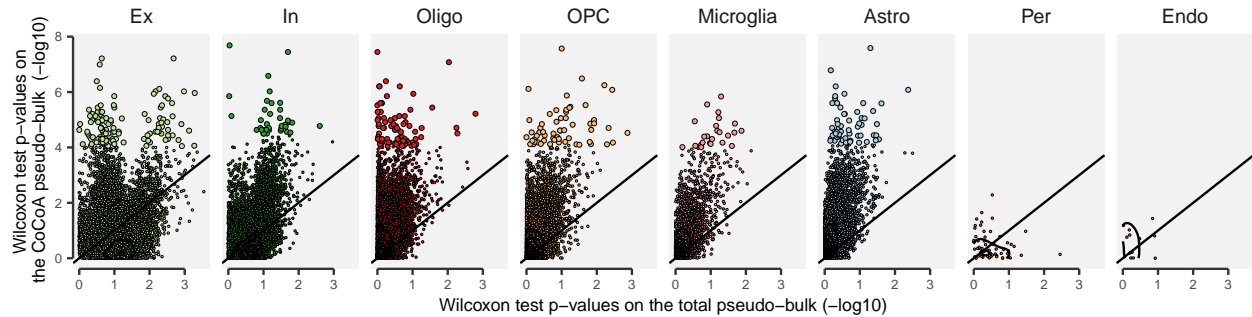## Differential expression analysis

CoCoA identified 377 genes are differentially-regulated between different AD conditions
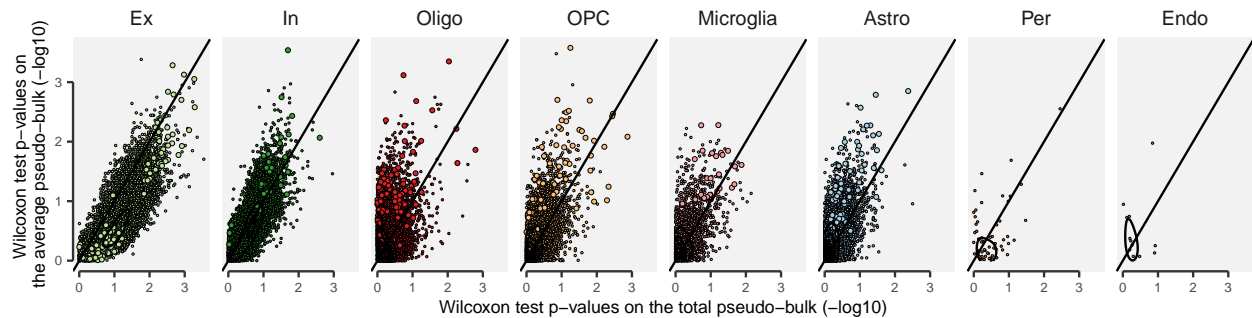
- 377 genes (2.72%) consist of 123 found in Ex, 36 found in In, 86 found in Oligo, 58 found in OPC, 24 found in Microglia, 57 found in Astro
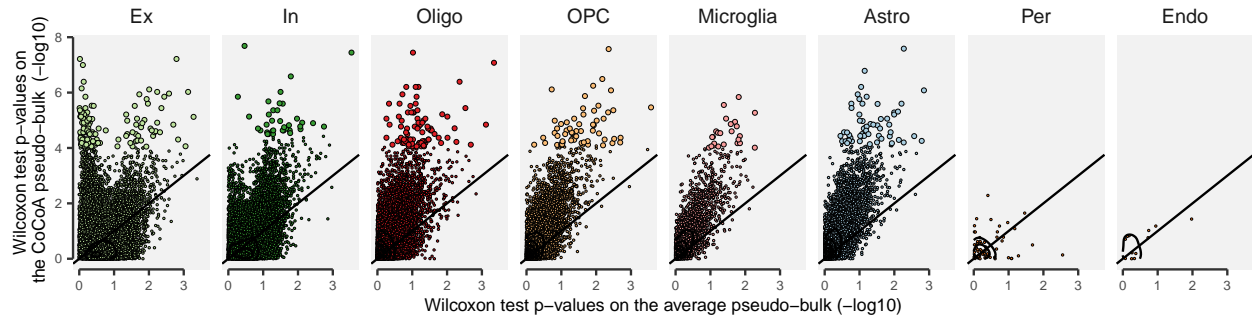
PDF

PDF

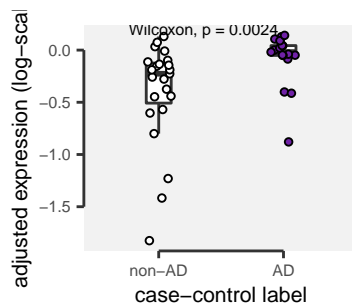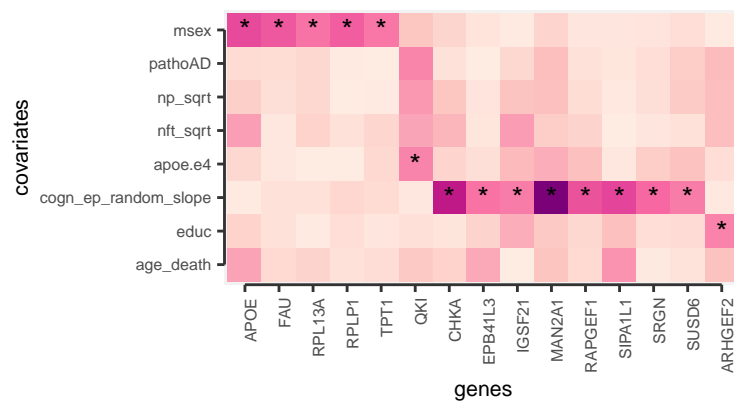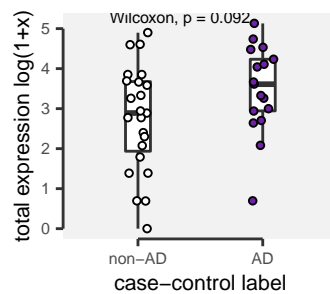
PDF

```
## pdf
##   2
```

p-value distribution

Why do we find more genes with CoCoA?
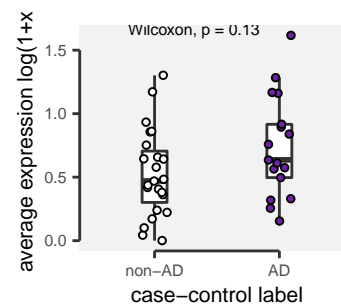
Take the examples of Microglia genes

3

```
## 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

APOE

PDF

APOE (p = 3.6e−03)
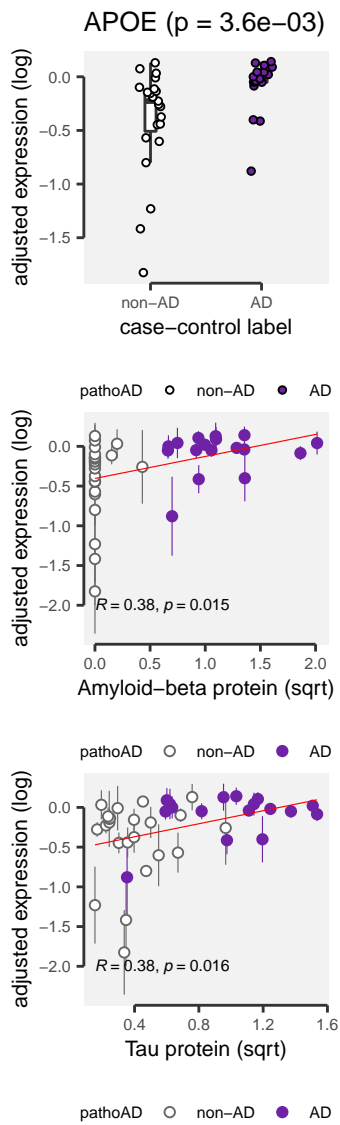
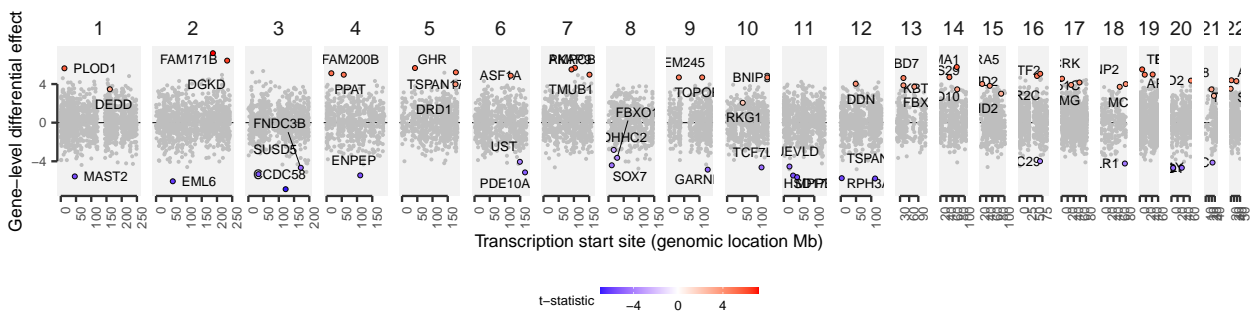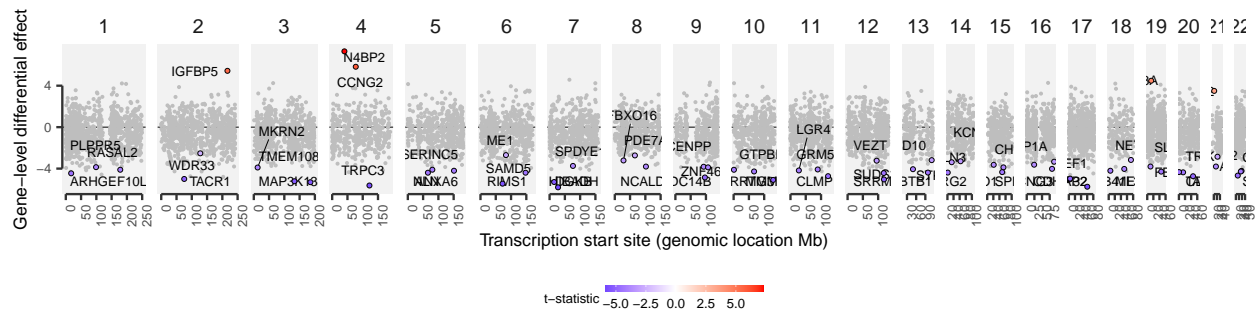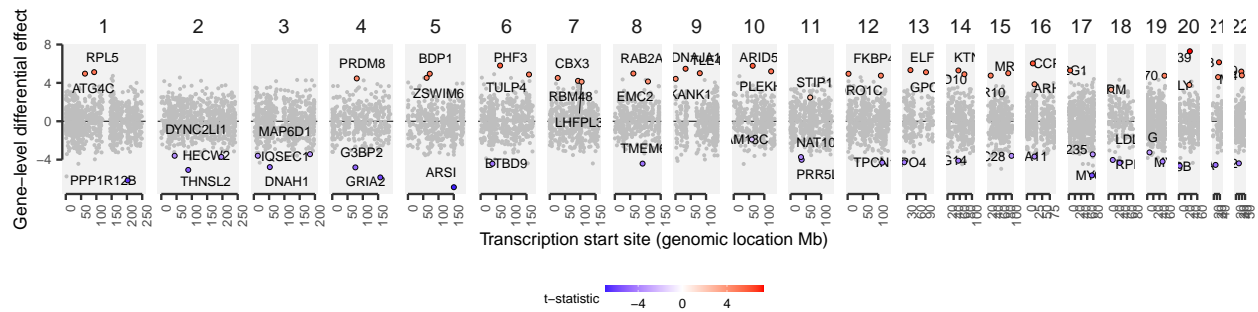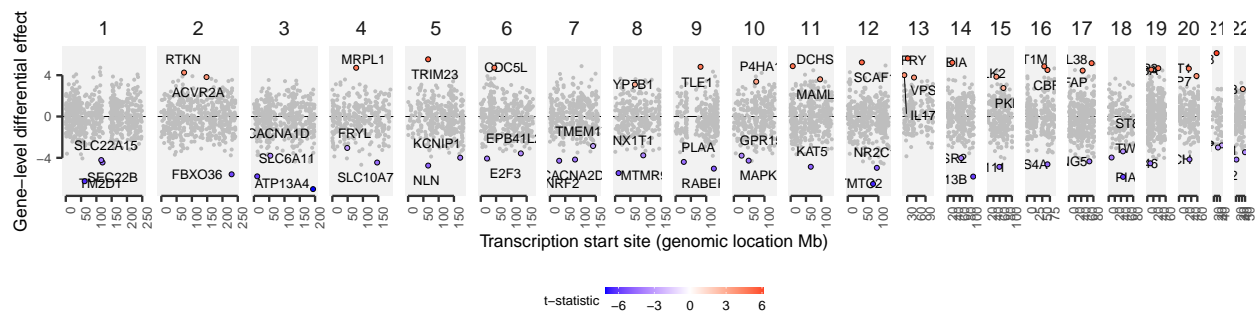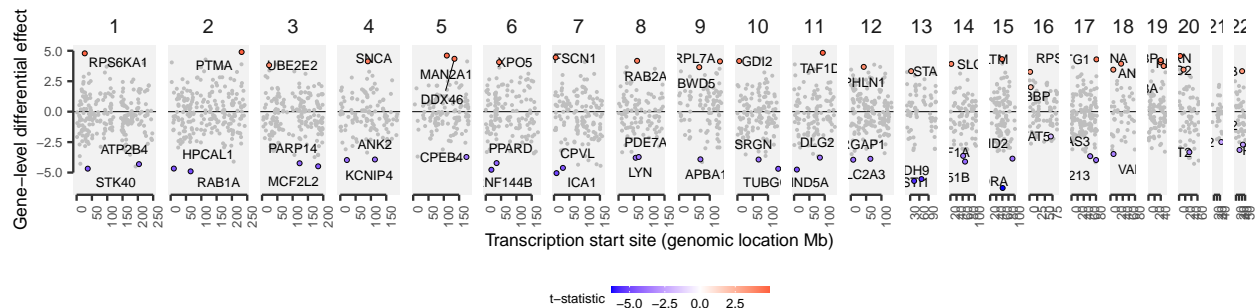PDF

Genomic views

Ex

Ex top genes

In

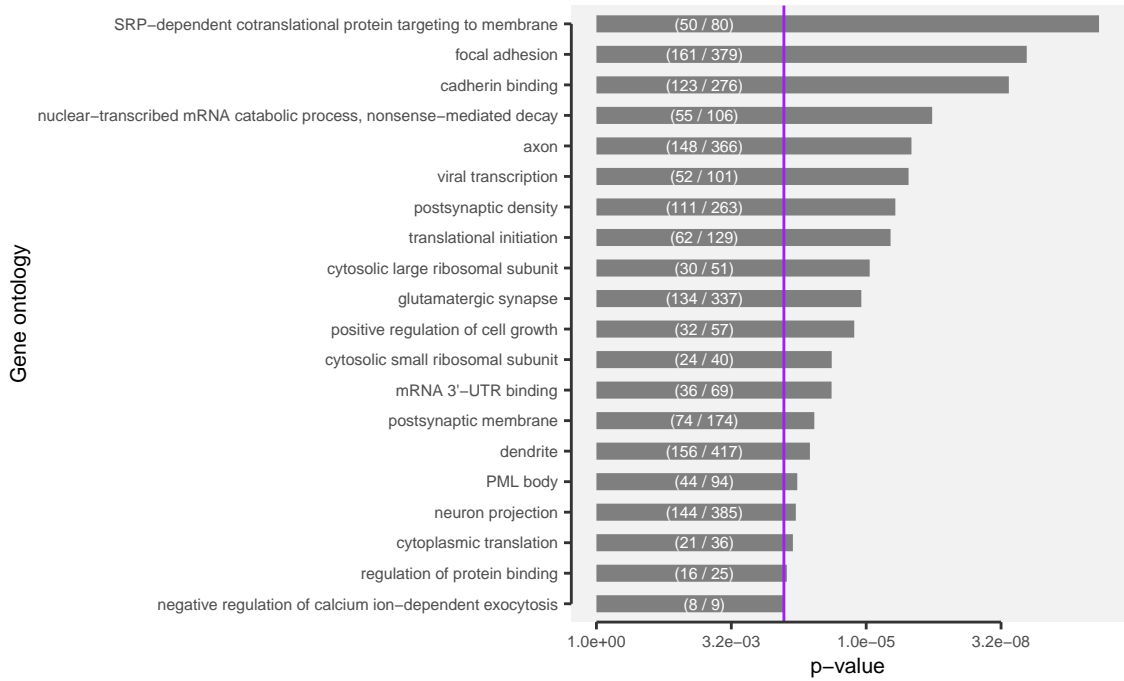In top genes

Oligo

Oligo top genes

OPC

OPC top genes

Microglia

Microglia top genes

Gene−level differential effect

Transcription start site (genomic location Mb)

t−statistic   −6   −3   0   3   6

Astro

Astro top genes

Gene−level differential effect

Transcription start site (genomic location Mb)

Per

Per top genes

Gene−level differential effect

Transcription start site (genomic location Mb)

Endo

Endo top genes

7

Gene ontology enrichment

PDF