

Djuna von Maydell

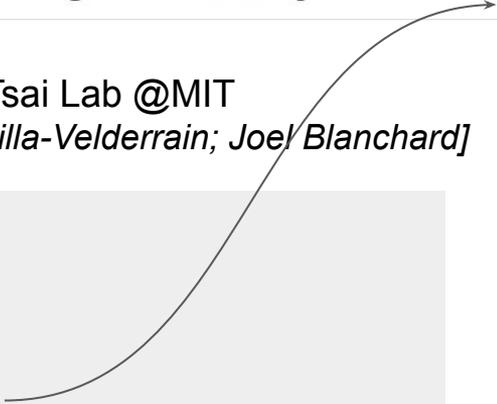
The first publicly available dataset of single-cell gene expression from *post-mortem* human brain tissue of patients who are carriers of APOE4, the major Alzheimer's disease risk gene

Open Data
@ MIT

APOE4 impairs myelination via cholesterol dysregulation in oligodendrocytes

Djuna von Maydell, Tsai Lab @MIT
[Leyla Akay; Jose Davilla-Velderrain; Joel Blanchard]

Big Dataset



APOE4 impairs myelination via cholesterol dysregulation in oligodendrocytes

Djuna von Maydell, Tsai Lab @MIT
[Leyla Akay; Jose Davilla-Velderrain; Joel Blanchard]

Big Dataset

?

New Discoveries

The Dataset



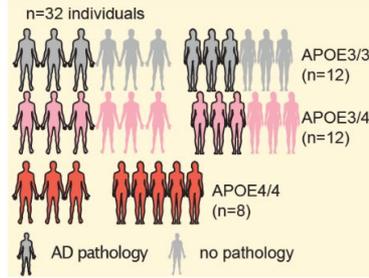
→ Predict effects of APOE4 on molecular states of human brain cells

The Dataset



→ Predict effects of on molecular states of human brain cells

(1) Humans Donated Brain Tissue post-mortem



Why make these data open?

circa 20,000 x 200,000
→ cell-states

per-subject
metadata

(2) Only followed up on a single prediction
→ *APOE is the strongest genetic risk variant for Alzheimer's*
→ *many other mechanisms of APOE4 risk*

(3) Ask completely different questions!

(4) Creating the dataset is really expensive!

In practice, what does this mean?

Open in theory \neq Accessible and usable in practice

Making sense of the data...

TWCC-AD10-AD10-lib1_S2_L002_...

TWCC-AD12-AD12-lib1_S1_L003_...

TWCC-C11-C11-lib1_S5_L003_R2...

TWCC-AD3-27768-lib1_S4_L001_...

TWCC-AD1-AD1-lib1_S2_L004_R...

TWCC-C12-C12-lib1_S3_L003_R2...

TWCC-AD1-AD1-lib1_S2_L004_R...

TWCC-AD8-AD8-lib1_S5_L001_R...

TWCC-AD10-AD10-lib1_S2_L002_...

TWCC-AD3-27768-lib1_S4_L001_...

TWCC-AD5-AD5-lib1_S3_L001_R...

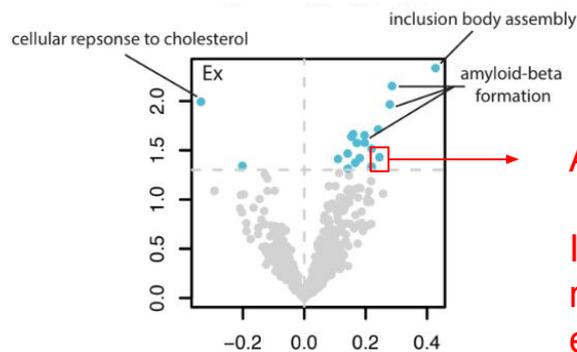
TWCC-AD11-AD11-lib1_S2_L003_...

TWCC-AD11-AD11-lib1_S2_L003_...

TWCC-AD9-AD9-lib1_S4_L001_R...

TWCC-AD12-AD12-lib1_S1_L003_...

Using the data...



And this one?

I guess I'll have to re-write and re-run the entire pipeline...

How we ended up making the data open:

README.md

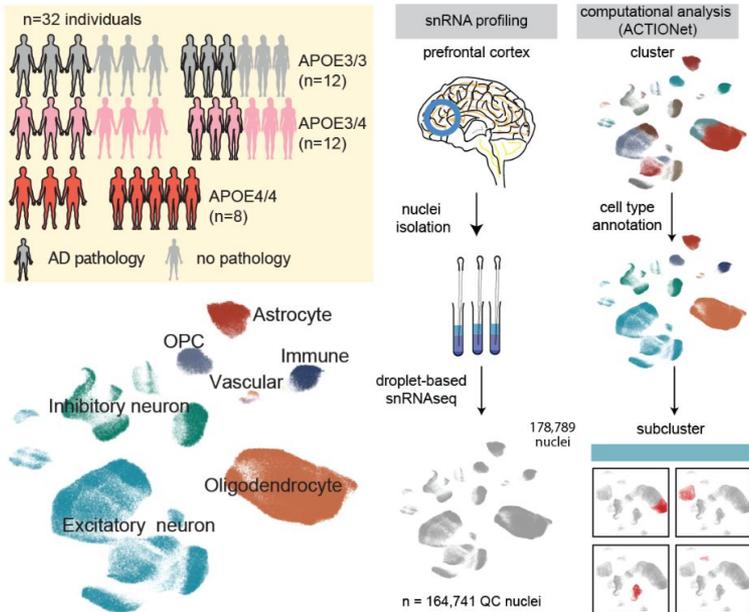
This repository contains instructions for open data access and code to reproduce the analyses presented in

APOE4 impairs myelination via cholesterol dysregulation in oligodendrocytes

Find the corresponding paper [here](#). paper is accepted for publication, link coming soon

snRNAseq & lipidomic Data availability

snRNAseq database of APOE genotype effects in human brain



a) Detailed descriptions of data files

- Download the single-cell- and lipidomic-related data from [Synapse](#) and add these data to the `./data` directory according to the directories given in the table below. This includes the following files: N.B. The file `ROSMAP_clinical.csv` needs to be downloaded separately [from here](#).

Data File	Description / Origin
<code>single_cell_data/individual_level_averages_per_celltype/Ast.csv</code>	run <code>./scripts/get_individual_level_averages</code>
<code>single_cell_data/individual_level_averages_per_celltype/Ex.csv</code>	run <code>./scripts/get_individual_level_averages</code>
<code>single_cell_data/individual_level_averages_per_celltype/In.csv</code>	run <code>./scripts/get_individual_level_averages</code>
<code>single_cell_data/individual_level_averages_per_celltype/Mic.csv</code>	run <code>./scripts/get_individual_level_averages</code>

b) Code for data pre-processing & analyses

- Now run the following code snippets to recapitulate the analysis:

Create Environment:

```
conda env create -f ./environment/apoe_env.yml
```

Check it out and let us know how we can improve usability!

github.com/djunamay



Thank you!

Li-Huei Tsai (MIT)

Joel Blanchard (Mt. Sinai)

Leyla Akay (MIT)

Jose Davila-Velderrain (Human Technopole)

Religious Orders Study
and Rush Memory Aging Project
(Human Samples & Metadata)

David A. Bennett (Rush)

Ihab Hajjar (Emory)

ROSMAP donors & their families

*A very large lipidomic dataset
from human brains that we used
in our analysis; open access on
the AD knowledge portal*

The Tsai Lab