

Avoiding dissociation bias with single-nucleus RNA-seq

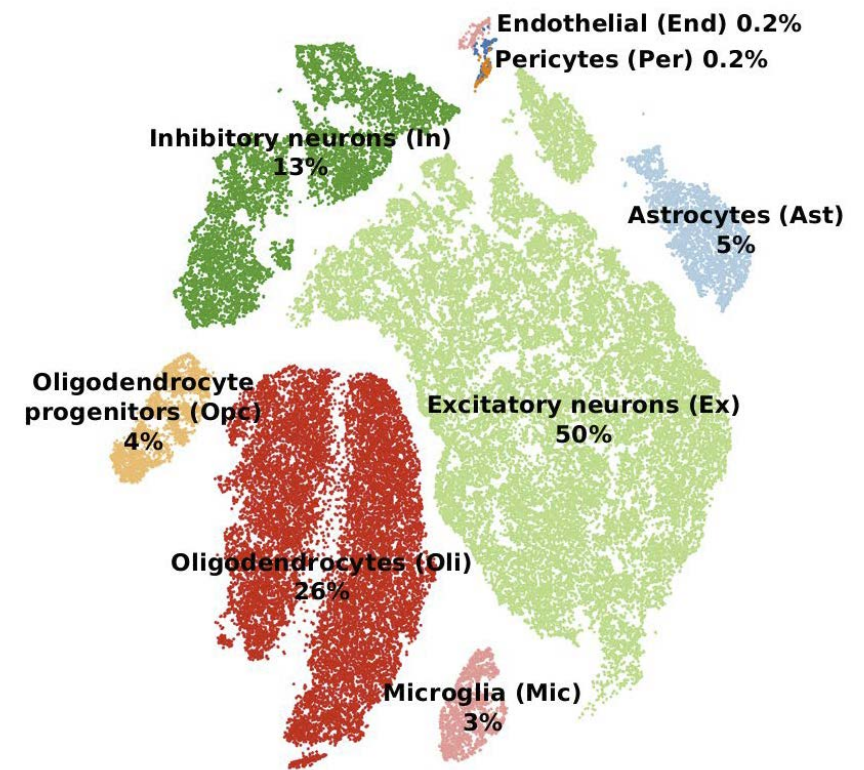
Technical Journal Club

Lukas Frick

9 December 2019

Single-cell RNA-seq (scRNA-seq) and single-nucleus RNA-seq (snRNA-seq) are competing technologies with the same goal

- scRNA-seq/snRNA-seq profile **global mRNA expression** in **individual cells**.
- They share a disadvantage: Only a **small sample** of mRNAs from each cell can be captured.
 - In each cell, many genes will be missed (count of 0), even though they are expressed.
 - This can be compensated by **clustering cells** to find subtypes, and then analyzing all cells in that group.



tSNE plot

Mathys *et al*

Other technologies (bulk RNA-seq, ribosome profiling) produce a **complete transcriptome**, but they cannot examine single cells!

- Single cell-transcriptomics has unique advantages:
 - It can define new and rare cell subtypes, or new differentiation states (example: disease-associated microglia).
 - It correctly identifies a gene that is upregulated in one cell type, but downregulated in another!
 - It can distinguish between a gene that is truly **overexpressed**, and a gene that seems upregulated because a particular **cell subtype** is **more abundant**.

Single-cell RNA-seq requires **enzymatic digestion** of fresh tissue, which can introduce technical artefacts



[dx.doi.org/10.17504/protocols.io.m9sc96e](https://doi.org/10.17504/protocols.io.m9sc96e)

Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations

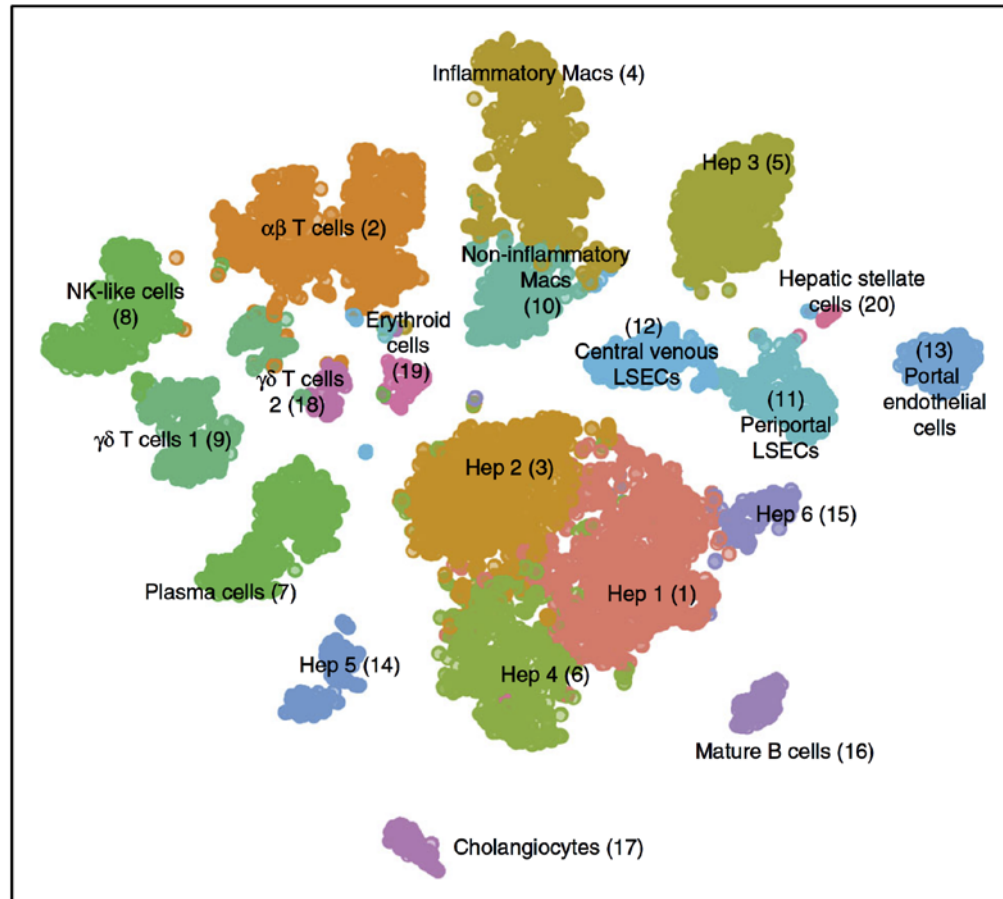
MacParland et al, *Nat Comm*, 2018

Digestion of human liver for scRNA-seq:

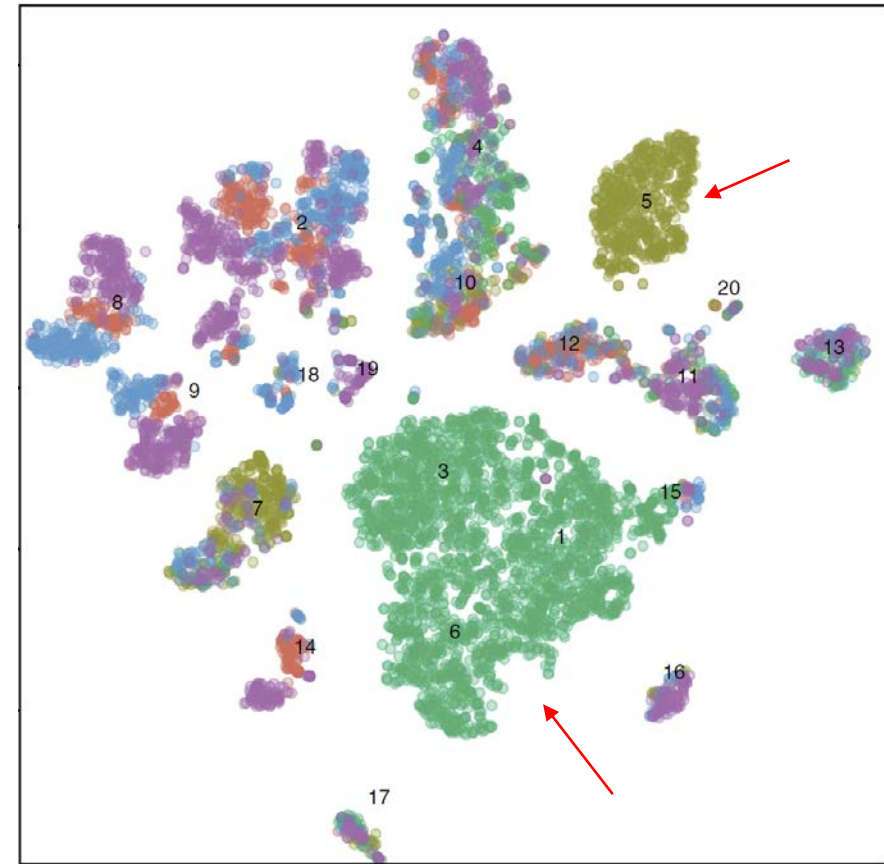
- The caudate lobe was resected before transplantation.
- It was cannulated via 2-3 exposed vessels in the cut surface, and perfused with collagenase and neutral protease.

In scRNA-seq, cells from multiple individuals are pooled ...

If a cluster comes from a single individual, this points to technical issues!



MacParland et al, *Nat Comm*, 2018



Liver 1 Liver 2 Liver 3 Liver 4 Liver 5

Protocols that involve perfusion with collagenase cannot be applied to small biopsies or frozen tissue

More recent paper:

A human liver cell atlas reveals heterogeneity and epithelial progenitors

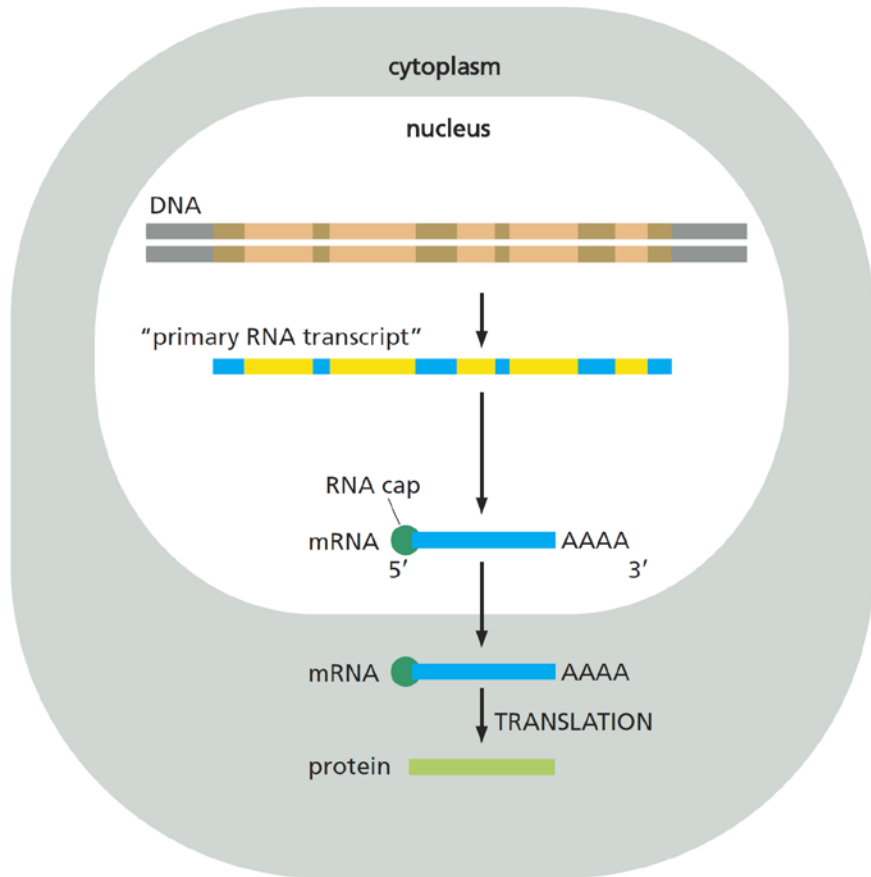
Aizarani et al, *Nature*, 2019

12 samples from liver resections were again **perfused** with collagenase

Next challenges for scRNA-seq:

- **Routine clinical samples**
- **Pathology**

Single-nucleus RNA-seq can be done on **frozen tissues** and **without** the need for **enzymatic digestion**

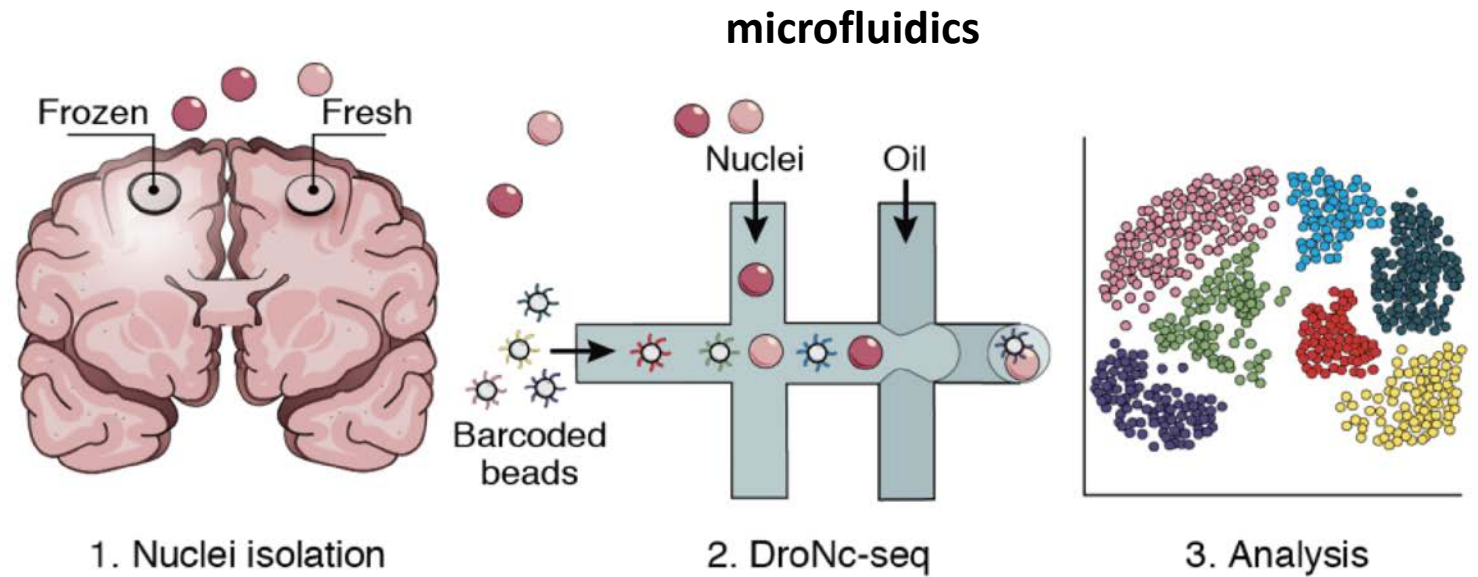


- The tissue is homogenized to **lyse cells** while leaving **nuclei intact**.
- About 10% of cellular RNA is located in the nucleus.
 - Both pre-mRNAs and mature mRNAs are counted.

The invention of DroNc-seq (snRNA-seq with droplet technology) enabled snRNA-seq on a larger scale

Massively parallel single-nucleus RNA-seq with DroNc-seq

Habib et al, *Nature Methods*, 2017



- mRNAs from single nuclei are captured on **barcoded** beads.
 - The barcodes uniquely identify **each sample** and **each cell**.
- Samples are pooled for reverse transcription (RT) and sequencing.

ARTICLE

Single-cell transcriptomic analysis of Alzheimer's disease

Hansruedi Mathys^{1,2,10}, Jose Davila-Velderrain^{3,4,10}, Zhuyu Peng^{1,2}, Fan Gao^{1,2}, Shahin Mohammadi^{3,4}, Jennie Z. Young^{1,2}, Madhvi Menon^{4,5,6}, Liang He^{3,4}, Fatema Abdurrob^{1,2}, Xueqiao Jiang^{1,2}, Anthony J. Martorell^{1,2}, Richard M. Ransohoff⁷, Brian P. Hafler^{4,5,6,8}, David A. Bennett⁹, Manolis Kellis^{3,4,11*} & Li-Huei Tsai^{1,2,4,11*}

Nature

June 2019

→ First single-cell view of AD pathology

Samples were selected from the Religious Order Study (ROS)

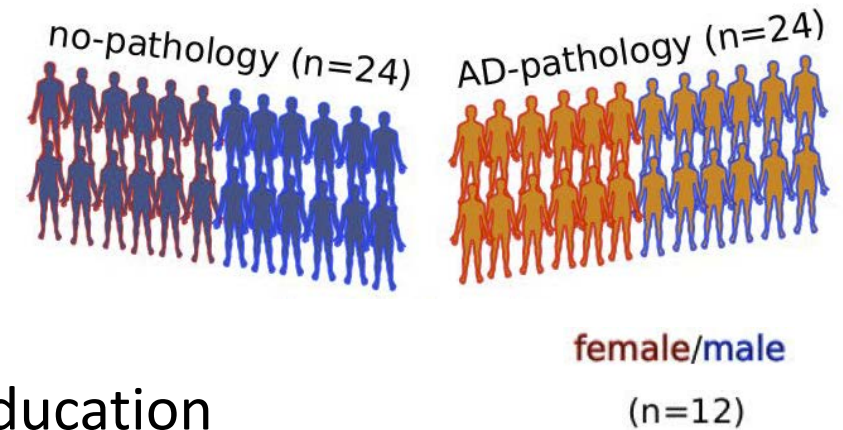
... a longitudinal cohort study of elderly nuns, priests and brothers.

- 48 subjects:

- 24 without pathology
- 24 with Alzheimer's disease (mild to severe β -amyloid)
- Matched for sex, age (~87 years), and years of education

- Extensive prior data was available:

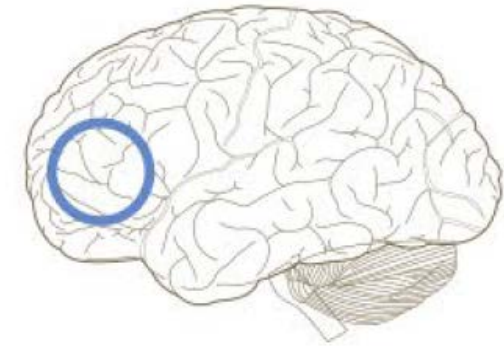
- Clinical data, cognitive tests, detailed post-mortem
- Bulk RNA-seq, genetics, epigenomics, proteomics, metabolomics...



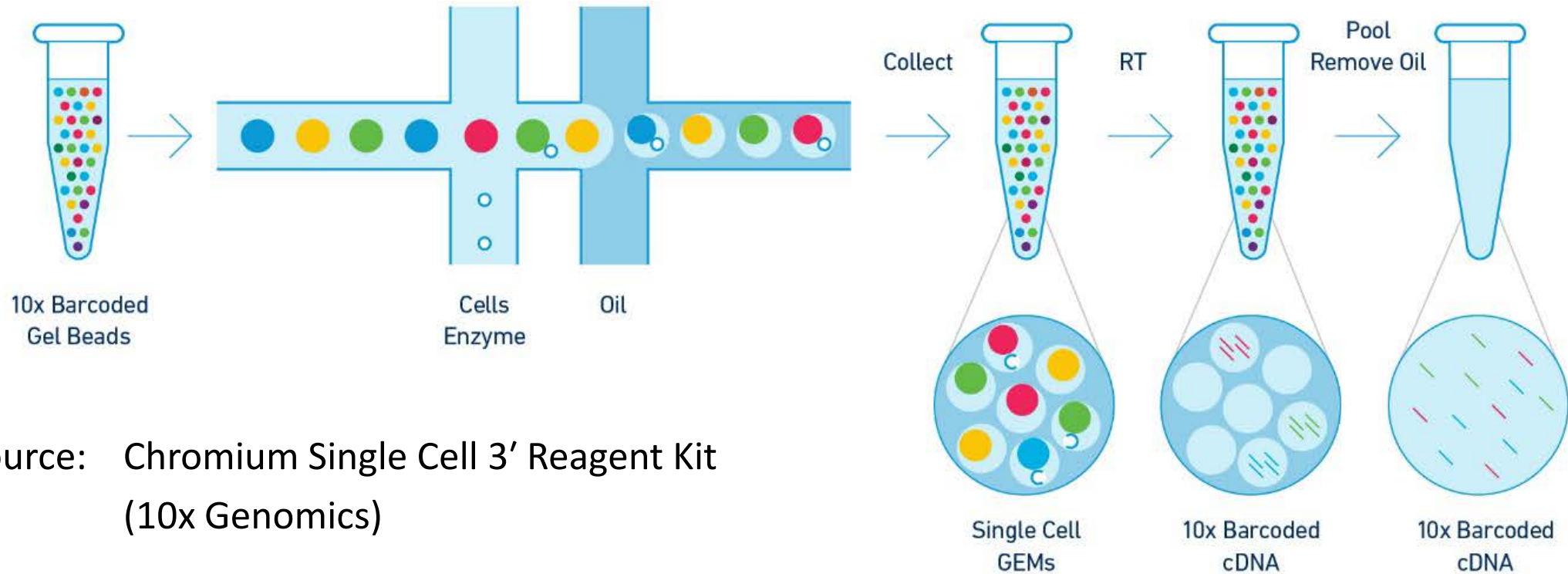
Methods: Tissue processing

- Frozen post-mortem samples from the **prefrontal cortex** were used.
- Protocol:
 - All procedures are carried out on ice.
 - Homogenize using a Dounce tissue grinder (10 strokes with the loose pestle).
 - Filter through a 40 μm cell strainer.
 - Load on top of OptiPrep density gradient and separate nuclei by centrifugation.
 - Count nuclei and dilute to 1000 nuclei/ μL .

Prefrontal cortex



Methods: Droplet-based single-nucleus RNA-seq (**DroNc-Seq**)



Source: Chromium Single Cell 3' Reagent Kit
(10x Genomics)

Cells are delivered at a limiting dilution, such that the most (~90-99%) of droplets contain no cell, while the rest largely contain a single cell.

Methods: **Data pre-processing**

- Read depth was equalized between the 48 libraries before merging.
- Low-quality cells were excluded:
 - Cells with <200 detected genes
 - Cells with a high ratio of mitochondrial to endogenous RNAs
→ indicates dead or stressed cells

→ Final dataset:

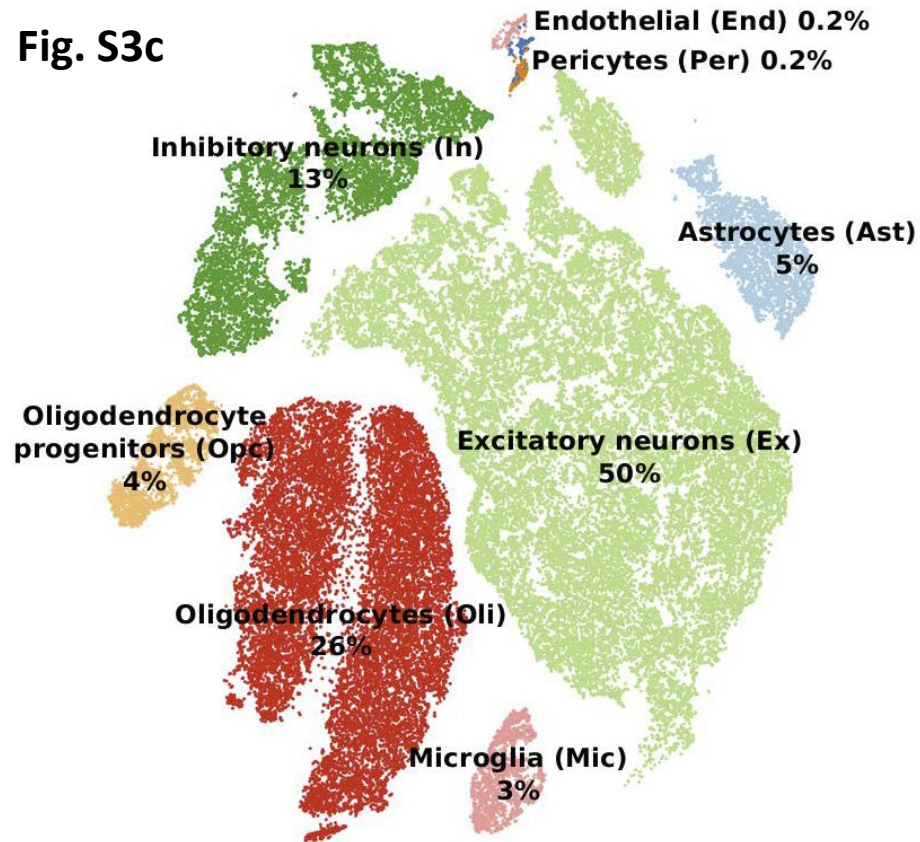
- 75'060 nuclei
- 17'926 protein-coding genes



80,660 cells

Methods: Cluster identification

Fig. S3c



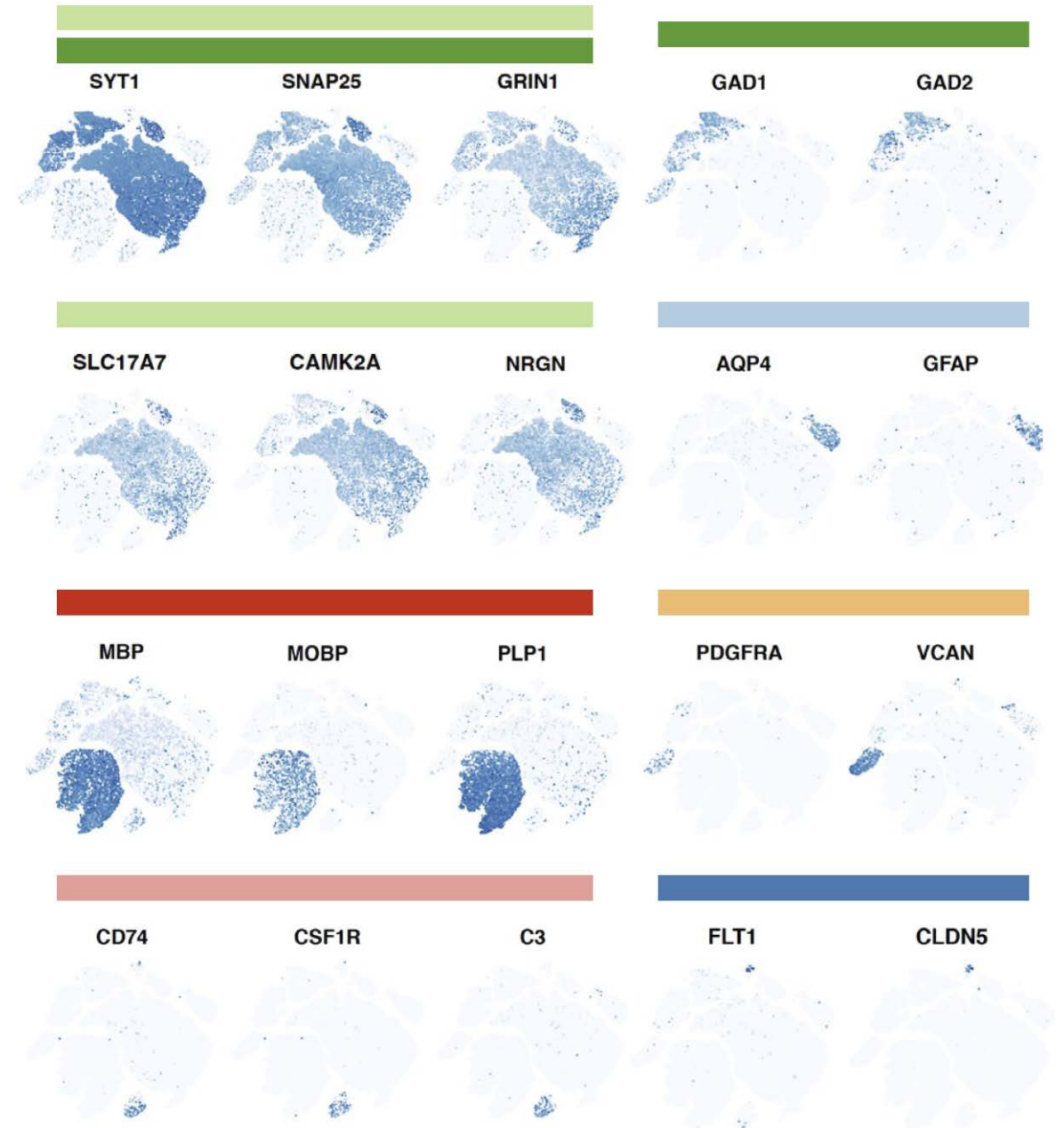
1. PCA (principle component analysis) on 3188 highly variable genes
2. t-SNE (t-distributed stochastic neighbor embedding) on the top 10 principle components
3. *Louvain* algorithm → identify clusters and sub-clusters
4. Use known marker genes to assign clusters to cell types

Spurious clusters (doublets, low-quality cells) were excluded → 70'634 cells remained

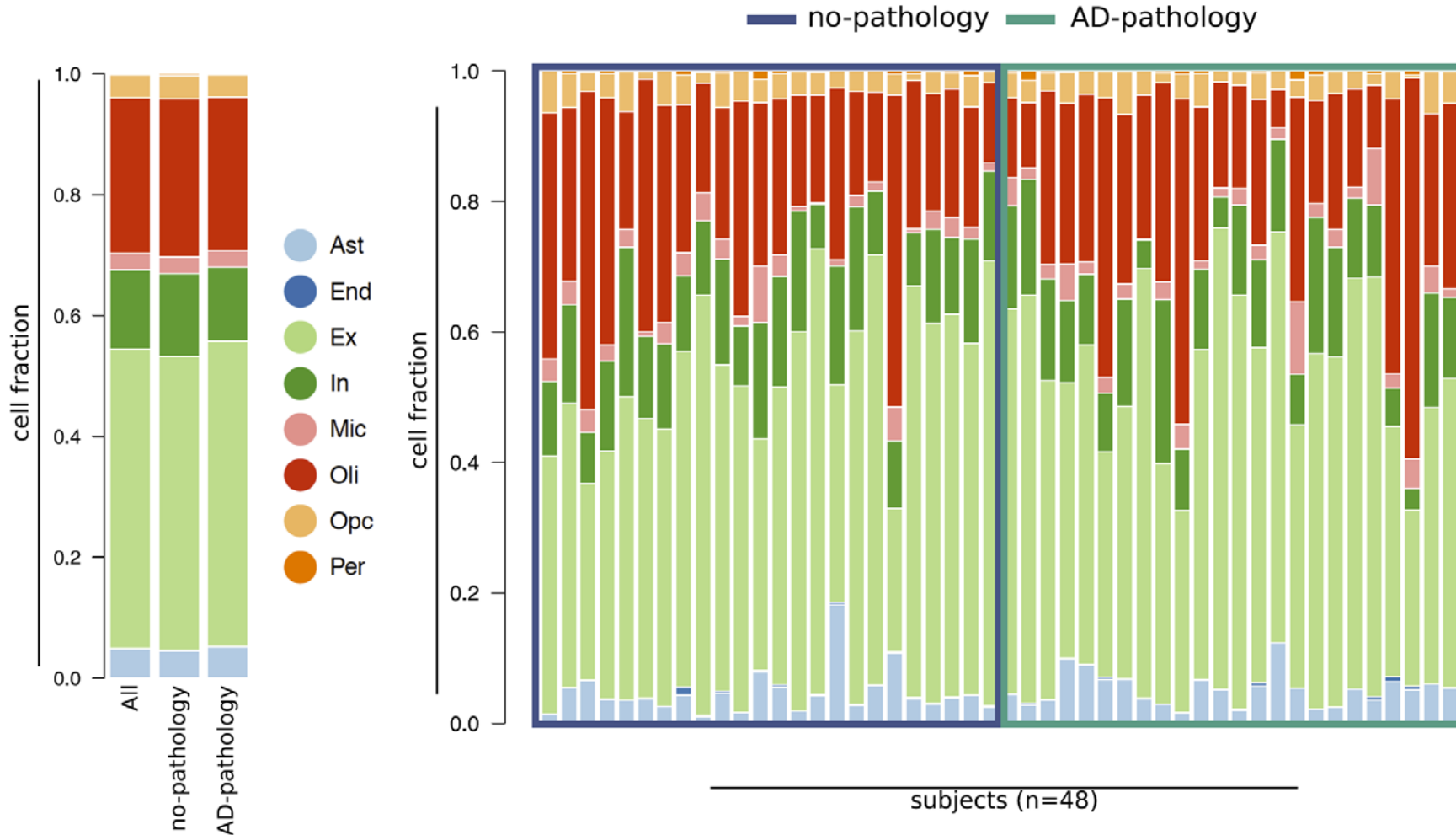
Methods: **Quality check**

Known cell-type specific marker genes were found in the expected clusters.

(Owing to their low cell counts, pericytes and endothelial cells populations were not analyzed for differential expression.)



Methods: Quality check



There was little variation in the proportions of cell subtypes between groups (AD vs. no AD) and different individuals.

Most highly upregulated genes

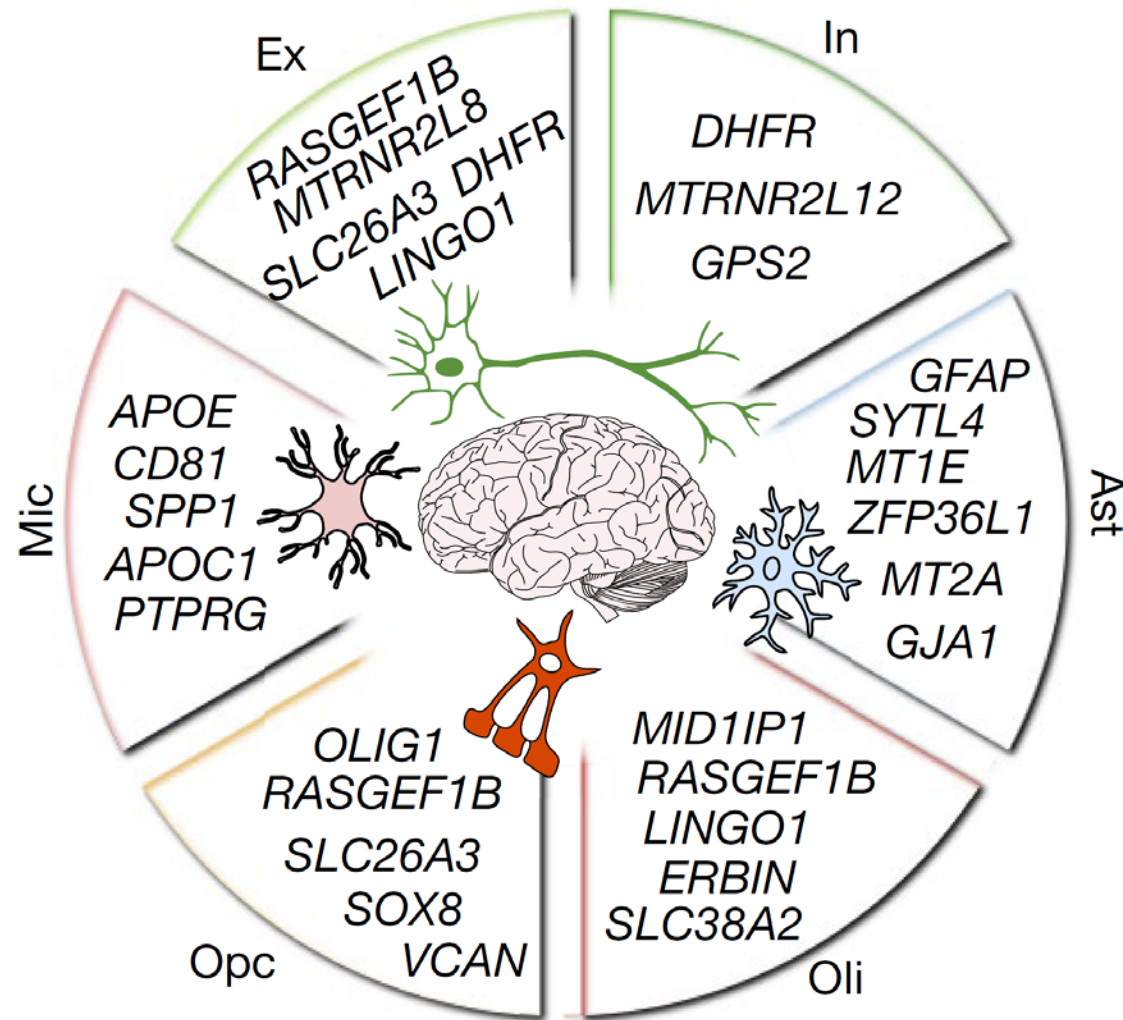


Fig. 1a

Alzheimer's-associated DEGs
(differentially expressed genes)
were **found in all cell types!**

- Excitatory neurons
- Inhibitory neurons
- Astrocytes
- Oligodendrocytes
- Oligodendrocyte precursor cells
- Microglia

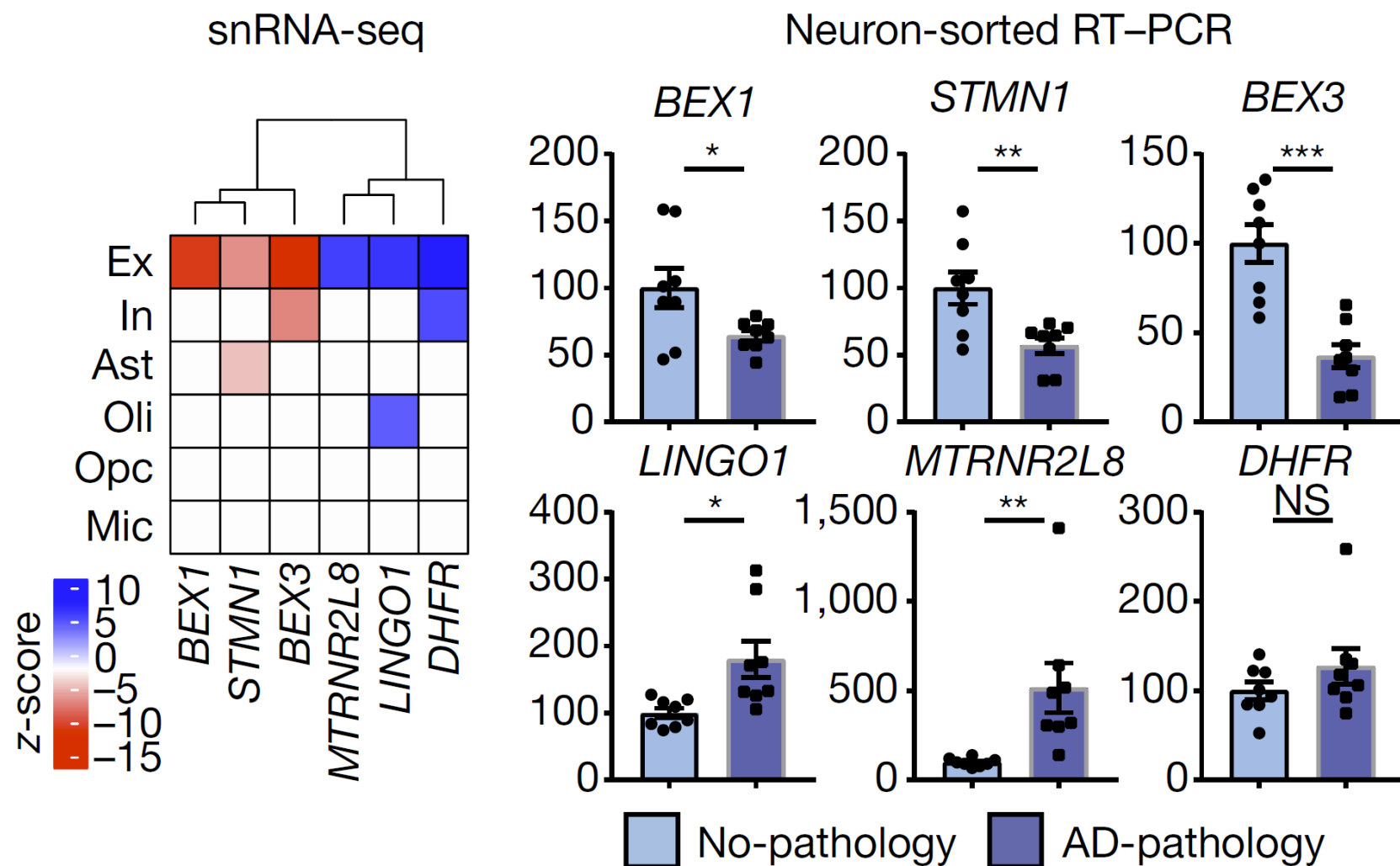
95% of DEGs were perturbed only
in neurons or a single glia type.

	DEGs	
Ex	565	191
In	51	3
Ast	32	37
Oli	71	102
Opc	18	10
Mic	13	22
	Down	Up

Fig. 1b

Neurons showed transcriptional **repression** (75% of DEGs in excitatory and 95% in inhibitory neurons were **downregulated**).

In contrast, most DEGs in oligodendrocytes, astrocytes and microglia were **upregulated** (53-63%).



Six neuronal DEGs were chosen for validation by qPCR.

NeuN+ nuclei were isolated by FACS.

5 of 6 DEGs could be confirmed.

Fig. 1c

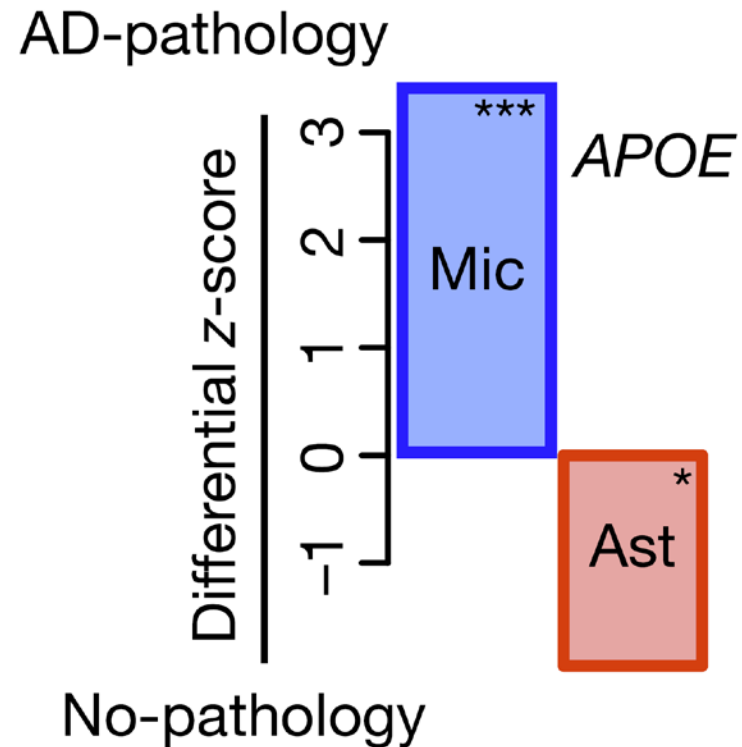


Fig. 1f

The ***APOE*** gene was upregulated in microglia, but downregulated in astrocytes!

Microglia:

n = 955 AD cells

n = 965 no-pathology cells

Astrocytes:

n = 1830 AD cells

n = 1562 no pathology cells

snRNA-seq versus bulk RNA-seq

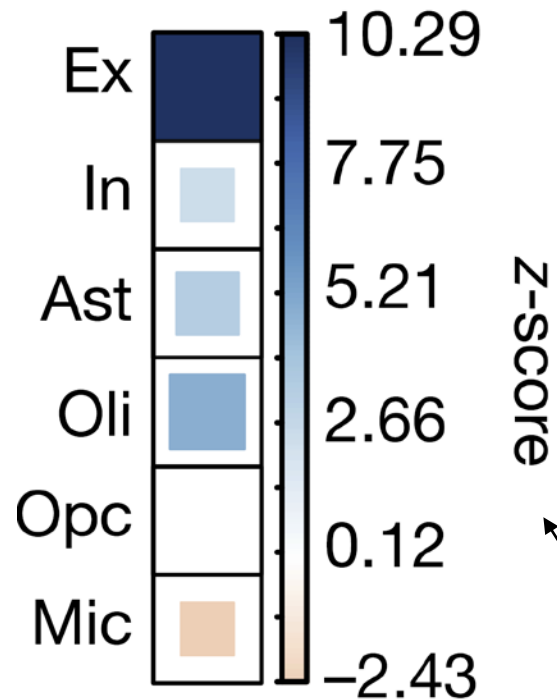


Fig. 1f

snRNA-seq data was compared with **bulk RNA-seq data**.

Bulk data are **dominated** by **excitatory neurons** and **oligodendrocytes**.

In contrast, expression changes in **microglia** were **poorly represented**.

A rank permutation test was done to correlate the DEG rankings in each cell type with the DEG rankings of bulk data.

Clustering based on clinic-pathological traits was done to divide AD patients into **early** and **late** stages

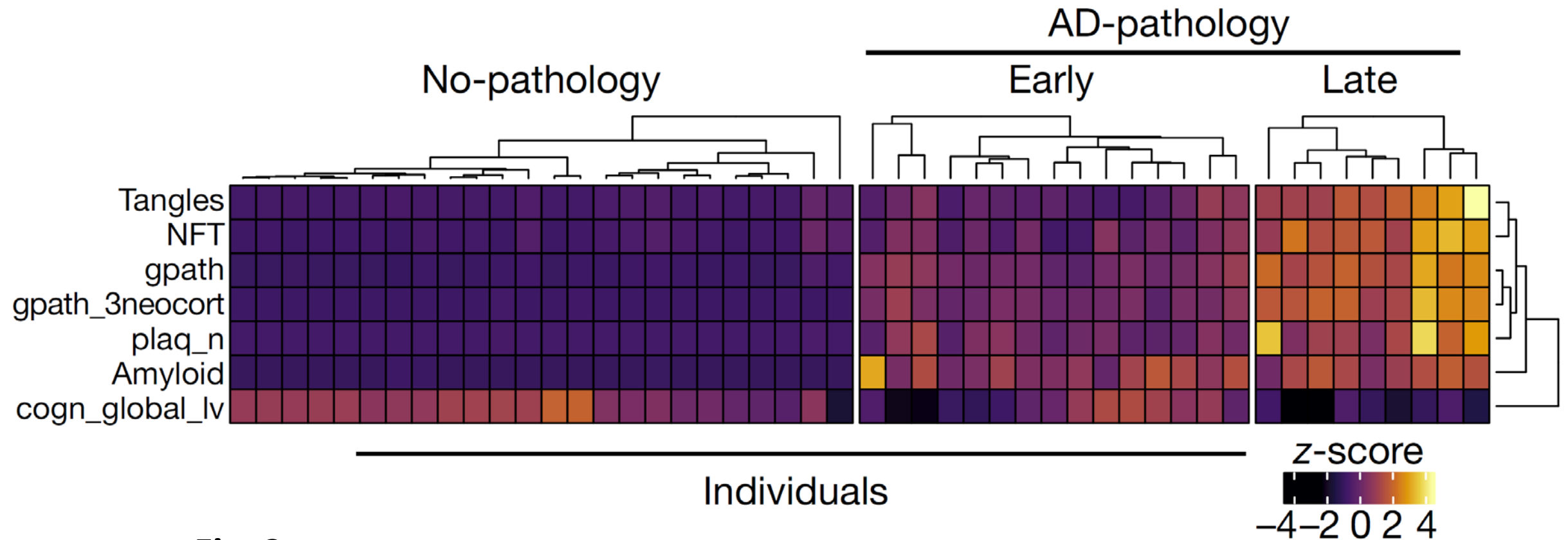


Fig. 2a

Analysis of **early stage AD** revealed that many transcriptional changes occur **before** severe pathology develops

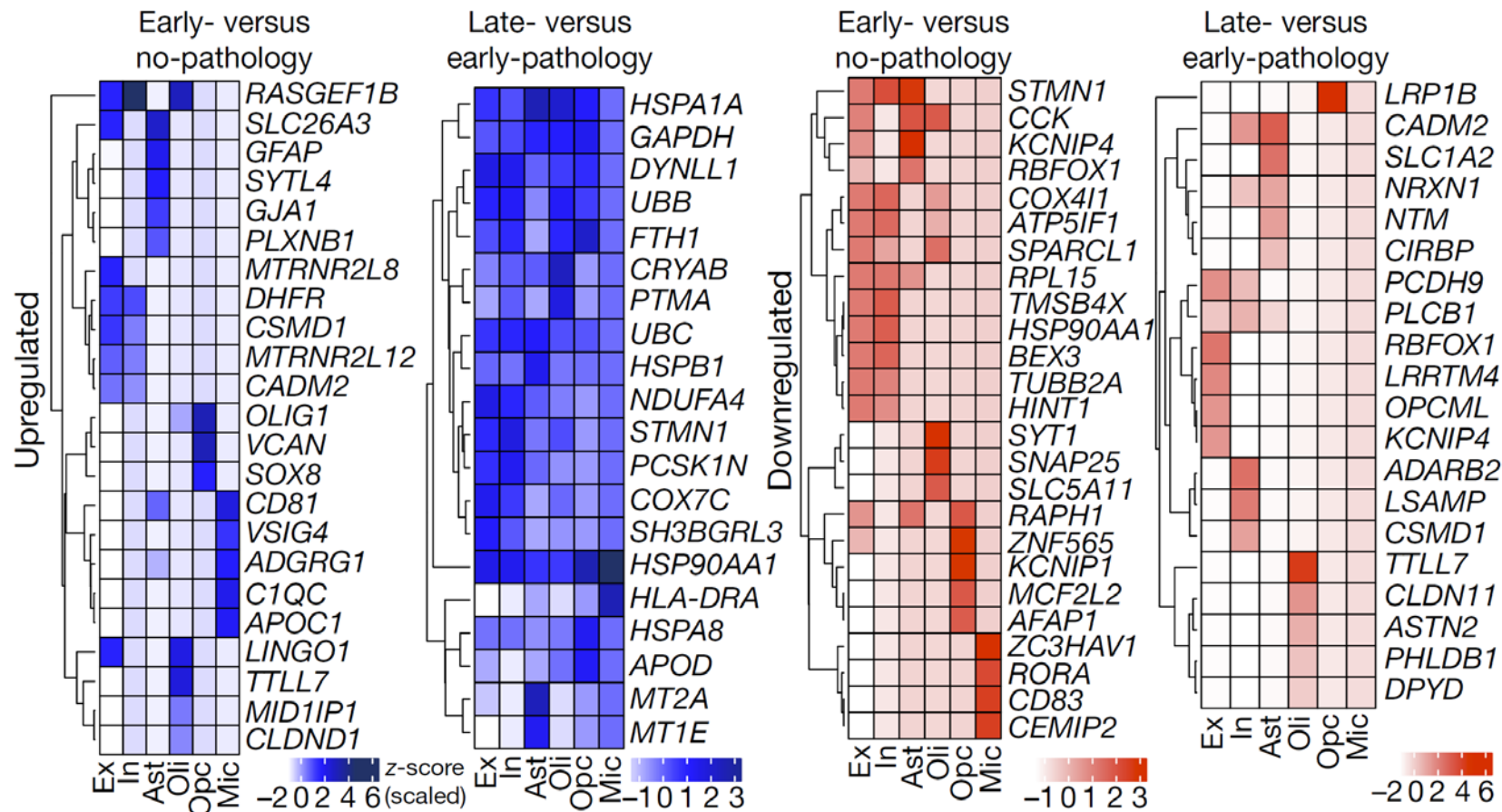


Fig. 2c

In **late AD**, many genes were **upregulated** across **all cell types**.
Many of these genes were involved in the **proteostasis network**.

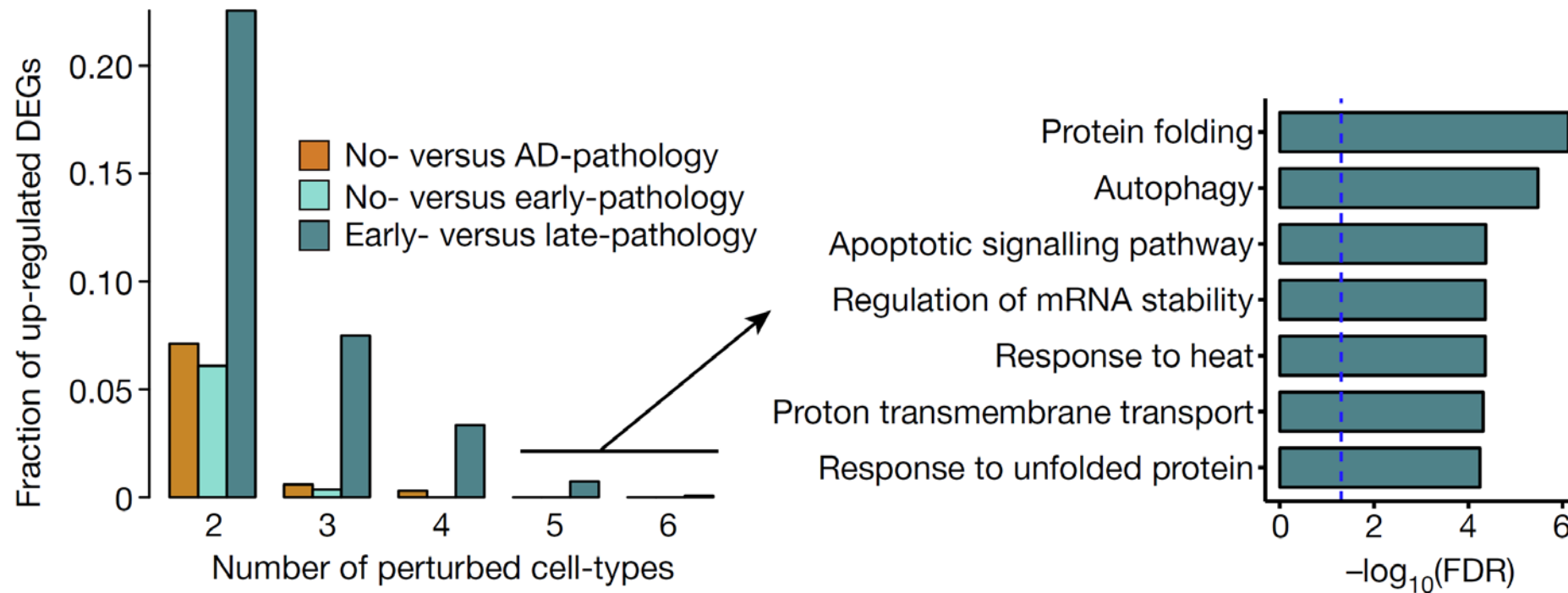


Fig. 2d, 2e

Shared upregulated genes included molecular chaperones: *HSP90AA1*, *HSPA1A*, *HSPB1*, *CRYAB*

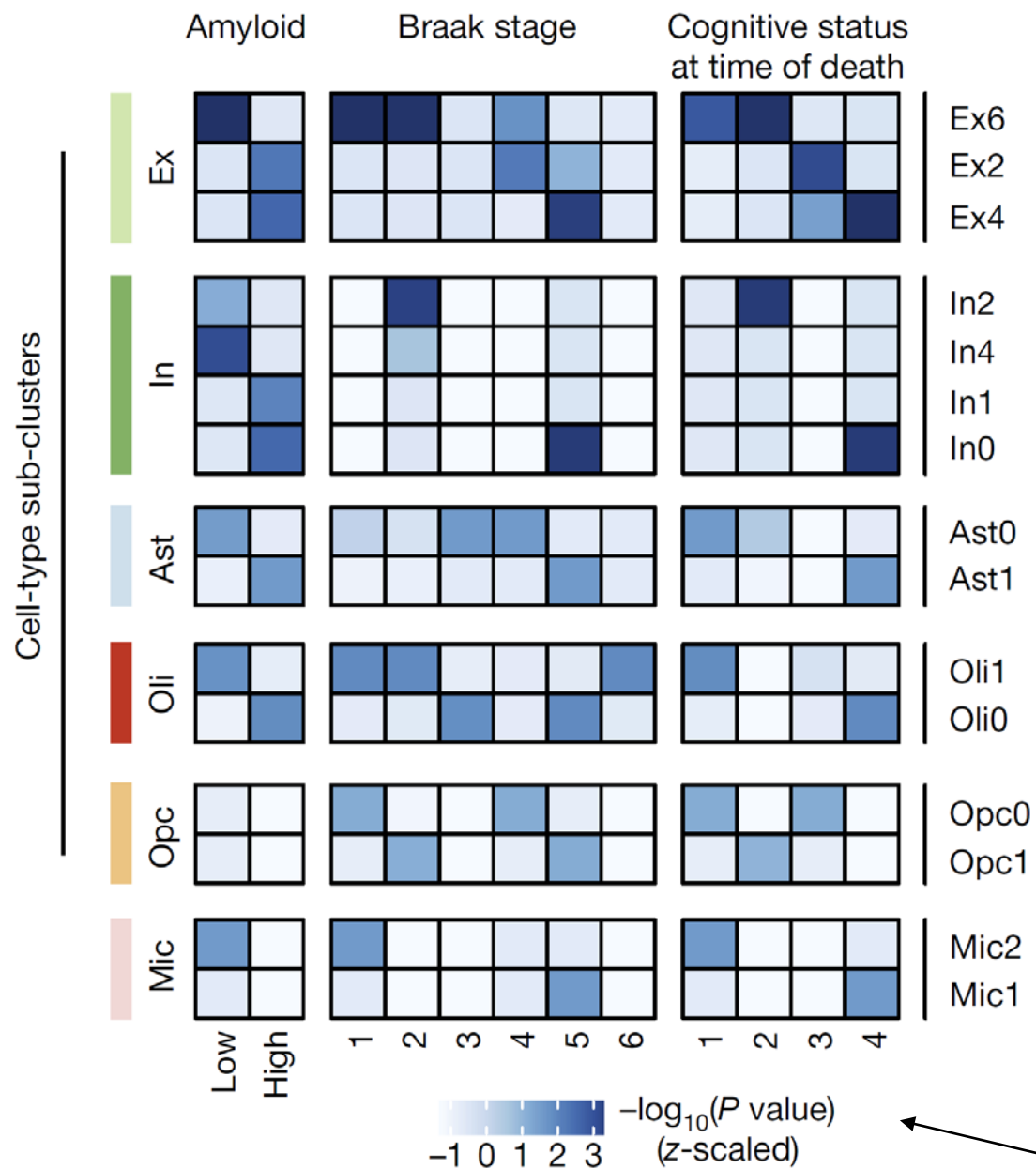


Fig. 3

- The main cell types were further divided into **sub-clusters**.
- Did some cell sub-types contain **more or fewer AD cells** than expected?
 - Ex4, In0, Ast1, and Oli0 were **associated with AD**
 - Ex6, In2, Ast0, Oli1 with **absence of pathology**

hypergeometric test

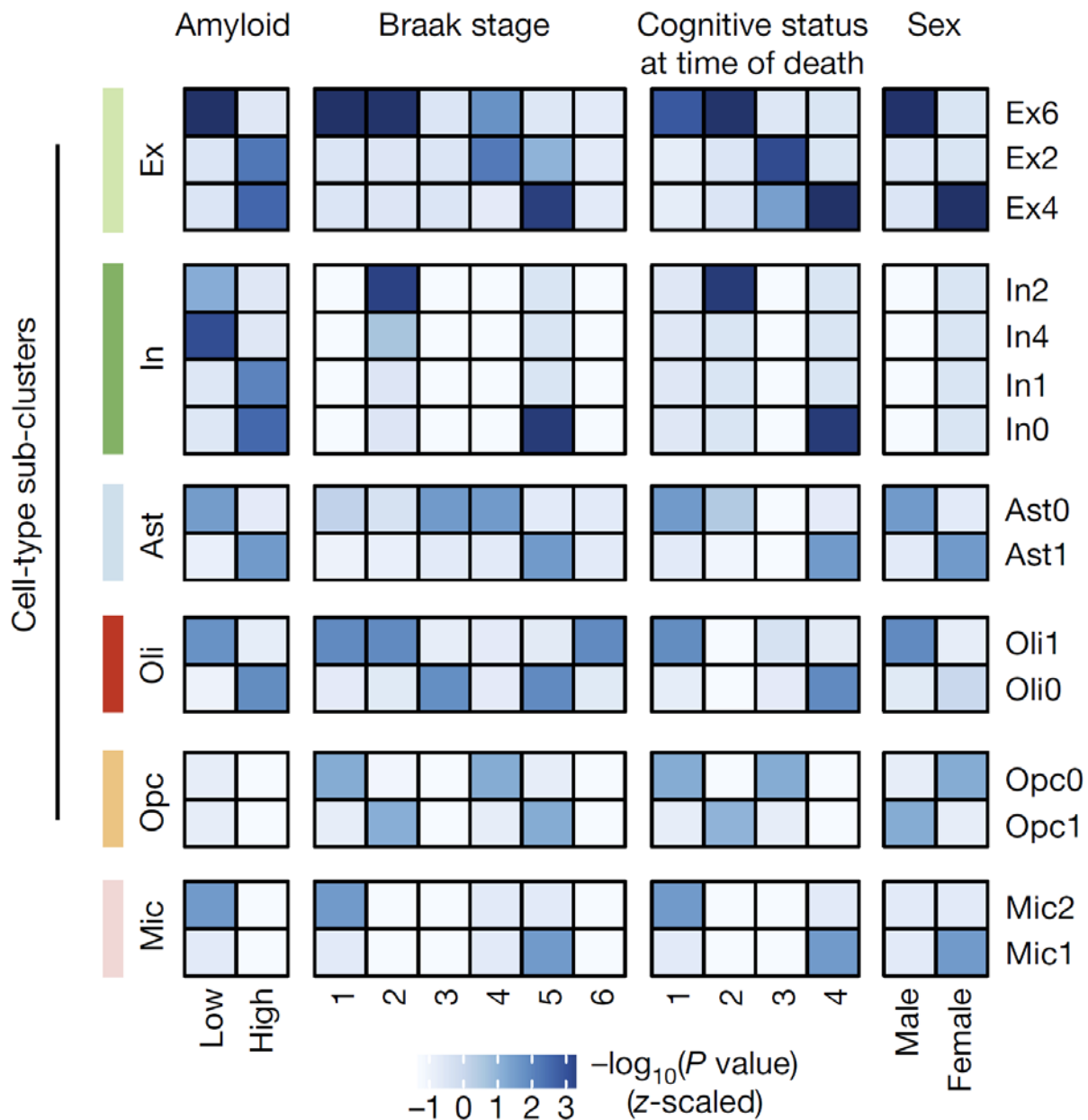


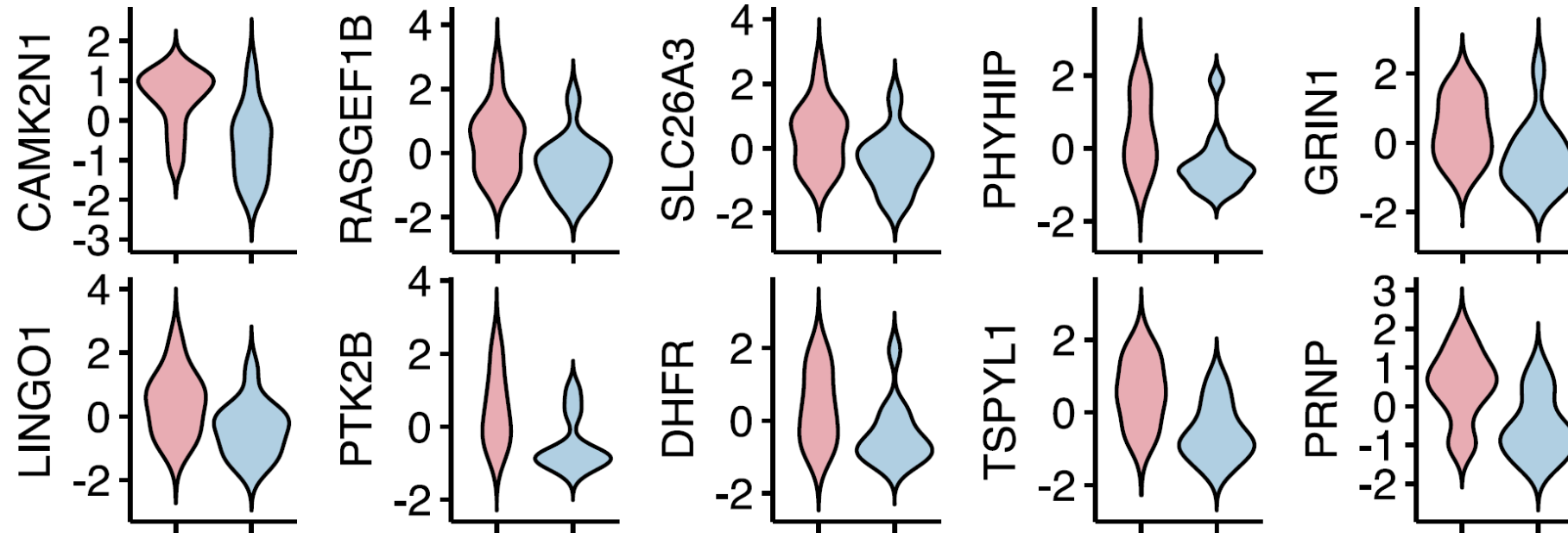
Fig. 3

Cell sub-types that were associated with AD pathology were also enriched in cells from **female** subjects.

Women are twice as likely to develop AD compared to men.

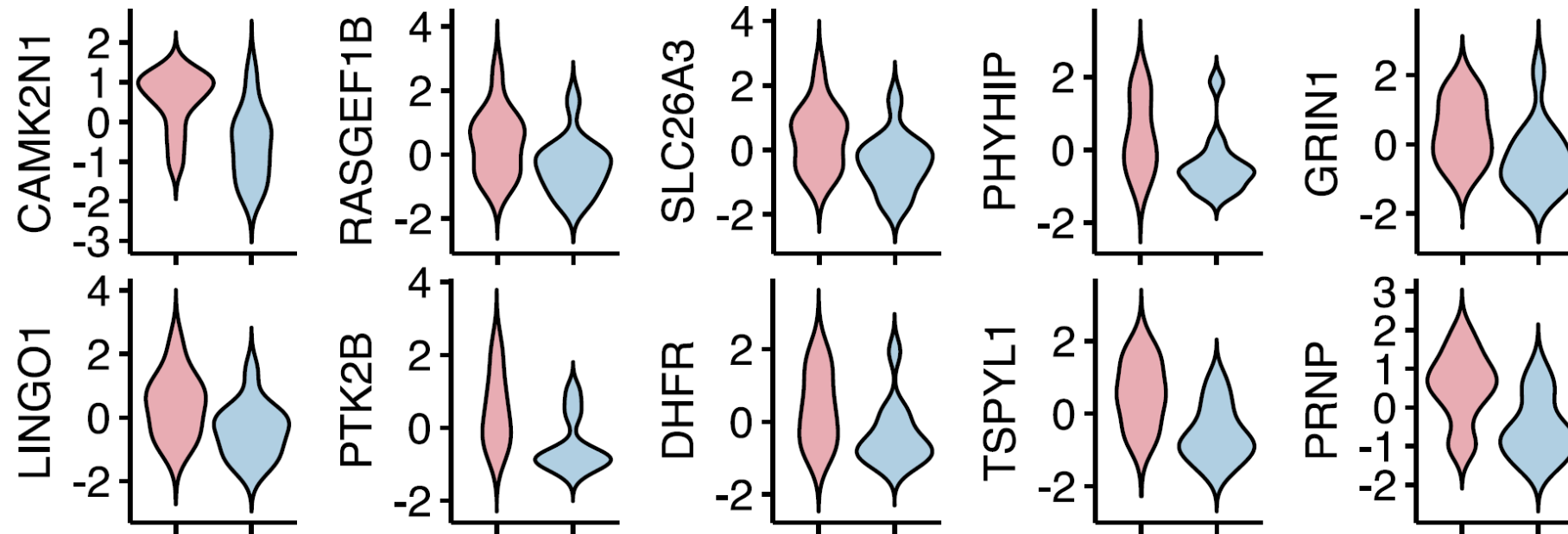
- (Hormone replacement therapy reduces the risk of AD.)
- Might transcriptomics data provide new insights into sex differences?

Ex4 markers expression in AD-pathology subjects



Expression levels across all excitatory neurons of the top 10 marker genes for the AD-associated subpopulation **Ex4** revealed higher expression in cells from females.

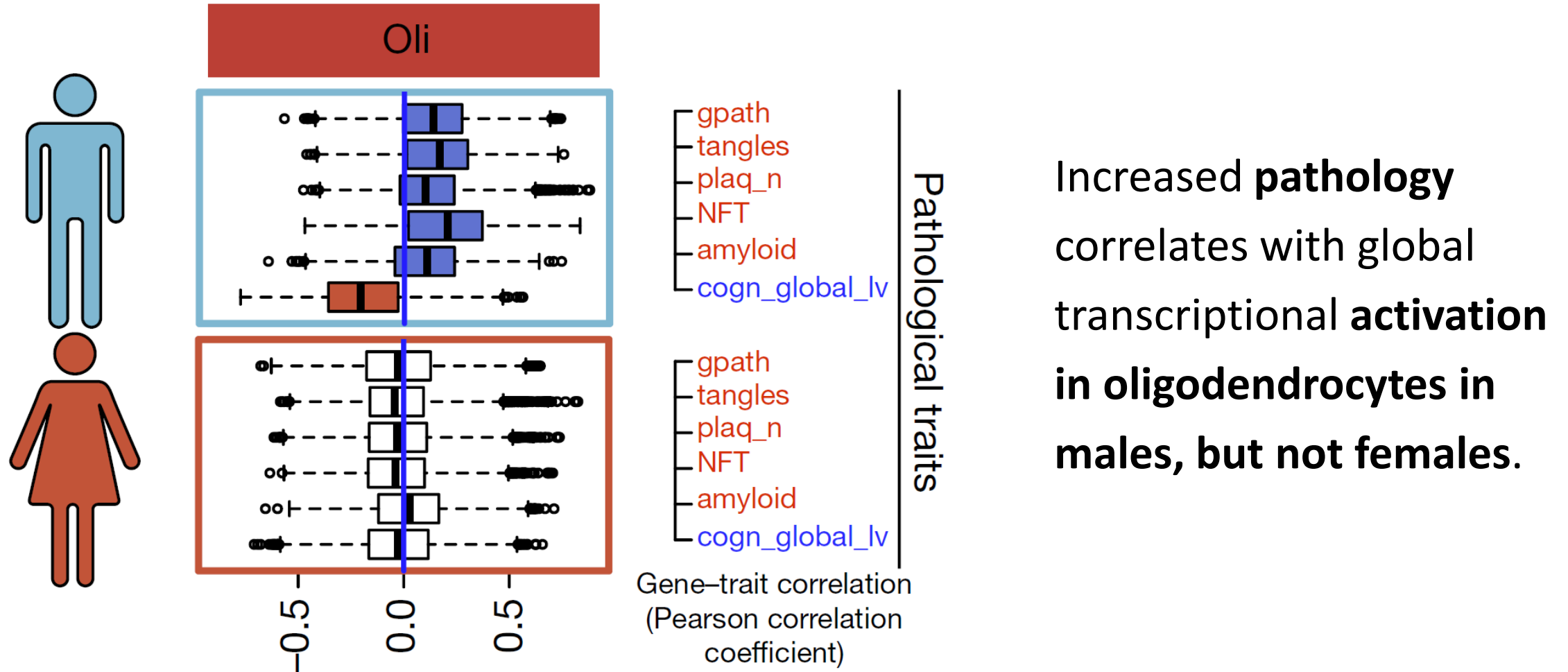
Ex4 markers expression in AD-pathology subjects



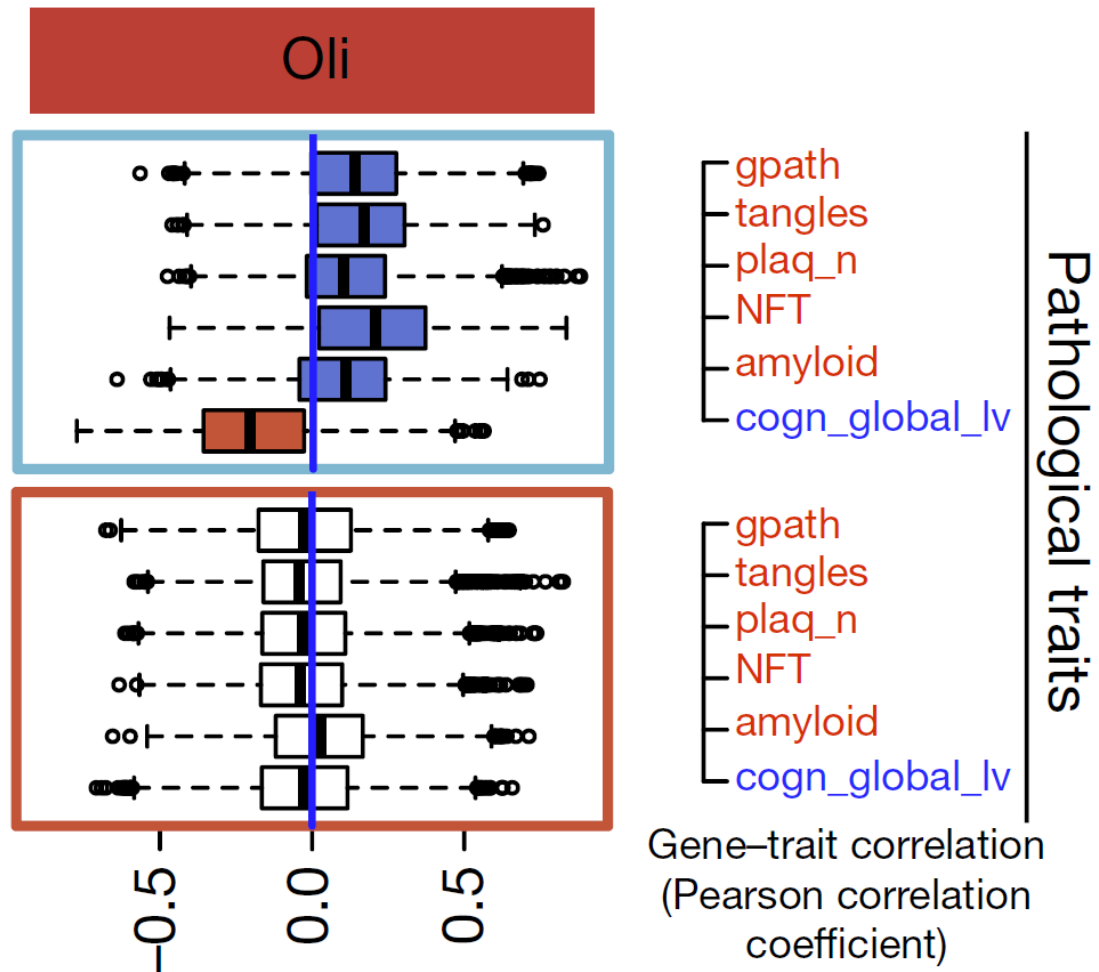
Myelination-related genes were recurrently perturbed not only in oligodendrocytes and OPCs, but also in most major cell types.

Multiple myelin-related genes were perturbed in neurons:
PRNP, CNTNAP2, ERBIN, NEGR1, and BEX1.

Female and male subjects show different cellular responses to AD



Female and male subjects show different cellular responses to AD



To assess whether a cell type “responds” to AD by up- or downregulating a gene:

- Calculate the average expression level of that gene for all cells of that subtype.
- Correlate this value with an Alzheimer’s disease score across all 48 subjects.
- This is done for all 17’926 genes to determine the distribution of Pearson correlation coefficients.

Clinicopathological variables used for the analysis



gpath
tangles
plaq_n
NFT
amyloid
cogn_global_lv

- **Global AD pathology burden** (based on 5 regions)
- **Neuronal neurofibrillary tangles** density (8 brain regions, IHC for abnormally phosphorylated tau)
- **Neuritic plaque burden** (5 regions, silver stain)
- **Neurofibrillary tangle summary** (5 regions, silver stain)
- **Amyloid- β protein** (8 regions, IHC and image analysis)
- Last **global cognitive function score** (average of 19 tests)

Female and male subjects show different cellular responses to AD

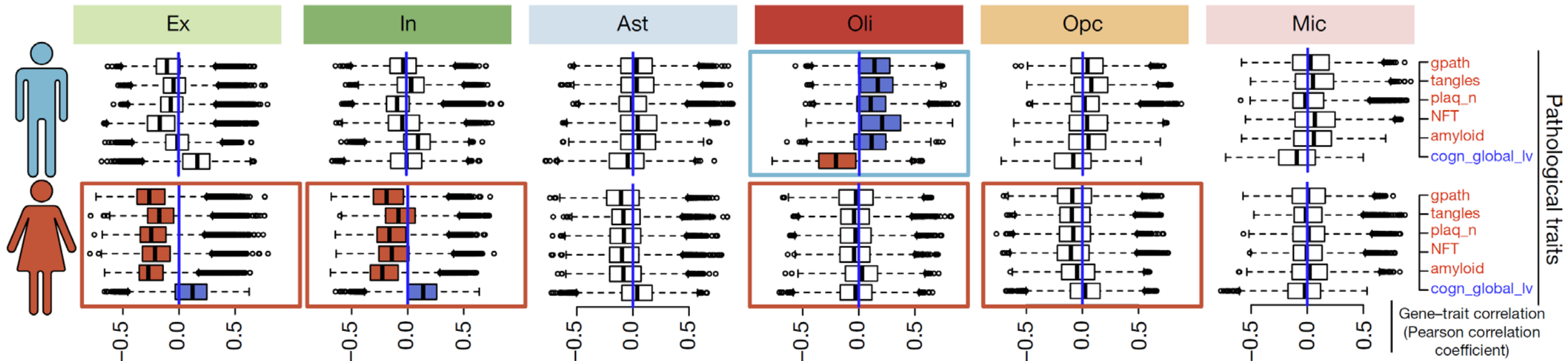


Fig. 4

Conclusions





Single-cell transcriptomic analysis of Alzheimer's disease

Hansruedi Mathys^{1,2,10}, Jose Davila-Velderrain^{3,4,10}, Zhuyu Peng^{1,2}, Fan Gao^{1,2}, Shahin Mohammadi^{3,4}, Jennie Z. Young^{1,2}, Madhvi Menon^{4,5,6}, Liang He^{3,4}, Fatema Abdurrob^{1,2}, Xueqiao Jiang^{1,2}, Anthony J. Martorell^{1,2}, Richard M. Ransohoff⁷, Brian P. Hafler^{4,5,6,8}, David A. Bennett⁹, Manolis Kellis^{3,4,11*} & Li-Huei Tsai^{1,2,4,11*}

- sn-RNAseq was able to produce a high-quality dataset from frozen post-mortem samples from Alzheimer's patients.
 - Known cell types and subtypes were well recapitulated.
- Much of the complexity of AD pathology, especially in early AD, would have been lost in bulk sequencing.
- Since it was purely descriptive, the study **cannot distinguish between responsive vs. driving changes**, that is, neuroprotection vs. pathogenicity.

ARTICLE

A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys

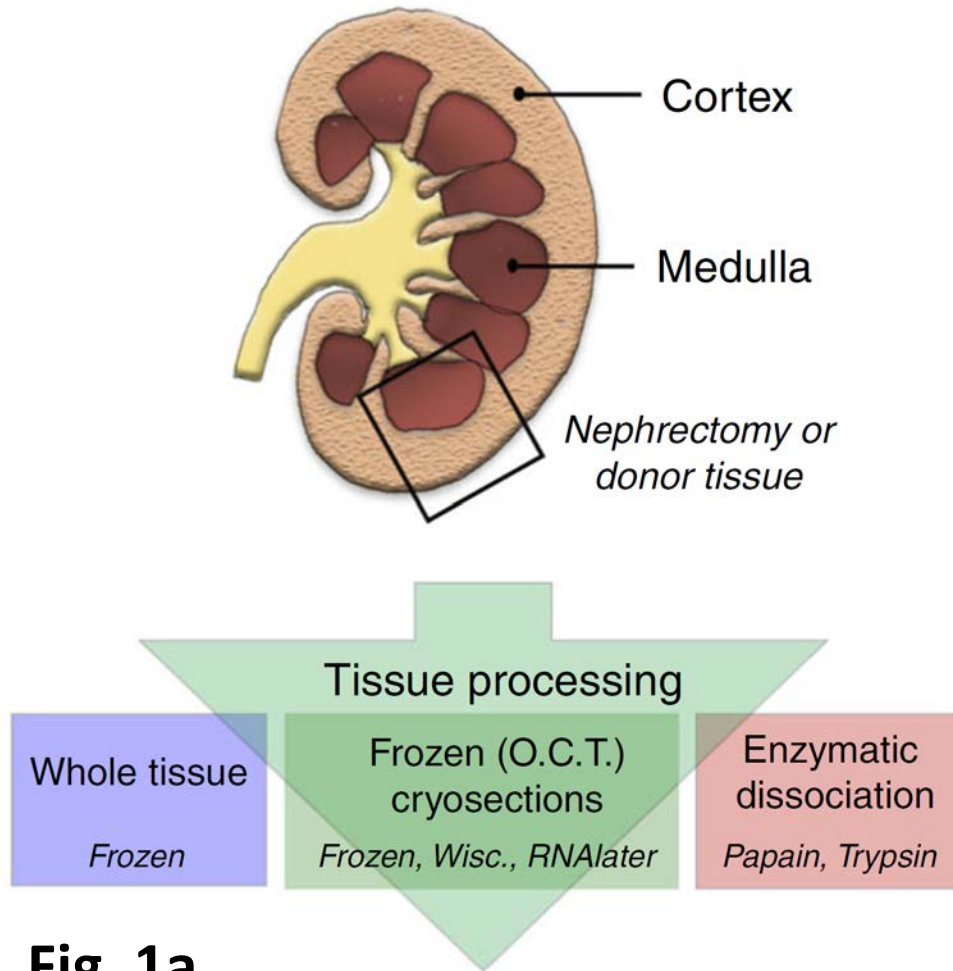
Blue B. Lake ^{1,6}, Song Chen^{1,6}, Masato Hoshi ^{2,6}, Nongluk Plongthongkum^{1,3,6}, Diane Salamon^{2,4}, Amanda Knoten², Anitha Vijayan², Ramakrishna Venkatesh⁵, Eric H. Kim⁵, Derek Gao¹, Joseph Gaut², Kun Zhang ¹ & Sanjay Jain ^{2,4}

June 2019

Human kidney samples were collected in two centres

- Washington University:
 - Patients undergoing partial or total nephrectomy
 - Discarded deceased kidney donors
 - University of Michigan:
 - Patients undergoing tumour nephrectomies
- Libraries were prepared from 94 patients
- 15 patients were included in the final analysis
(processed using the optimized protocol)

The authors tested various tissue processing methods



- Whole fresh-frozen tissue proved **unsuitable** for snRNA-seq of kidney tissue.
- Nuclei isolated from fresh tissues that were enzymatically dissociated (with papain/collagenase or trypsin/collagenase) were compatible with snRNA-seq.
- However, the **best data** quality was seen in samples that were embedded in O.C.T. and then **cryosectioned** before isolation of nuclei.

Fig. 1a

Tissue samples were cut to the right size, cleaned of blood, and then embedded in optimal cutting temperature (O.C.T.) blocks

A. Cryomold with OCT to bathe tissue at room temperature



B. Empty cryomold prechilled on powdered dry ice



A'. Bathe biopsy in OCT - 2 min in cryomold in A



B'. Transfer biopsy to prechilled cryomold in B, cover with OCT



C. Frozen OCT with embedded tissue



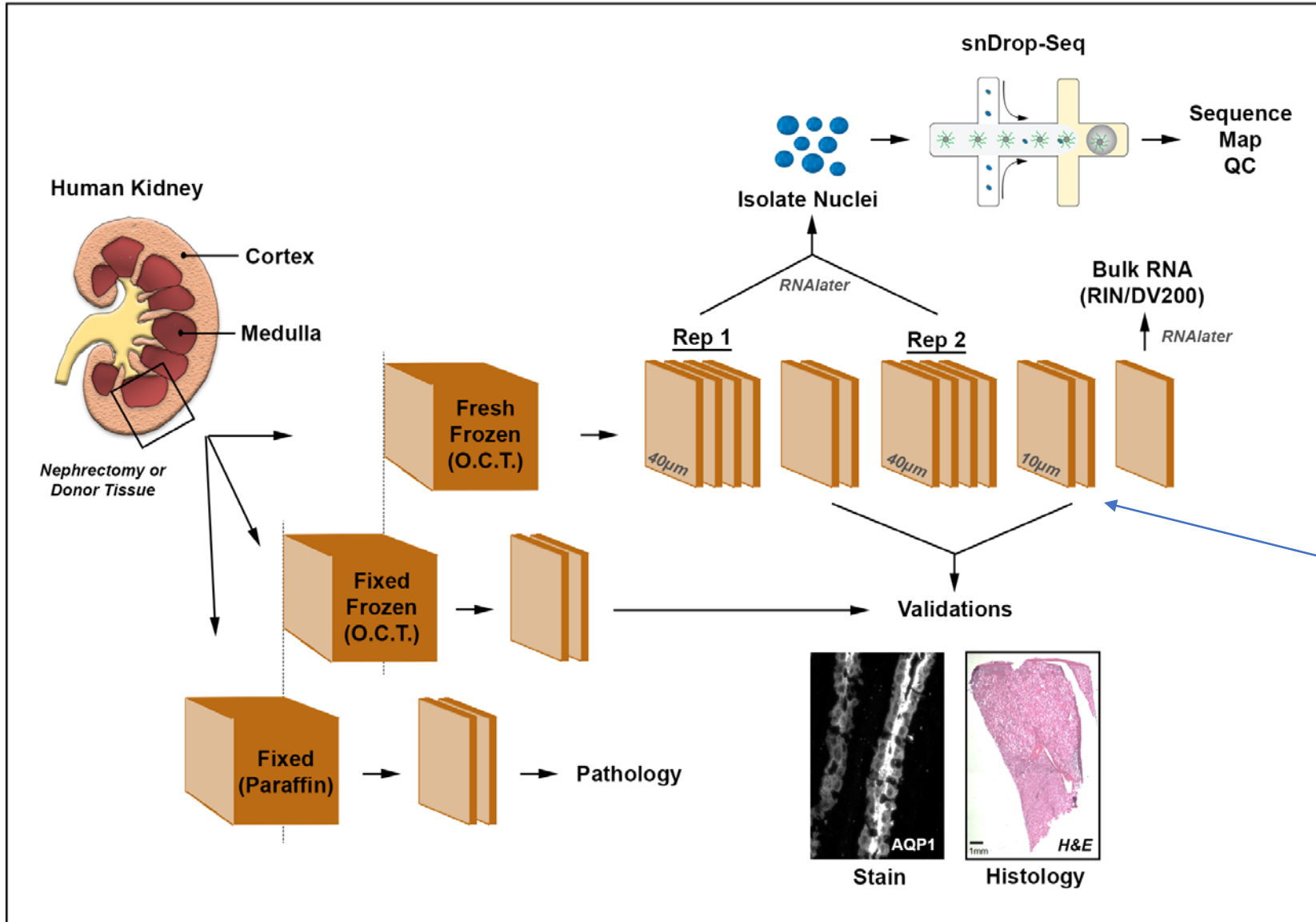
→ Storage at -80°C

Next, O.C.T. blocks were cryosectioned ($7 \times 35 \mu\text{m}$ cryosection rolls) and collected in tubes containing **RNAlater** solution



- Storage at -20°C until isolation of nuclei
- **>90% success rate**

Tissue Processing Pipeline



Adjacent thin sections
can be used for
assessment of tissue
integrity and correlation
with protein expression.

Fig. S12

Using their optimized pipeline, the authors observed **better representation** of the very many kidney cell types than in previous scRNA-seq datasets.

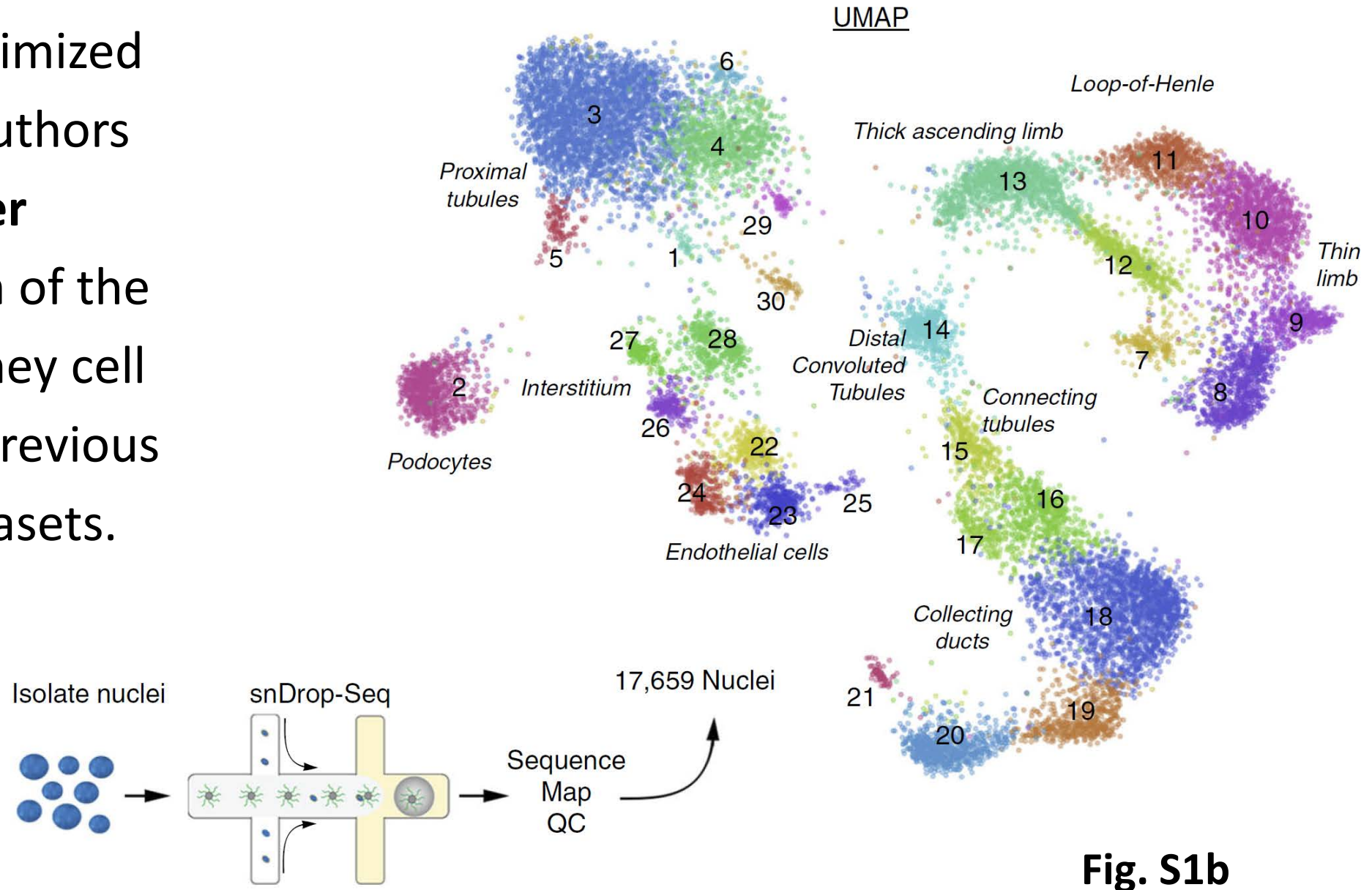
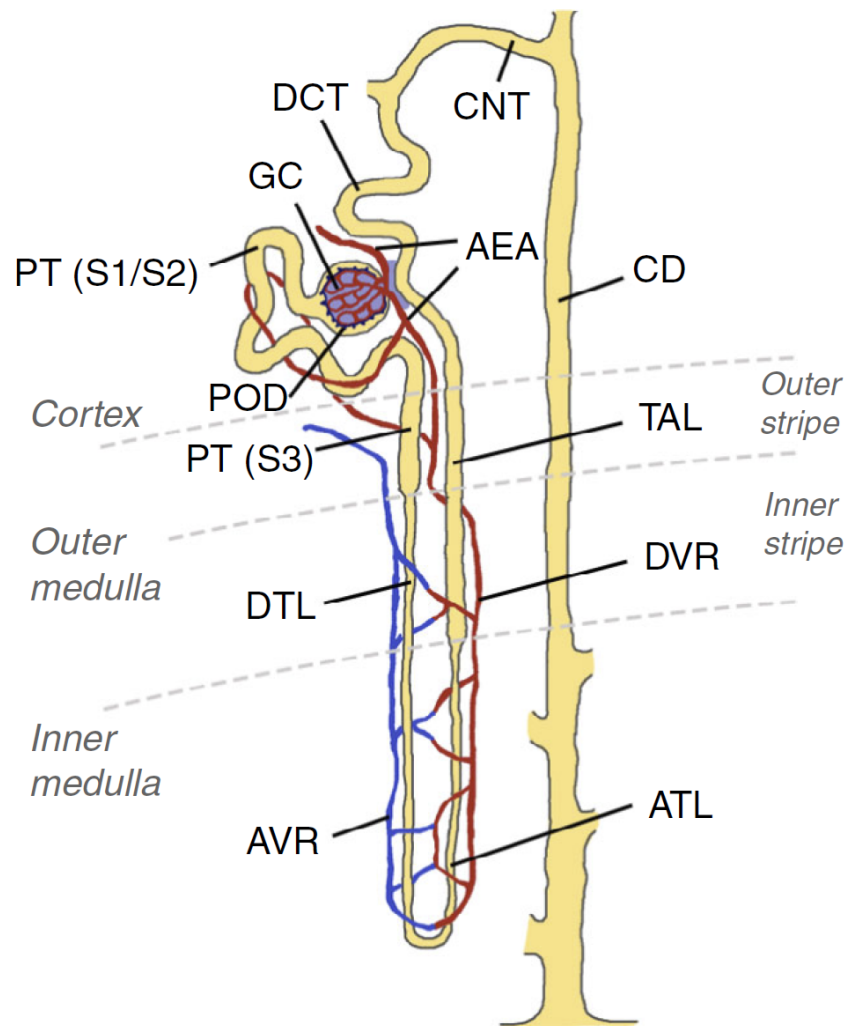


Fig. S1b



Cluster	Abbn	Annotation	# Nuc	Region sampled
1	EPC	Epithelial cells	55	0 1
2	POD	Podocytes	859	0 1
3	PT-1	Proximal tubule Ep. cells (S1/S2)	3238	0 1
4	PT-2	Proximal tubule Ep. cells (S2)	920	0 1
5	PT-3	PT Ep. cells—inflam. response	124	0 1
6	PT-4	PT Ep. cells—fibrinogen+ (S3)	86	0 1
7	PT-5	Proximal tubule Ep. cells (S3)	219	0 1
8	DTL	Descending thin limb	844	0 1
9	ATL-1	Thin ascending limb	431	0 1
10	ATL-2	Thin ascending limb	1325	0 1
11	ATL-3	Thin ascending limb	623	0 1
12	TAL-1	Thick ascending limb—surgery	536	0 1
13	TAL-2	Thick ascending limb	1217	0 1
14	DCT	Distal convoluted tubule	568	0 1
15	CNT	Connecting tubule	395	0 1
16	PC-1	Collecting system—PCs (cortex)	663	0 1
17	PC-2	Collecting system—PCs—stressed	208	0 1
18	PC-3	Collecting duct—principal cells	1970	0 1
19	IC-A2	Collecting duct—intercalated cells	540	0 1
20	IC-A1	Collecting duct—intercalated cells	772	0 1
21	IC-B	Collecting duct—intercalated cells	84	0 1
22	EC-1	Endothelial cells—glomerular cap.	344	0 1
23	EC-2	Endothelial cells—AVR	366	0 1
24	EC-3	Endothelial cells—AEA & DVR	281	0 1
25	EC-4	Endothelial cells	42	0 1
26	MC	Mesangial cells	215	0 1
27	vSMC/P	Vas. Sm. muscle cells/pericytes	168	0 1
28	INT	Interstitial	410	0 1
29	Unk	Unknown—novel PT (S2) Ep. cells?	68	0 1
30	IMM	Immune cells—macrophages	88	0 1

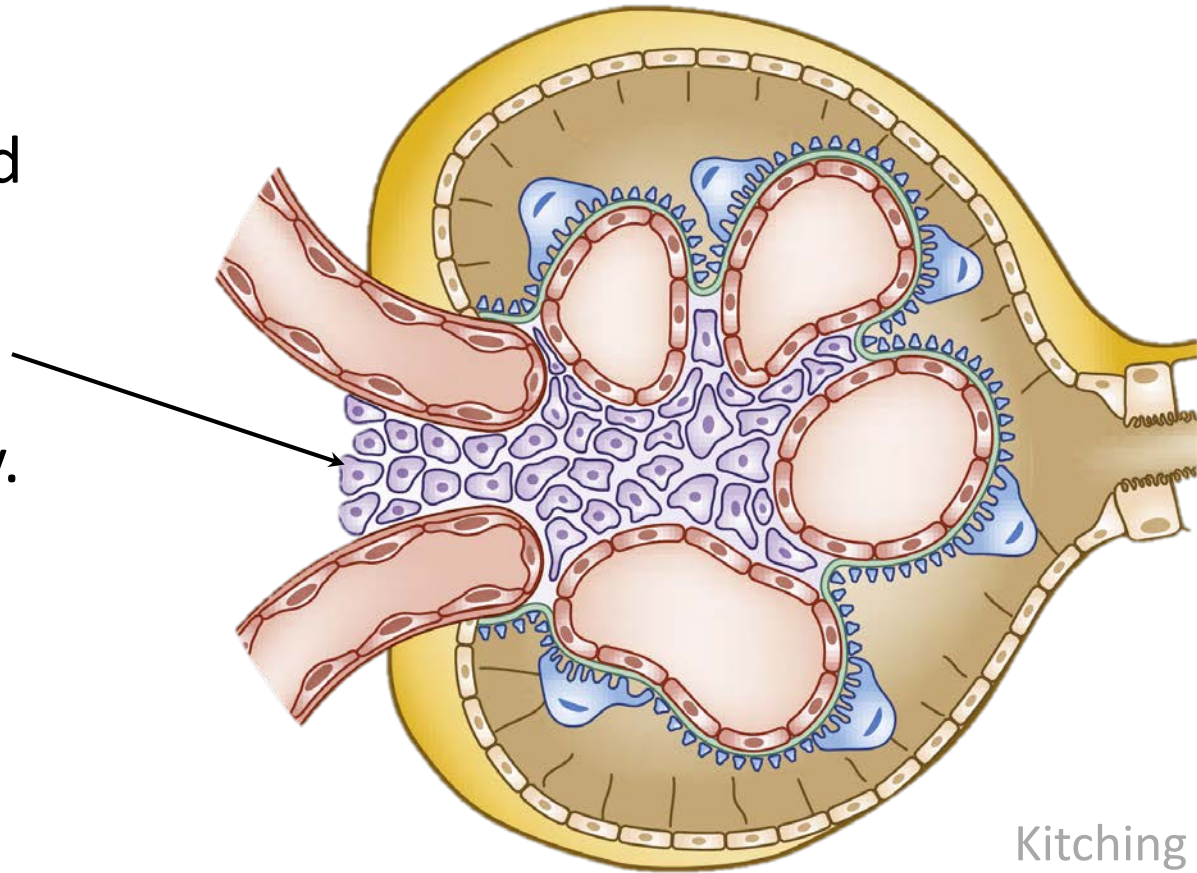
Cortex
 Medulla
 Both

Fig. 1c, 1d

PIEZO2, which encodes a stretch-gated ion channel involved in **mechanosensation**, was identified as a **new marker of mesangial cells**

Mesangial cells are specialized glomerular pericytes that are crucial for maintaining appropriate filtration capacity.

Glomerular cells are under-represented in scRNA-seq datasets (Wu et al, J Am Soc Nephrol, 2019).



Kitching and Hutton,
2016

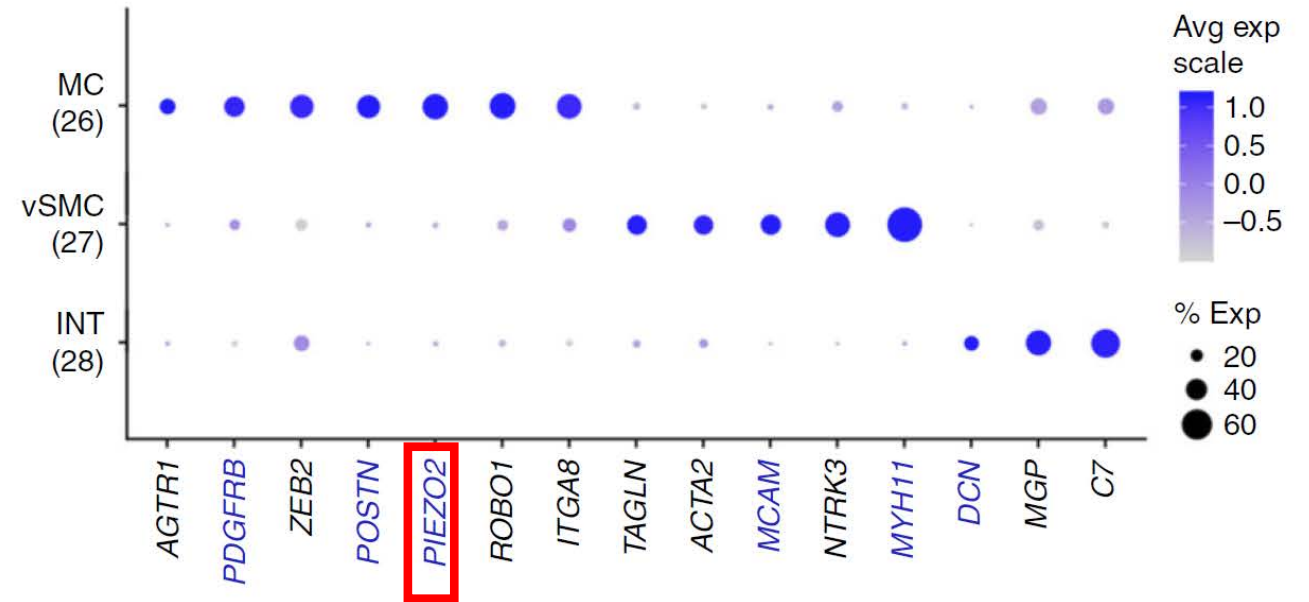
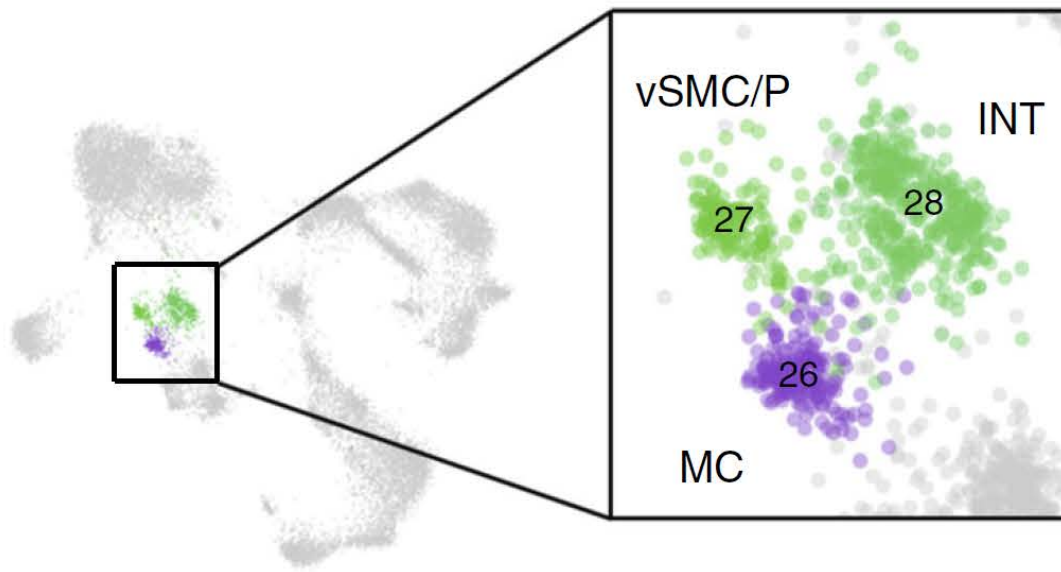


Fig. 5e, 5f

vSMC/P = Vascular smooth muscle cells/pericytes; INT = interstitial cells; MC = mesangial cells

Interestingly, ***PIEZO2*** has recently been identified as a novel **hypertension locus**.

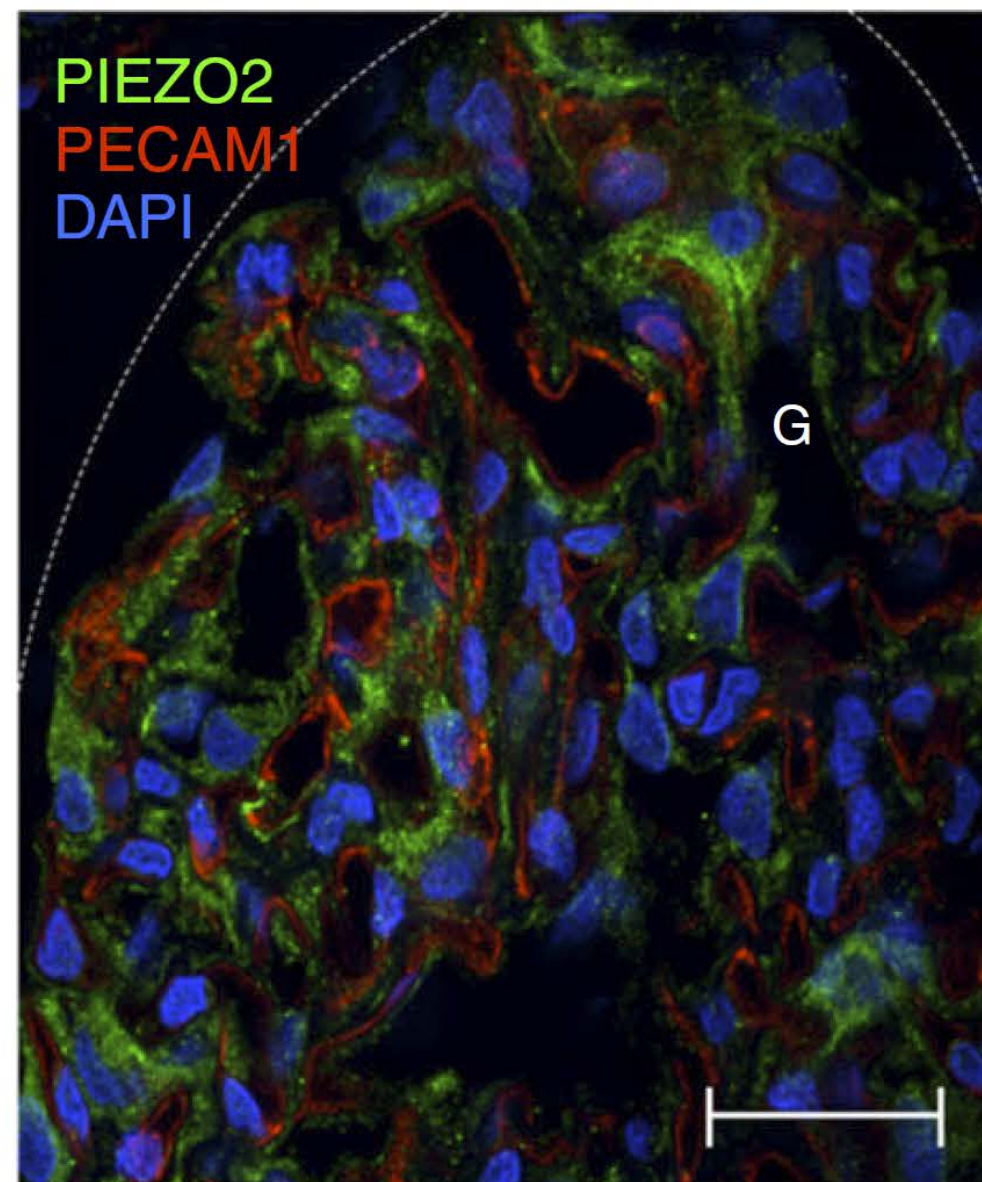
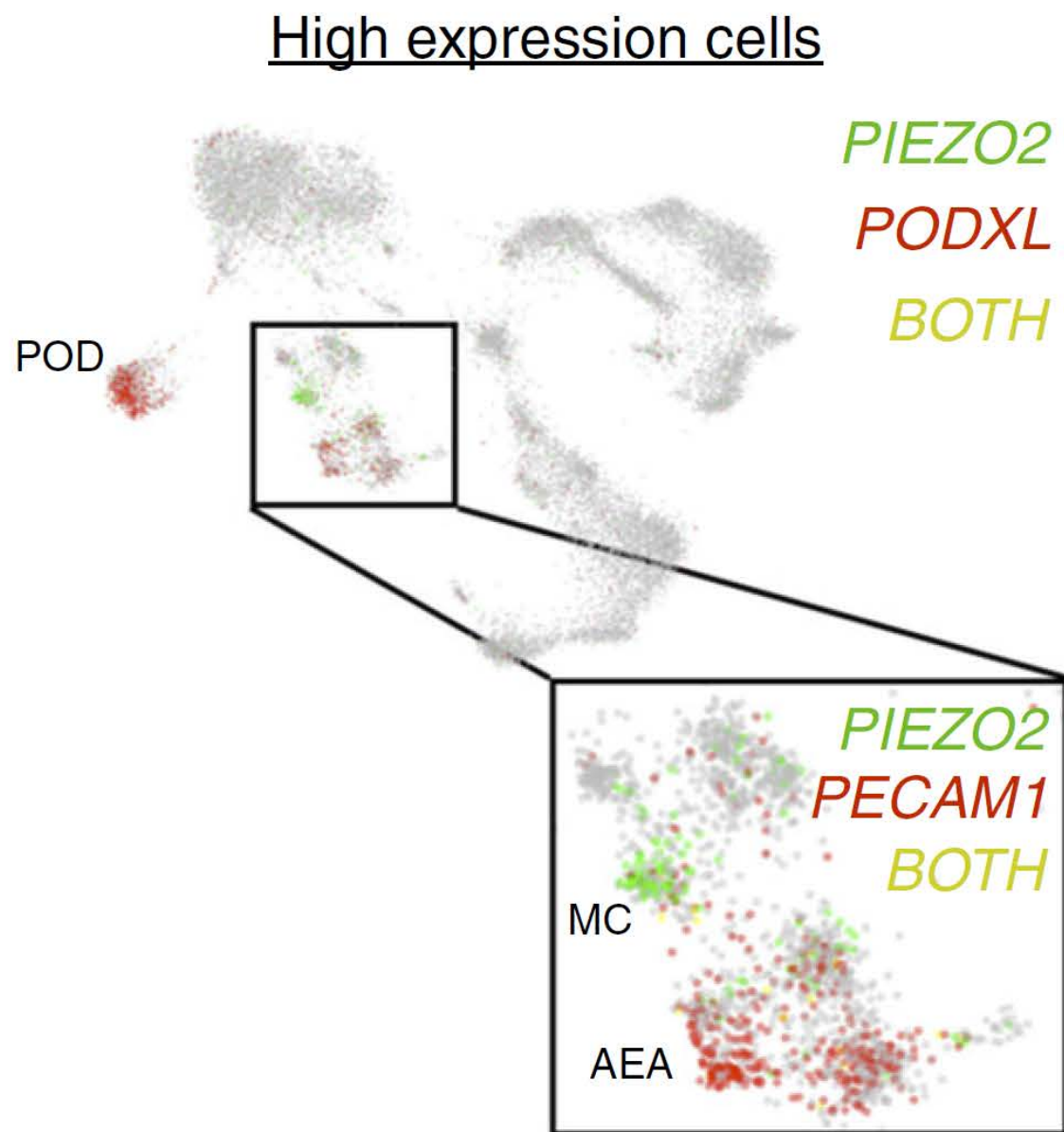


Fig. 5h, 5i

ARTICLE

A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys

Blue B. Lake^{1,6}, Song Chen^{1,6}, Masato Hoshi^{2,6}, Nongluk Plongthongkum^{1,3,6}, Diane Salamon^{2,4}, Amanda Knoten², Anitha Vijayan², Ramakrishna Venkatesh⁵, Eric H. Kim⁵, Derek Gao¹, Joseph Gaut², Kun Zhang¹ & Sanjay Jain^{2,4}

Conclusions

- Careful optimization of the protocol is required when applying snRNA-seq to a new tissue.
- For human kidney, snRNA-seq seems to produce data of high quality:
 - Clusters more accurately represent the expected histological composition compared to scRNA-seq.
 - Fewer stress-induced transcriptional changes are observed (no enzymatic dissociation at 37°C is required).
 - Gene detection sensitivity is sufficient.

Thank you for your attention!