Single Nuclei RNA Sequencing Reveals Gene Expression Alterations in Alzheimer’s Disease

Master’s Thesis

Presented to

The Faculty of the Graduate School of Arts and Sciences
Brandeis University
Department of Neuroscience
Dr. Avital Rodal, Advisor

In Partial Fulfillment
of the Requirements for the Degree

Master of Science
in
Neuroscience

by
Zhuyu Peng

May 2017
Copyright by
Zhuyu Peng

© 2017
Acknowledgments

I wish to express my heartfelt gratitude to Dr. Li-Huei Tsai for giving me this wonderful opportunity and welcoming me in her lab for my master’s thesis.

I also wish to express my most genuine appreciation to Dr. Avital Rodal for being my master’s thesis supervisor. Your insightful guidance and suggestions are invaluable to me.

I will forever be thankful to my amazing mentor Dr. Hansruedi Mathys, who stood by me in every step. Thank you for your patience, remarkable support and brilliant comments. You are a fantastic role model.

Special thanks to my parents. Thank you for allowing me to be ambitious.
Carried out in the Tsai Laboratory
Massachusetts Institute of Technology
Dr. Li-Huei Tsai, Advisor
ABSTRACT

Single Nuclei RNA Sequencing Reveals Gene Expression Alterations in Alzheimer’s Disease

A thesis presented to the Department of Neuroscience

Graduate School of Arts and Sciences
Brandeis University
Waltham, Massachusetts

By Zhuyu Peng

Alzheimer’s disease is a severe neurodegenerative disorder that results in loss of memory and cognition (Canter et al. 2016). In order to (1) investigate whether and how gene expression in the molecularly distinct cell types of the human brain is affected in Alzheimer’s disease and (2) to map the cellular diversity of the aged human brain at single-cell resolution on a genome-wide scale, we analyzed human post-mortem brain tissue of 12 AD patients and 12 aged-matched normal individuals using single-nucleus RNA-sequencing. A total of 3264 adult human postmortem prefrontal cortex nuclei were sequenced. We identified six distinct cell types and two pathology associated cell clusters in the human postmortem prefrontal cortex tissue. Finally, data analysis revealed dramatic differences in gene expression in both neurons and oligodendrocytes of individuals carrying one or more copies of the APOE4 allele compared to subjects that do not carry an APOE4 allele.
# Table of Contents

ACKNOWLEDGMENTS ........................................................................................................ III
ABSTRACT ............................................................................................................................ V

I. INTRODUCTION ................................................................................................................ 1
  1.1 ALZHEIMER’S DISEASE ............................................................................................ 1
    1.1.1 Epidemiology ................................................................................................. 1
  1.2 NEUROPATHOLOGICAL HALLMARKS ................................................................... 2
    1.2.1 Amyloid plaques ............................................................................................ 2
    1.2.2 Neurofibrillary tangles .................................................................................. 4
    1.2.3 Synaptic loss .................................................................................................. 5
    1.2.4 Genetic Landscape ......................................................................................... 5
  1.3 NEUROINFLAMMATION ....................................................................................... 7
    1.3.1 Microglia ........................................................................................................ 8
    1.3.2 Astrocyte ...................................................................................................... 9
    1.3.3 Oligodendrocyte ........................................................................................... 10
  1.4 SINGLE CELL RNA SEQUENCING .................................................................... 11
    1.4.1 Single Cell RNA Sequencing workflow ....................................................... 12

II. RESULTS ......................................................................................................................... 14
  2.1 CELL ISOLATION AND LIBRARY PREPARATION ..................................................... 14
  2.2 SIX DISTINCT CELL TYPES ARE IDENTIFIED IN CONTROL GROUP ..................... 16
  2.3 TWO PATHOLOGY-ASSOCIATED CLUSTERS ARE IDENTIFIED ........................... 20
  2.4 DIFFERENTIAL EXPRESSED GENES IN PATHOLOGY-ASSOCIATED CLUSTERS .... 24
  2.5 SCDE ON PATHOLOGY INDIVIDUALS AND CONTROL GROUP ......................... 30
  2.6 APOE ....................................................................................................................... 31

III. MATERIAL AND METHODS ...................................................................................... 35
  3.1 BRAIN TISSUE ....................................................................................................... 35
  3.2 TISSUE DISOCIATION ......................................................................................... 35
  3.3 SINGLE CELL CAPTURE VIA FLUORESCENCE ACTIVATED CELLS SORTING ........ 36
    3.4.1 RNA SPRI Beads Clean-up ........................................................................... 38
    3.4.2 Reverse Transcription ................................................................................... 39
    3.4.3 Whole Transcriptome Amplification ............................................................. 39
    3.4.4 DNA SPRI Beads Clean-up ........................................................................... 40
    3.4.5 Construct Nextera XT Sequencing Library Using Nextera® XT Library Prep Kit .... 41
    3.4.6 DNA SPRI Clean-up and Pooling ................................................................ 41

IV. DISCUSSION ............................................................................................................... 46
V. SUPPLEMENTARY FIGURE ....................................................................................... 49
VI. REFERENCES .......................................................................................................... 54
I. Introduction

1.1 Alzheimer’s Disease

At the 37th Meeting of South-West German Psychiatrist in Tubingen on November 3, 1906, Dr. Alois Alzheimer presented the case of his patient Auguste Deter, a 51-year-old female who exhibited unusual symptoms including aggression, memory disturbance, and disorientation. Dr. Alzheimer also reported the characteristic anatomical features found in her brain autopsy but excited little interest among the audience. However, today, this peculiar disease named after Dr. Alzheimer, has become one of the most devastating causes of death and affects an estimated 44 million people worldwide ("WHO | Dementia", 2016).

Alzheimer’s disease (AD) is a fatal progressive neurodegenerative disorder characterized by cognitive dysfunction, severe memory loss, and semantic impairment. Pathologically, intracellular neurofibrillary tangles and senile plaques have been identified as core hallmarks for AD, while the underlying biological mechanisms that trigger AD still remain controversial. Substantial evidence suggests that AD pathogenesis is not restricted to neurons but also involves a dysfunction in glial cells (Heneka et al., 2015).

1.1.1 Epidemiology

Alzheimer’s disease is a progressive, neurodegenerative disorder that affected nearly 44 million people around the world in 2015. More than 5.5 million people are living with AD in the US, and this number is predicted to rise to 16 million by 2050 (Alzheimer Dement, 2008).
Research showed that the life expectancy of AD patients after diagnosis varies from 4 to 8 years and the mortality rate rises dramatically as the age increases. From 2000 to 2013, the mortality rate of AD as the underlying cause of death has increased 23 percent for patients aging from 75 to 84.

1.2 Neuropathological hallmarks

1.2.1 Amyloid plaques

As one of the pathological hallmarks of Alzheimer's Disease, amyloid plaques are abnormal extracellular protein aggregates, which are believed to drive AD pathogenesis (Cummings et al., 1996). There are two categories of amyloid plaques: diffuse Aβ plaques (Aβ Ps) and neuritic plaques (NPs). Having been widely associated with synaptic dysfunction and neuronal death, neuritic plaques are spherical protein aggregation with a dense core composed of Aβ (Knowles et al. 1999). In the AD brain, microglia and astrocytes are often found in the vicinity to the neuritic plaques. In contrast, no evidence suggest microglia/astrocytes activation in the presence of diffuse plaques (DaRocha-Souto et al., 2011).

Aβ is a 38 to 43 amino acid peptide, generated via sequential proteolytic cleavage of the amyloid beta precursor protein (APP) into a heterogeneous group of peptides of varying lengths. Three crucial catalyzing enzymes are involved in this enzymatic process: α-, β-, γ-Secretase. When α-Secretase cuts APP followed by γ-Secretase, non-amyloidogenic products are generated, including sAPPbeta, p3, C83 and ACID50 (Tanabe et al., 2006).
Fig. 1 Proteolytic mechanisms of APP (Images from Vardy et al., 2005).

However, in Alzheimer’s Disease, APP is often cut first by β-Secretase followed by γ-Secretase, yielding AD pathology-associated Aβ peptides. In this case, because the cleavage site of γ-Secretase can be different, the length of Aβ can vary from 39 to 43 amino acids, among which Aβ_{40} and Aβ_{42} are dominant forms in the brain. Although Aβ_{40} is more abundant in cerebral spinal fluid, the more insoluble Aβ_{42} has been demonstrated to be the major component of senile plaques in AD, while Aβ_{40} is only found in a few set of plaques (Miller et al., 1993). Hypotheses suggest that Aβ_{42} is more hydrophobic and more amyloidogenic, thus having a higher propensity to aggregate, but the mechanisms of the preferential deposition remain to be elucidated.

Moreover, emerging studies showed that changes in the Aβ_{42} to Aβ_{40} ratio dramatically alter the properties of the mixture. For instance, a higher Aβ_{42}/Aβ_{40}
ratio seems to be associated with the early onset of presenilin mutation induced AD (Dorey et al., 2015). Although the presence of \( \beta \)-amyloid plaques and NFT are standards for AD diagnosis, several studies indicate that the level of \( \beta \)-amyloid plaques poorly correlates with severity of cognitive deficits (Vardy et al., 2005).

1.2.2 Neurofibrillary tangles

As another characteristic marker for AD, neurofibrillary tangles (NFT) are intracellular depositions of hyperphosphorylated and misfolded tau proteins. Tau is a phosphoprotein regulating the polymerization and stabilization of microtubules (Kirschner et al, 1975), an important structure for intra-neuronal cargo trafficking. In the human brain, there exist six isoforms of tau protein, expressions of which are regulated developmentally by alternative splicing of pre-mRNA generated from the tau gene. Phosphorylation of tau at normal level regulates its function to bind microtubules, while hyperphosphorylation can convert tau into an inhibitory molecule, which sequesters other microtubule-associated proteins from interacting with microtubules. Furthermore, hyperphosphorylation at certain sites, such as Ser199/Ser202/Thr205, can detach tau from microtubules and increase its propensity to aggregate, thus forming NFTs. Hypotheses suggest that the detachment of tau further contributes to the destabilization of microtubules, which breaks down axonal flow, eventually leading to severe neuronal death. Likewise, studies from autopsied AD brains suggest that the hyperphosphorylation level of tau in AD brain is three to fourfold higher than in normal brains (Yen et al., 2009). Interestingly, studies suggest the absence of tau in tau knockout animals has no
drastic effect on brain development and function (Dehmelt et al., 2005). In general, although correlations have been established between the sequence of neurofibrillary pathological progression and clinical symptoms propagation, other factors may be involved in propelling pathological cascades of AD.

1.2.3 Synaptic loss
The dramatic synapse loss in neocortex and hippocampal formation is another major neuropathological feature in the AD brain. Patients diagnosed with early Alzheimer’s Disease showed 44% synaptic reduction at the outer molecular layer of the dentate gyrus (Scheff et al., 2007). Various studies showed multiple synaptic proteins are downregulated in the brains of AD patients (Ingelsson et al., 2004). It’s now known that amyloid-beta mediates synaptic loss through NMDAR. The activation of NMDAR increases the Ca$^{2+}$ influx. This enhances the activation of a Ca$^{2+}$-activated phosphatase (CaN), which, in turn, alleviates the synaptic loss in AD brains (Wu et al., 2010). Hypothesis suggests that the overload of ER Ca$^{2+}$ may compensate for downregulation of synaptic function in AD (Popugaeva & Bezprozvanny, 2014), which indicates a potential therapeutic target for AD. Furthermore, substantial evidence demonstrates a strong correlation between the severity of synaptic loss and cognitive decline in AD (Scheff & Price, 2003).

1.2.4 Genetic Landscape
It is believed that both AD and AD progression are affected by inherited and environmental factors (Gatz et al., 2006). Aging (Wyss-Coray, 2016) is the most important known risk factor for AD. In addition, low educational level (Whalley et al.,
2000), traumatic brain injuries (Gottlieb, 2000), excessive alcohol consumption, and drug abuse (Gray et al., 2016) have been suggested to enhance the AD susceptibility. Around 25% of all AD cases are familial forms of Alzheimer's Disease (FAD). Early-onset FAD (EOFAD) is diagnosed in the families where members exhibit consistent early onset before 55 years old (Atwood&Bowen, 2015). Mutations in three genes (PSEN1, PSEN2, and APP) have been recognized to contribute to EOFAD, where patients start to develop AD symptoms younger than age 65.

Located on chromosome 21 in human, the APP gene encodes for amyloid precursor protein, the mutation of which strongly associates with AD. Studies showed 10% - 15% of EOFAD can be attributed to the mutation of APP. Although the function of APP is not clear, APP knockout animals exhibit reduced locomotor activity, memory and learning deficits (Wang et., 2010). 32 different mutations of APP are suggested to cause autosomal dominant EOFAD. In addition, locus duplication of APP has also been strongly associated with EOFAD (Rovelet-Lecrux et al., 2006). Mutations seem to affect N- or C-truncated peptides and increase the production of Aβ42. Interestingly, the coding mutation A673T in the APP gene was shown to protect against AD and related cognitive decline (Jonsson et al., 2012).

PSEN1 and PSEN2 are homologous genes, both encoding proteins critical for γ-Secretase function. 179 PSEN1 and 14 PSEN2 mutations lead to autosomal dominant EOFAD. Mutations of PSEN1 lead to the most aggressive type of FAD, which can occur even 25 to 30 years earlier (Wang et al., 2015).
APOE is a class of apolipoprotein which has three isoforms: APOE2, APOE3 and APOE4. Composed of 299 amino acids, APOE plays pivotal roles in lipid transportation, which is essential for axonal growth and synaptic formation. The malfunction of lipid metabolism has been widely shown to relate to AD pathogenesis. Moreover, the clearance of Aβ 40 from the CNS is inhibited in an allele-specific manner in knockout mice (APOE4>APOE3), suggesting the possibility that APOE4 might mediate the reduced Aβ clearance in AD brains. In fact, APOE ε4, located at human chromosome 19, has been demonstrated as a strong risk factor for sporadic AD. Individuals carrying one copy of APOE ε4 have a three-fold higher risk to develop AD, and the number increases to 15 fold for homozygous carriers (Ungar et al., 2014).

1.3 Neuroinflammation

Inflammation is a biological process of tissue reacting to injuries. Acute inflammation is defensive, while chronic neuroinflammation is detrimental and often associated with neurodegenerative disorders. Emerging evidence suggests the critical role of neuroinflammation in AD pathogenesis. First, genes encode for immune receptors, including TREM2 and CD33, were found to be upregulated in AD (Wang et al., 2015). Second, chronic activation of microglia and astrocytes in AD brain leads to the overexpression of main inflammatory mediators, which is a central feature of neuroinflammation. Meanwhile, elevation of immunological mediators, such as chemokines and cytokines, can be detected in the cerebrospinal fluid of AD patients. Next, epidemiologic evidence suggests that prolonged treatment with anti-inflammatory drugs reduces the risk of AD development (Sastre
et al. 2003). Specifying reactive cells and their AD-related signaling pathways can bring deeper insights into this important disease-promoting mechanism.

1.3.1 Microglia

Microglia, a macrophage in the CNS (Ginhoux et al., 2013), play a key role in defending against antigens (Rock et al., 2004), remodeling synapses (Aarum et al., 2003), and maintaining neuronal circuits plasticity (Aguzzi et al., 2013). Derived from myeloid cells, microglia distribute uniformly across the brain regions (Lawson et al., 1990). When presented with pathological antigens, activated microglia undergo dramatic morphological changes and migrate to the site of lesion, triggering innate immune responses. Meanwhile, surface molecules of microglia such as chemokine receptors and the major histocompatibility complex (MHC) are upregulated to facilitate the immune signaling pathways. In AD, microglia and astrocytes are found clustering at Aβ plaques (Wang et al., 2015), suggesting that Aβ plaques seem to attract and recruit microglia and microglial precursor cells (Review by Streit et al., 2004). The activation of microglia can be triggered by a variety of factors including pro-inflammatory cytokines, lipopolysaccharide, and accumulated neuronal debris. Therefore, microglia activation can be both neuroprotective and detrimental. On one hand, microglia removes apoptotic debris to reduce further stimulation to the neuroinflammatory pathways. Specifically, microglial phagocytosis plays an essential role in Aβ clearance (Li et al., 2014). On the other hand, endogenous microglia are demonstrated to be inefficient in removing Aβ, which can advance the neuroinflammatory responses. Overactivation
of microglia leads to overproduction of pro-inflammatory cytokines, which further enhance microglia activation. In addition, inflammatory cytokines, in turn, inhibit microglial phagocytosis by downregulating Aβ phagocytosis receptors (Yamamoto et al., 2008). Thus, microglial activation triggers a vicious cycle of ongoing neuroinflammatory responses.

![Diagram of microglial activation cycle](image)

**Fig 2. Reactive microgliosis and progressive neurotoxicity.**

*(Image from Block et al., 2007)*

1.3.2 Astrocyte

Astrocytes are CNS-resident cells, critical in nutrients provision to nervous tissue and repair mechanisms in the CNS. In the normal human brain, an estimated 30% to 50% neural cells can be represented by astrocytes. Functionally, astrocytes involve in a series of critical CNS processes, which are dysregulated in AD. In AD brains, astrocytes undergo morphological and functional alterations, including changes in gene expression, protein composition and cell activity. Furthermore, evidence
suggests that astrocytes activation is well correlated with the cognitive decline in AD (Perez-Nievas et al., 2013). Similar to microglia, astrocytes are found clustering around amyloid plaques in both human and animal models (Meyer-Luehmann et al., 2008). Astrocytes undergo Aβ plaque related astrogliosis followed by atrophy in AD, leading to significant decrease in glutamate transmission. Researchers found R75NTR play crucial role in astrocyte and neuronal communication. By applying TGF-beta to astrocytes, a small fragment of protein R75NTR was found released, which was able to bind nucleoporins. By binding to nuclear pore complex, R75NTR enhanced molecule flow of Smad2, a molecule triggering astrocytes activation (Schachtrup et al., 2015), indicating the significant role of astrocytes in neuronal communication. Moreover, astrocytes release molecules involve in antioxidant defense and trophic support when communicating with neighboring cells. In different cellular environments, astrocytes can release both pro- and anti-inflammatory cytokines.

Fig3. Changes of microglial and astroglial in AD brain.
(Images obtained from Heneka et al., 2015)

1.3.3 Oligodendrocyte

Oligodendrocytes, a type of highly specialized and markedly heterogeneous neuroglia cell, provide myelin sheath for axon insulation in the CNS. In MS, oligodendrocytes and myelin are known as targets of immune reactions. In AD,
myelin disruption is hypothesized to associate with pathophysiological changes of oligodendrocytes. In the gray matter of AD patients, demyelination is found in both AD mice model and AD patients (Mitew et al., 2010). Researchers hypothesized that under oxidative stress oligodendrocytes are susceptible due to their low glutathione content and high iron concentration, making them easily impaired under oxygen radicals. Other evidence suggests that oligodendrocytes express complement components, the increasing expression of which is suggested to contribute to neuroinflammation.

1.4. Single Cell RNA Sequencing

Within almost all the organisms, including humans, every cell contains the same set of genes. But various cells have distinct expression patterns, giving rise to different functional behaviors and properties of different cell types, in both disease and healthy conditions. Therefore, comparing and understanding transcriptomes gains us profound insight into the functional components in the genome, constituent components of certain cell populations, and how differential gene expression leads to or reflects pathology.

In conventional methods of gene expression profiling, such as microarrays and next generation sequencing technologies, large numbers of cells are profiled simultaneously, and consequently only the average expression level of genes are manifested. These approaches not only require a large amount of RNA to start with, signals of interest are also likely to be masked by the averaging effect.

Revolutionary advances in RNA sequencing now enable us to profile transcriptomes of individual cells in an unbiased, efficient and high-resolution manner, thereby
revealing the cellular complexities and cell-specific functionalities. Single-cell RNA sequencing is a powerful technique used to measure the genome-wide expression of genes at single cell resolution (Lun et al., 2016).

Compared to conventional bulk-RNA sequencing, single-cell RNA sequencing has several advantages: First of all, single-cell RNA sequencing makes it possible to (1) examine cell-specific changes at the transcriptomic level, thus allowing us to identify cell diversity within a complex and large cell population; (2) to discover new cell types; (3) to examine transcriptomic changes in disease progression and during normal development at high resolution. Secondly, single-cell RNA sequencing allows us to profile rare cell populations, which are previously inaccessible via traditional profiling methods. In general, single cell RNA sequencing can achieve resolution that bulk RNA sequencing is unable to.

1.4.1 Single Cell RNA Sequencing workflow

In the past few years, various single-cell RNA sequencing (scRNA-seq) protocols were designed and but rationales behind are generally similar.
**Fig4. Single-cell RNA-sequencing workflow.**


Firstly, tissue is dissociated and individual cells are isolated. Despite the unprecedented progress in scRNA-seq, it still remains unattainable to sequence RNA directly. Therefore, it is a common approach to capture RNA by lysing the individual cells. Reverse transcription converts the captured RNA into first strand cDNA. And then PCR is conducted for cDNA amplification. An alternative approach to amplify the first strand cDNA is in-vitro transcription. Next, a cDNA library is constructed for sequencing (Kolodziejczyk et al., 2015).

For transcriptional heterogeneity analysis, it is essential to quantify gene expression variations across single cells. Principal Component Analysis (PCA), a linear transformation algorithm, which projects high-dimensionality vectors (data) onto lower dimensional space, is frequently used for dimensionality reduction while retaining most of the variations of the data (Hotelling et al., 1933). In this approach, dissimilar variables are kept far apart. T-distributed Stochastic Neighbor Embedding (tSNE), a non-linear dimensionality reduction technique, is also frequently used to reduce and visualize high-dimensional vectors, where similar variables are modeled in small pairwise distances (Hinton&Roweis, 2002).
II. Results

We performed single-nucleus RNA-sequencing on human post-mortem brain tissue of 12 AD patients and 12 aged-matched normal individuals, using the SMART-Seq2 protocol (Picelli et al., 2014). A total of 3135 adult human postmortem prefrontal cortex nuclei were sequenced. We generated over 2.5 billion total reads and an average of 790,000 reads per cell nuclei were obtained.

2.1 Cell Isolation and Library Preparation

Human post-mortem frozen tissue was first homogenized, and nuclei were purified using an OptiPrep density gradient (protocol adapted from Swiech et al., 2015 Nature Biotechnology). Purified nuclei were labeled with a nuclear stain (Hoechst 33342) before FACS sorting. Then, a single nucleus was sorted into each well of a 96-well plate via flow cytometry. A numerical boundary (gate) was set according to optical homogeneity and size of the cells to gate out potential non-nuclei debris (Fig.4). Single nuclei were selected based on the Hoechst signal intensity (Fig.4b).
**Fig4. Diagram of FACS sorting.**

a. Frozen human postmortem brain tissue; Single nuclei under microscope; FACS sorting.

b. Forward-scattered light FACS scatter plot.

c. Nuclei stained by DAPI. d. Side-scattered light FACS scatter plot.

After the final sequencing library has been constructed, the DNA fragment size distribution of library products was measured via Agilent High-Sensitivity DNA BioAnalyzer (figures provided by MIT BioMicro Center) to ensure the quality of the library. Among all the parameters, the concentration of the sample is an important criterion to proceed to the sequencing steps. Normally the DNA concentration ranges distribute from 2 to 100 ng/μl after Nextera library has been pooled and cleaned up.

**Fig7. Electropherogram showing the DNA fragment size distribution of a final sequencing library.**
The fragment size distribution of the final libraries was unimodal and well-defined. In the example, we obtained 17.5490 ng/μl and an average fragment size 401bp, which is within the ideal range for sequencing.

2.2 Six Distinct Cell Types Are Identified in Control Group

To examine intrinsic heterogeneity within the data set, several dimensionality reduction algorithms were implemented for the following purposes: (1) to reduce data noise (2) for feature extraction in a data-driven manner (3) for better visualization and unbiased conceptualization of the experimental results. Genes expressed in less than 50 cells were excluded in order to reduce the noises due to random transcript loss caused by technical variation (Ramskold et al., 2012). To control the potential confounding effect associated with pathology, only cells (1531 nuclei) from the control group were selected for cell type identification.

Firstly, Principal Component Analysis (PCA) was performed on the FPKM matrix of RNA-Seq data to investigate variations across cells. PCA, a multivariate projection algorithm, projects vectors onto PC axes along the direction where the maximal variation of the data is achieved. In general, PCA takes the dataset and re-organizes the data according to new variables to reveal the hidden structuredness. After PCA analysis, the number of principle components used as input for the t-distributed stochastic neighbor embedding was empirically determined based on the variance explained by each principal component. Eventually, T-Distributed Stochastic Neighboring Embedding (t-SNE) was implemented for data visualization. Principal component analysis and t-distributed stochastic neighbor embedding (t-SNE) were
performed using Seurat, an R package designed for the analysis of single-cell RNA-sequencing data. For clustering the cells, the FindCluster() function in Seurat was called.

Among 1531 nuclei from the control group, eight distinct clusters were identified (Fig.8). In general, the groups of cells determined by the tSNE algorithm matched the clusters determined by the FindCluster() function very well. However, the tSNE algorithm grouped cluster 0, cluster 1, cluster 2 and cluster 3 together, while the FindCluster() function indicated that these cells are part of distinct clusters. This discrepancy suggests the possibility that these four clusters are 4 subtypes of one cell type. However, later analysis failed to identify cell subtypes within cluster 1 to cluster 4.

Fig.8 Identification of six major classes of cells in the aged human brain. The tSNE plot shows a two-dimensional representation of global gene expression profile relationships among 1531 nuclei isolated from the brain of control subjects. Data points are colored by cluster identity as determined by the FindCluster() function of Seurat.
To identify marker genes of each cluster, the FindAllMarkers() function in the Seurat was implemented and the results were visualized in scatter plots generated via Matlab. Cell type specific gene signatures were found in each cluster.
Fig. 9 Marker genes define distinct cell types.

For instance, gene MBP, an oligodendrocyte-specific marker, was found highly expressed in cluster 0, 1, 3, 4. NRGN, a known to be neuron-specific gene, was found highly and exclusively expressed in cluster 4. To further confirm cell cluster identity, we compared our marker genes to marker gene lists identified in other published studies (Darmanis et al., 2015, Lake et al., 2016, Foldy et al., 2016, Zeisel et al., 2015, Poulin et al., 2016) and we observed an exclusive pattern of overlaps between our clusters and identified cell types form other groups.

<table>
<thead>
<tr>
<th></th>
<th>Cluster0</th>
<th>Cluster1</th>
<th>Cluster2</th>
<th>Cluster3</th>
<th>Cluster4</th>
<th>Cluster5</th>
<th>Cluster6</th>
<th>Cluster7</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPCs</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oligodendrocyte</td>
<td>4.25%</td>
<td>2.17%</td>
<td>1.32%</td>
<td>5.17%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Astrocyte</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.26%</td>
</tr>
<tr>
<td>Microglia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.13%</td>
</tr>
<tr>
<td>Neurons</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.43%</td>
<td>3.20%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Endothelia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 10 Table showing the overlap of marker genes of clusters 0 to 7 and marker genes of the major brain cell types as determined by Darmanis et al.
By overlapping marker genes from Darmanis et al. to our maker genes in every cluster, Oligodendrocyte maker genes from Darmanis et al., were found exclusively in Cluster0. 1, 2, 3 in our dataset. Astrocyte marker genes identified by Darmanis et al. also showed exclusively in cluster 7, which further confirms cluster identities. Following the same method, we identified cluster0, cluster1, cluster2, cluster3 to be oligodendrocyte clusters, cluster 4 and cluster 5 are neuron clusters, which can be further partitioned to be the excitatory neuron cluster and the inhibitory neuron cluster respectively. Cluster 6 is the microglia cluster and cluster 7 is the astrocyte cluster.

2.3 Two Pathology-associated Clusters Are Identified
To examine to potential impact of AD pathology, same cell clustering techniques were performed on the entire cell population (3135 nuclei) and a total of 11 distinct cell clusters were identified.
Fig.11 The tSNE plot shows a two-dimensional representation of global gene expression profile relationships among 2934 nuclei isolated from the brain of control subjects and subjects with amyloid pathology. Data points are color-coded according to cluster identity as determined by the FindCluster() function of Seurat.

To determine cell type in each cluster, we first determined marker genes for each cluster using the FindAllMarkers() function in Seurat and examined their expression patterns across all cells. Next, we examined the expression pattern of marker genes found in other studies in our dataset via Matlab (Fig.12). We found that each cluster was exclusively enriched for maker genes for one specific cell type.

Oligodendrocyte-specific markers (for example MBP, MOBP) were identified in five clusters (Cluster 0,1,2,3,4,8). Neurons were also detected in multiple clusters. Inhibitory neuronal markers (Gad1, Gad2, GABRA1) were exclusively detected in cluster 6. Cluster 7 was identified as excitatory neuron cluster, due to the detection of previously defined marker genes (CAMK2A,SATB2). CD74 were detected exclusively and highly expressed in cluster 9, which was identified as microglia cluster. Astrocyte-specific genes (GFAP,GFGR3,AQP4) were specially enriched in cluster 10.

OPCs  Oligodendrocyte
Fig.12 Figures show the expression patterns of marker genes identified in previous studies (Darmanis et al., 2015) in our dataset. Cell type specific marker genes were highly expressed in a cluster specific manner.
However, no cell-type specific marker genes exclusively characterized cluster 5. Marker genes extracted by the FindAllMarkers() Function were neither cell-type specific nor exclusively expressed in cluster 5. Furthermore, Clusters 0,1,2,3,8 and cluster 4 were grouped together by the tSNE solution and common marker genes were detected (MBP), but most oligodendrocytes marker genes showed variability in their expression level between oligodendrocytes subclusters (Clusters 0,1,2,3,4,8). In particular, we found that most oligodendrocytes marker genes were highly expressed in cluster 0,1,2,3,8, but the expression levels were relatively moderate in cluster 4, suggesting other factors might affect cellular phenotype.

By tracing the cell origins, we confirmed the unbiased contribution of cells from each individual to each cluster. Interestingly, cells from the AD pathological group were dominant in number in both cluster 5 (86%) and cluster 4 (82.1%). Therefore we concluded that cluster 5 and cluster 4 represent novel amyloid pathology-associated cell states that are almost exclusively observed in the brains of subjects with amyloid pathology and are very rare in the brain of control subjects.
Fig. 13 The tSNE plot as shown in Figure 11. Cells isolated from the brain of control subjects are shown in black and cells isolated from the brain of subjects with amyloid pathology are shown in red.

2.4 Differential expressed genes in pathology-associated clusters

To learn how the cells in the amyloid pathology-associated clusters 4 and 5 differ from cells in the healthy brain, we performed a differential gene expression analysis followed by Gene Ontology Enrichment analysis. For the differential gene expression analysis, we used the single-cell differential expression (SCDE) software package (Kharchenko et al. 2014), which has been specifically designed to determine gene expression differences between groups of cells based on single-cell RNA-sequencing data.

Both t-SNE results and the expression of marker genes demonstrated the association between cluster 4 to the oligodendrocytes cluster (cluster 0,1,2,3,8). In
addition, previous analysis suggested the presence of neuronal markers in cluster 5 (Fig.14). Therefore, we compared cluster 5 to the neuron clusters (cluster 6 and cluster 7); and cluster 4 to the oligodendrocyte clusters (cluster 0,1,2,3,8) respectively by performing the differential gene expression analysis.

![Fig.14 The Presence of neuronal markers in cluster 5.](image)

In the comparison of cluster 5 to cluster 7 (excitatory neuron cluster), we used an expression difference Z-score (corrected for multiple hypothesis testing using Holm procedure (cZ) of larger than 4 as a significance cutoff. We found a total of 1678 genes to be differentially expressed, with 825 genes upregulated and 853 genes downregulated. To learn what the functions of these genes are and what cellular processes might be affected by the changes in their expression level, we performed a Gene Ontology analysis using the Enrichr Suite (http://amp.pharm.mssm.edu/Enrichr/). We found many genes downregulated in cluster 5 (compared to cluster 7) demonstrated Ca2+ mediated synaptic functions (SYT1, CALM1, CAMK2A, VSNL1, MAP18), and roles in neurotransmission (MAO1B, SNAP25, CHN1). Also, many of the genes upregulated in cluster 5 (compared to cluster 7) have oligodendrocytes-related functions (QIK, MYRF, ASPACDH1, DSG1),
indicating that some cells of cluster 5 may be oligodendrocytes. Indeed, the tSNE scatter plot shows that a subset of the cells of cluster 5 locates in the large group of cells that contains the oligodendrocyte clusters. Moreover, genes associated with the term "Electron transport chain" (MT-ATP6, MT-ATP8, MT-CO2, MT-CO3, MT-ND2, MT-ND3, MT-ND4L, MT-ND5, MT-ND6) were also clearly upregulated in Cluster 5.

![Volcano plot showing SCDE results of Cluster 5 and Cluster 7](image)

**Fig.15 a. SCDE results of Cluster 5 and Cluster 7 shown in volcano plot.**

**b. GO analysis of downregulated genes associated Biological processes in cluster 5.**

The same SCDE and GO analysis were conducted on cluster 5 and cluster 6 with a significance cutoff cZ>3. As observed for cluster 7, the cells of the inhibitory neuron
cluster (cluster 6) expressed genes associated with synaptic functions at a higher level than the cells of cluster 5. Similarly, many upregulated genes (MBP, PLP1, ASPA, CLDN5, CLDN11) in cluster 5 demonstrated oligodendrocytes-associated functions such as myelination and axon ensheathment.

In summary, compared to cluster 6 and 7, the amyloid pathology associated cluster 5 clearly demonstrated a downregulation of genes associated with synaptic functions such as synaptic transmission and regulation of synaptic plasticity.

Fig.16 a. SCDE results of Cluster 5 and Cluster 6 shown in volcano plot.

b. GO analysis of downregulated genes associated Biological processes in cluster 5.
Cells were separated by gender for performing the SCDE analysis on cluster 0,1,2,3,8 versus cluster 4 (cZ>4). The reason for doing so was that the working memory of the computer used was not sufficient to perform the SCDE analysis of all the cells of the clusters 0,1,2,3,4, and 8. However, intriguingly, we found many genes that are upregulated in cluster 4 showed pronounced association with positive activation of innate immune response (TLR2, CD36, NOD2, IL1B2) and positive regulation of cytokine production, which are both known phenotypes in AD pathology. In addition, many downregulated genes were associated with oligodendrocytes functions, suggesting that the potential impairment of oligodendrocytes-related functions in the brains of amyloid pathology individuals.

For gender 1 (men), upregulated genes demonstrated innate immune responses associated functions, though not as pronounced as in gender 0 (women). Although this may be a real biological difference between men and women, it could also be due to the difference in the cell number. Lower cell number would almost certainly reduce the significance. To further test the source of variation, the differential expression analysis could be performed to compared men and women in only pathology cases.
b.
- regulation of growth of symbiont in host (GO:0044126)
- negative regulation of growth of symbiont involved in interaction with host (GO:00444146)
- negative regulation of growth of symbiont in host (GO:00444130)
- modulation of growth of symbiont involved in interaction with host (GO:00444144)
- positive regulation of myeloid leukocyte cytokine production involved in immune response (GO:0061081)
- positive regulation of cytokine production involved in immune response (GO:0002720)
- acrosome assembly (GO:0001675)
- regulation of macrophage cytokine production (GO:0010935)
- regulation of interferon-gamma biosynthetic process (GO:0045072)
- positive regulation of interleukin-12 production (GO:0032735)

c.

d.
Fig. 17 a. SCDE of Cluster 4 and Cluster 01238 in gender 0.

b. GO analysis of upregulated genes associated biological process in cluster 4 gender 0.

c. SCDE of Cluster 4 and Cluster 01238 in gender 1.

d. GO analysis of upregulated genes associated biological process in cluster 4 gender 1.

2.5 SCDE on Pathology Individuals and Control group

To further investigate how AD pathology may affect different cell types, we conducted SCDE and GO analysis comparing cells from individuals with amyloid pathology to cells from control subjects in each cell cluster. Likely due to the small number of cells, very few genes reached the threshold of cZ>3 in cluster 6, 7, 9, and 10. Therefore, a significance cutoff cZ>2 was used.

In cluster 9, because there are only 13 cells it could be that just by chance these transcripts were not detected. The low cZ score also indicates that we should not ignore this possibility. Downregulated genes seem to be associated with protein metabolic process and cellular transferase activity (PPP2C1, PDK1, MEN1, PSMD2).

In cluster 10, among the genes that are most significantly upregulated in the cells from the pathology group, we found some of the genes associated with the terms "Electron transport chain/Oxidative phosphorylation" that were also found to be
upregulated in Cluster 5. This is also true for both Cluster 6 (MT-ATP6, MT-ND3, MT-ND4L, MT-ND6) and cluster 7 (MT-ND3, MT-ND4L, MT-ATP6, MT-ATP8). This observation provides additional evidence, that these genes are really upregulated in neurons and other cell types in individuals with AD pathology.

Fig. 18 GO analysis of upregulated genes associated biological processes in cluster 10 pathology group

2.6 APOE

Independent of Single nuclei RNA Sequencing results, we found an over-representation of APOE genotype in pathology-associated clusters.
b.

APOE\(\varepsilon 4\) is well-known risk factor for late-onset sporadic Alzheimer’s Disease. To explore the impact of the APOE genotype and its potential effect on different gender, SCDE and GO analysis were performed to compare differential gene expression in
individuals carrying APOE ε4 allele and individuals who do not. This analysis was performed on oligodendrocytes (clusters 0,1,2,3,4,8) and neurons (clusters 5,6,7) respectively in a gender specific manner.

Using a significance cutoff of cZ>4, we found genes were significantly differentially expressed in neurons that contain at least one APOE4 allele. In neurons of gender 0 (women), upregulated genes in cells containing an APOE ε 4 allele were associated with the innate immune response pathway (PLCG2, SULT1b1, SLC13A1). In gender 1 (men), we only detected one upregulated gene in cells containing the APOE ε 4 allele (MTRNR2L12). However, in both gender, we found many common enriched terms associated with the downregulated genes are related to synaptic functions, which is consistent with previous observations.

In oligodendrocytes (cluster 0,1,2,3,4,8) of the female individuals carrying an APOE ε4 allele, many enriched terms related to the downregulated genes were associated with myelination and microtubule related functions.

Generally, in both neurons and oligodendrocytes the number of significantly changed genes seems to be much higher in gender 0 than in gender 1. To further illustrate the gender differences, we generated scatterplots by first selecting the genes that are significantly changed in gender 0/gender 1. And then we plotted for these genes the fold change in gender 1 versus the fold change in gender 0.

a.
Fig. 20  

b. Differential gene expression differences between gender in cluster 012348.  

(Images generated by Dr. Mathys).

These data indicate that the gene expression dysregulation in neurons and oligodendrocytes by the APOE4 allele is gender specific. More genes significantly changed in women and many genes only downregulated in women, implicating impact of the e4 allele on gene expression in female is more widespread.
III. Material and methods

3.1 Brain Tissue

Human postmortem brain tissue were donated by 24 individuals from Religious Orders Study and Rush Memory and Aging Project.

3.2 Tissue dissociation

Before tissue dissociation, five solutions were prepared.

1. Homogenization Buffer: 0.008ml 0.1mM Ethylenediaminetetraacetic acid pH 7.5 (EDTA pH 7.5; Sigma-Aldrich®, USA, E6635-500G), 0.2ml 5mM Calcium Chloride (CaCl₂, Sigma-Aldrich®, USA), 0.12ml 3mM Magnesium Acetate (Mg(Ac)₂; Sigma-Aldrich®, USA), 0.4ml 10mM Tris-hydrochloride(Tris- HCl pH 8.0;TEKNOVA, T0228), 0.4ml 10% NP40(IGEPAL®CA-630 ) and 0.002797ml 1mM b-mercaptoethanol (Sigma-Aldrich®, USA) in 26.0962ml H₂O.

2. Diluent: 0.012ml 0.6mM Ethylenediaminetetraacetic acid pH 7.5 (EDTA pH 7.5; Sigma-Aldrich®, USA, E6635-500G), 0.3ml 30mM Calcium Chloride (CaCl₂, Sigma-Aldrich®, USA), 0.18ml 18mM Magnesium Acetate (Mg(Ac)₂; Sigma-Aldrich®, USA), 0.6ml 60mM Tris-hydrochloride(Tris- HCl pH 8.0;TEKNOVA, T0228), and 0.004196ml 6mM b-mercaptoethanol (Sigma-Aldrich®, USA) in 8.903 ml H₂O.

3. Working solution: 6.67ml Diluent in 33.33ml OptiPrep(Sigma-Aldrich®, USA,D1556).

4. Gradient Solution:
(1) 29% OptiPrep Solution: 9.24ml Homogenization Buffer in 12.76ml Working Solution.

(2) 35% OptiPrep Solution: 4.8ml Homogenization Buffer in 11.2 ml Working Solution.

5. Lysis buffer: 7ml Buffer TCL (Qiagen, 1031576) and 70µl 2-mercaptoethanol (Sigma-Aldrich®, USA, M6250-10ML).

Human postmortem brain tissue was dissected in a petri dish (Eppendorf, Germany, 0030701011) on dry ice and 2ml Homogenization buffer was added. Next, tissue was transferred into a Wheaton Dounce Tissue Grinder(Sigma-Aldrich®, USA, D8938) and grinded with loose pestle for 25 times followed by 25 times with tight pestle. Mixture was transferred into a test tube (FALCON® A Corning Brand, USA). 3ml Homogenization Buffer was added for 5 minutes and incubation took place on ice. Then, 5ml Working Solution was added. After that, tissue homogenate was mixed and loaded onto 29% OptiPrep solution to in ultracentrifuge tube with 5ml 35% OptiPrep gradient. Cells were ultra-centrifuged at 9000rpm at 4°C for 20min. Finally, nuclei were collected from the layer between 29% OptiPrep Solution and 35% OptiPrep Solution. Nuclear Stain DAPI(4,6-diamidino-2-phenylindole) was added before FACS sorting. (Protocol adapted from Swiech et al., 2015 Nature Biotechnology)

3.3 Single Cell Capture via Fluorescence activated Cells Sorting

Flow Cytometry is a powerful tool of individual cell analysis for a heterogeneous cell population. With hydrodynamic focusing, prepared sample stream flowed with saline solution into the flow cell via fluidics system, to ensure only one cell pass through the
interrogation point at a time. At the interrogating point, single cells passed through the laser, where light scattering at all angles was collected and quantified by the detector. The magnitude of the light scattered forward is roughly proportional to the cell size. The collected light was converted into a voltage pulse to generate the forward scatter histogram, which presented the size distribution of cells in the sample. Light scattered to the side was collected via another detector. The generated scatterplot can be used to measure the morphological complexity of the cells. In this process, 200 cells were sorted into the A1 well and one single cell was distributed into each well of the PCR 96- well plate (Eppendorf, Germany, 951020401).


3.4 SMART-Seq2 single-cell RNA sequencing library preparation

In order to profile single cell and avoid averaging over population behavior, numerous single cell RNA sequencing protocol had been generated in the past decade. (Islam et al., 2010, Tang et al., 2010). SMART-Seq II is able to generate quantitative and reproducible
data from a small amount of starting material (RNA) at single cell resolution (Picelli et al., 2014, Trombetta et al., 2014)

3.4.1 RNA SPRI Beads Clean-up

All equipment and workbench was sprayed with RNaseZap® RNase Decontamination Solution (ThermoFisher Scientific, USA). RNA-SPRI beads (Agencourt® RNAClean® XP, Beckman Coulter, USA, A63987) which was preserved at 4°C in lab refrigerator, was taken out to room temperature for 30 minutes. Firstly, two reagents was prepared on ice:

(1) 3’Mix RT Mixture: 105.6 µl 3’RT primer (3’RT, 10µM, IDT®: 99µL of H2O and 11µL of 3’RT 100µM), Deoxynucleotide (dNTP) Solution Mix (New England BioLabs® Inc., N0447L), RDil (Takara®, 2313B: 99µL of H2O and 11µL of RDil 40U/µL), 94µL H2O and 11.52 µL ERCC Spike-ins.

(2) 5’RT Mixture: 79.2µL of H2O, 211.2µL of 5X Maxima RT buffer (ThermoFisher Scientific, USA, EP0752), 211.2µL of Betaine (Betain 5M, Sigma-Aldrich Corp., USA, B0300-1VL), 95.04 of Magnesium chloride solution (MgCl2: 100mM, Sigma-Aldrich Corp., USA, M1028: 90µL of H2O + 10µL of MgCl2), 105.6 µL of 5’RT primer (10µM, Exiqon®: 99µL of H2O + 11µL of 5’RT), 26.4µL of Recombination RNase Inhibitor (40U/µL, TaKaRa, 2313A) and 10.56µL of Maxima RNaseH- minus RT (200U/µL, ThermoFisher Scientific, USA, EP0752).

After being thoroughly vortexed, 11µL RNA-SPRI beads were distributed into each well of the 96-well plate which contained single cells from FACS sorting and pipetted up and down for 15 times. Then the covered plate was incubated at room temperature for ten
minutes before being incubated on the magnet (DynaMag-96 side-skirted magnet) for another five minutes. After that, supernatant were carefully removed while the plate stayed on the magnet. 100µL of freshly prepared 80% ethanol were distributed to each well for washing purpose and the plate was moved from one column of the magnet to another for 30 seconds. Then, ethanol was carefully removed from each well. The washing was repeated for three times before drying step where ethanol was removed completely from each well and the loosely-covered plate stayed on the magnet for 10 minutes.

3.4.2 Reverse Transcription
Dried beads in each well were then resuspended in 4µl of 3’ RT mixture and mixed up and down for 15 times before sealed by Microseal® ‘F’ foil Seals (Bio-Rad Laboratories, Inc., United Kingdom, MSF1001). Next, the plate was centrifuged at 800g for 1 minute at room temperature before the placed into qPCR machine for 3’RT primer annealing at 72°C for 3 min. Plate was then taken out and cooled on ice followed by adding 7 µl of 5’RT mix to each well of the plate. Finally, the plate was spun down at 800g for 1 minute at room temperature and sealed with Microseal before put into qPCR machine. Reverse transcription was conducted and the entire process took roughly 2.5 hours.

3.4.3 Whole Transcriptome Amplification
KAPPA-HIFI mixture was prepared right before reverse transcription was completed.

KAPPA HiFi + ISPCR primer Master Mix: 105.6µL of H₂O, 52.8µL of ISPCR Primer (10µM, IDT®) and 1320µL of KAPA HiFi HotStart ReadyMix (KapaBiosystems, USA, KK2601).
After reverse transcription, plate was taken out and cooled down in ice for 5 minutes. Then 14µL KAPPA-HIFI mixture was distributed into each well and pipetted up and down for 7 times. Next, the sealed plate was centrifuged for 1 minute at 800g at room temperature before placed into the qPCR machine for whole transcriptome amplification. Whole transcriptome amplification took around 2.7 hours.

3.4.4 DNA SPRI Beads Clean-up

After being thoroughly vortexed, DNA-SPRI beads (Agencourt® AMPure® XP, Beckman Coulter, USA, A63881) were preserved at 4°C in lab refrigerator, was taken out to room temperature for 30 minutes. 20µL DNA-SPRI beads were distributed into each well of the 96-well plate which contained single cells from RT-WTA and pipetted up and down for 15 times. Then the covered plate was incubated at room temperature for five minutes before being incubated on the magnet (DynaMag-96 side-skirted magnet) for another five minutes. After that, supernatant were carefully removed while the plate stayed on the magnet. 100µL of freshly prepared 80% ethanol were distributed to each well for washing purpose and the plate was moved from one column of the magnet to another for 30 seconds. Then, ethanol was carefully removed from each well. The washing was repeated for three times before drying step where ethanol was removed completely from each well and the loosely-covered plate stayed on the magnet for 10 minutes. Once the beads were dried, 20µl of 1× TE buffer (TE Buffer: 10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0, TEKNOVA, T0228) were distributed to each well and mixed up and down for 20 times.
3.4.5 Construct Nextera XT Sequencing Library Using Nextera® XT Library Prep Kit. Followed by DNA beads clean-up step, Nextera XT sequencing library was constructed using Nextera® XT Library Prep Kit (Illumina, Inc., USA, 15032354). In order to pool cells from one plate together without losing cell identity, each single-cell library needed to be individually barcoded with distinct index primers. Therefore, we generated 4 combinations of primers, making for distinct combined dual-index primer plates: SET A, SET B, SET C, SET D. 2.5μl Tagment DNA Buffer (TD Buffer; Illumina, Inc., USA, 15027866), 1.25μl WTA sample and 1.25μl Amplicon Tagment Mix (ATM; Illumina, Inc., USA, 15031560) were added to each well of a fresh 96-well plate. Then the mixture in each well were pipetted up and down for 5 times and centrifuged at 280g at room temperature for one minute. Next, tagmentation step was conducted in qPCR machine for 10 minutes at 55°C. Then plate was taken out and cooled down in ice for 2 minutes. After that 1.25 μl Neutralize Tagment Buffer (NT Buffer; Illumina, Inc., USA, 15031559) was added into each well and the plate was incubated for 5 minutes at room temperature to terminate tagmentation. Finally, 2.5μl index primer followed by 3.75μl Nextera PCR Master Mix (NPM; Illumina, Inc., USA, 15027920) were transferred into each well. Mixture were pipetted up and down for 5 times and centrifuged at 280g for 1 min at room temperature. Amplification was carried out in qPCR machine for 3 minutes at 72°C, 30 sec at 95°C, 12 cycles at 10 sec at 95°C, 30 sec at 55°C, 60 sec at 72°C, 5 min at 72°C before cooled down to 4°C.

3.4.6 DNA SPRI Clean-up and Pooling

DNA SPRI Beads(Agencourt® AMPure® XP, Beckman Coulter, USA, A63881) were obtained and placed at room temperature for 30 min. At first, plate from Nextera library
construction was first centrifuged for 1 min at 800xg at room temperature. Then all the content of each well of one sample were pooled together into a 1.5 ml Eppendorf tube. 0.9 volumes DNA SPRI beads were added into the tube and mixed by pipetting up and down for 20 times. After that, tube was incubated for 5 minutes before supernatant was carefully removed.

For washing step, 1 ml of 80% Ethanol were added to the tube and each tube were rotated for 15 times on the magnet before ethanol was aspirated. Washing step was performed for 3 times and the covered-tube was dried for 15 minutes. 30 µl TE buffer (TEKNOVA, T0228) were added to for DNA elution and the solution were pipetted for 100 times. Supernatant was transferred to a new, labeled 1.5 ml tube. Finally, the final constructed library was sent to MIT BioMicro Center for the sequencing.
3.5 Data Analysis

3.5.1 Tophat

Tophat aligns reads obtained from RNA-sequencing to reference genomes. Then alignments were processed through Cufflinks, where the abundance of transcripts was represented in FPKM (Reads Per Kilobase of exon per Million reads mapped).
3.5.2 Principal Component Analysis (PCA)

PCA is a dimension reduction algorithm, which calculates the most accurate linear construct to represent original data matrix based on newly-calculates variables (principal component). Objects with similar properties are assigned with new coordinates and automatically group together according to the similarities, thus making PCA a powerful tool for unbiased feature selection. It projects high-dimensionality vectors (data) onto lower dimensional space, while retaining most of the variations (Hotelling et al., 1933). Among the PCs, PC1 is the first constructed variable that represents for the greatest variation and PC2 has the second most variation. For analyzing single-nuclei RNA sequencing data in this study, PCA was performed on Seurat, an R package developed by Rahul Satija.

Normalized gene expression levels represented by fragments per kilobase per million mapped reads (FPKM) matrix was used as input. According to the calculation of the algorithm, 9 PCs were retained.

3.5.3 t-Distributed Stochastic Neighbor Embedding (t-SNE)

T-distributed Stochastic Neighbor Embedding (tSNE), a non-linear dimensionality reduction technique, where similar variables are modeled in small pairwise distance (Hinton&Roweis, 2002). Unlike PCA, euclidean distances between objects in high dimensionality are preserved to some extent when t-SNE calculates the similarities, thus making it powerful in visualizing high-dimensional data as well as feature selection. In this experiment, t-SNE was performed on Seurat.
3.5.4 Elbow Plot

Elbow Plot is a method, which visually demonstrates the variation of PCs. It was used in PCA for determining the number of PCs that could be retained.

3.5.5 Single Cell Differential Expression (SCDE)

Single Cell Differential Expression (SCDE) is a software package (Kharchenko et al. 2014) for analyzing Single Cell RNA-seq data. Normalized expression count matrix was used as an input and statistical scores, which quantify differential gene expression between groups. In the output, cZ scores represents distances between the value and the mean in the form of standard deviation numbers.

3.5.6 Gene Ontology Enrichment Analysis (GO Analysis)

Gene Ontology Enrichment Analysis was performed on Enrichr Suite. Briefly, in the Gene Ontology classification system genes are assigned to predefined groups (or terms) based on their function. A gene list of interest (for example a list of upregulated genes) can then be compared to all gene ontology groups and a statistical test is used to identify groups containing sets of genes that are overrepresented in the list of genes of interest.
IV. Discussion

Human brain is a complex system built from broad classes of functionally specialized cell types (Zeisel et al. 2015). We performed single nuclei RNA sequencing on human postmortem brain tissue of AD patients and gained unprecedented insight into the transcriptional dynamics of distinct cell types in response to AD pathology. For the first time, we investigated Alzheimer’s disease pathology at single-nuclei resolution and found altered gene expression profile in AD brain at single nuclei level. In this study, a total number of 2.5 billion reads were generated and over 60000 genes were defined for every cell type. We discovered novel and pathology-associated cell states and the identified the causality between APOE4 genotype and formation of novel cell state, due to the fact that genotype is determined prior to the formation of the cell states. Moreover, APOE4 genotype seems to cause dysregulation of gene expression in a gender specific manner. Thus an interesting interaction seems to exist between the female gender and the APOE genotype that causes a widespread dysregulation of gene expression in oligodendrocytes and neurons. Additionally, our GO analyses, in general, fit what had been reported previously. For example, we found downregulation of synaptic genes (Berchtold et al., 2014) and an upregulation of expression of genes involved in immune response, though only observed in Oligodendrocytes of female individuals.

Experimental results also demonstrated evidences that gene expression is affected in all major brain cell types in AD pathology (oligodendrocytes, neurons, microglia, astrocytes). However, this conclusion is less certain for microglia and astrocytes due
to the low number of cells in each cell cluster. Gene expression changes observed in these clusters also showed not reliable values of significance. In addition, no strong evidences suggested that observed gene expression changes correlate with amyloid pathology. We do not know whether amyloid pathology causes the changes in gene expression or whether the changes in gene expression result in amyloid pathology. It is also possible that the arrow goes in both directions. Moreover, we unexpectedly detected a large number of Oligodendrocytes, yet not being able to define Oligodendrocytes sub-types. However, previous work (Darmanis et al., 2015) only identified 38 Oligodendrocytes out of a total of 466 cells. One possibility may be the different relative contribution of each cell type to the cellular composition of the prefrontal cortex, which is more dominantly composed of white matter than tissue used in Darmanis et al. This discrepancy may also result from the initial sorting strategy, which favored the detection of Oligodendrocytes over neurons (which have larger nuclei). Furthermore, it is possible that the mechanical homogenization of the brain tissue introduced a bias in the cellular composition because the nuclei of some cell types may be more fragile than the nuclei of other cell types. Importantly, the number of analyzed individuals is relatively small when it comes to comparison between men and women or between individuals with APOE genotype and the control individuals. This study revealed hundreds if not thousands of genes that are dysregulated in various cell types of individuals with amyloid pathology. However, based on the RNA-sequencing data we cannot know whether these changes in gene expression have a causal effect on cell function. Future experiments could test the functional
consequence of interfering with the expression of genes identified to be
dysregulated in AD patients. This could be achieved by knocking out a gene of
interest in a cell type specific manner in an animal model. Alternatively, one could
use the CRISPR/Cas9 technology to generate an induced pluripotent stem cell line
lacking a gene of interest. The cell line could then be differentiated into the cell type
of interest and certain functional assays could be performed in vitro. The animal
model and the iPS cell line could then be used as Alzheimer's disease model systems
to identify and test strategies to reverse the cellular defects. Especially the iPS cell
line could be used for high throughput compound screening to identify molecules
that reverse the cellular defect. These experiments might uncover novel therapeutic
strategies for Alzheimer's disease. Additionally, since we detected only very few
astrocytes and microglia, one future experiment would be to try to label these nuclei
with an antibody recognizing a cell type specific marker and specifically sort only
microglia or astrocyte nuclei.
V. Supplementary Figure

a.

Cluster 0

Cluster 1

Cluster 2

Cluster 3

Cluster 4

Cluster 5

Cluster 6

Cluster 7
SFig1. Marker genes (obtained from FindCluster() ) define distinct cell types in the entire cell population.
SfIg.2 SCDE results of cluster 5 versus cluster 7 visualized via Matlab.
**Fig. 3** SCDE results within cluster 7 visualized via Volcano Plot.

**Fig. 4** SCDE results within cluster 10 visualized via Volcano Plot.
Sfig. 5 SCDE results within cluster 9 visualized via Volcano Plot.

Sfig. 6 SCDE results within cluster 6 visualized via Volcano Plot.
VI. References


34. Ziegenhain, Christoph et al.” Comparative Analysis of Single-Cell RNA Sequencing Methods”


