

RESEARCH ARTICLE

Open Access

# Sex dependent glial-specific changes in the chromatin accessibility landscape in late-onset Alzheimer's disease brains



Julio Barrera<sup>1,2†</sup>, Lingyun Song<sup>2†</sup>, Julia E. Gamache<sup>1,2</sup>, Melanie E. Garrett<sup>3</sup>, Alexias Safi<sup>2</sup>, Young Yun<sup>1,2</sup>, Ivana Premasinghe<sup>1,2</sup>, Daniel Sprague<sup>1,2</sup>, Danielle Chipman<sup>1,2</sup>, Jeffrey Li<sup>2</sup>, Hélène Fradin<sup>2</sup>, Karen Soldano<sup>3</sup>, Raluca Gordân<sup>2,4,5</sup>, Allison E. Ashley-Koch<sup>3,6\*</sup>, Gregory E. Crawford<sup>2,7,8\*</sup> and Ornit Chiba-Falek<sup>1,2\*</sup>

## Abstract

**Background:** In the post-GWAS era, there is an unmet need to decode the underpinning genetic etiologies of late-onset Alzheimer's disease (LOAD) and translate the associations to causation.

**Methods:** We conducted ATAC-seq profiling using NeuN sorted-nuclei from 40 frozen brain tissues to determine LOAD-specific changes in chromatin accessibility landscape in a cell-type specific manner.

**Results:** We identified 211 LOAD-specific differential chromatin accessibility sites in neuronal-nuclei, four of which overlapped with LOAD-GWAS regions ( $\pm 100$  kb of SNP). While the non-neuronal nuclei did not show LOAD-specific differences, stratification by sex identified 842 LOAD-specific chromatin accessibility sites in females. Seven of these sex-dependent sites in the non-neuronal samples overlapped LOAD-GWAS regions including APOE. LOAD loci were functionally validated using single-nuclei RNA-seq datasets.

**Conclusions:** Using brain sorted-nuclei enabled the identification of sex-dependent cell type-specific LOAD alterations in chromatin structure. These findings enhance the interpretation of LOAD-GWAS discoveries, provide potential pathomechanisms, and suggest novel LOAD-loci.

**Keywords:** ATAC-seq, Chromatin accessibility, Nuclei sorting, Late-onset Alzheimer's disease, snRNA-seq, Gene dysregulation

\* Correspondence: [alison.ashleykoch@duke.edu](mailto:alison.ashleykoch@duke.edu); [greg.crawford@duke.edu](mailto:greg.crawford@duke.edu); [o.chiba-falek@duke.edu](mailto:o.chiba-falek@duke.edu)

†Julio Barrera and Lingyun Song contributed equally to this work.

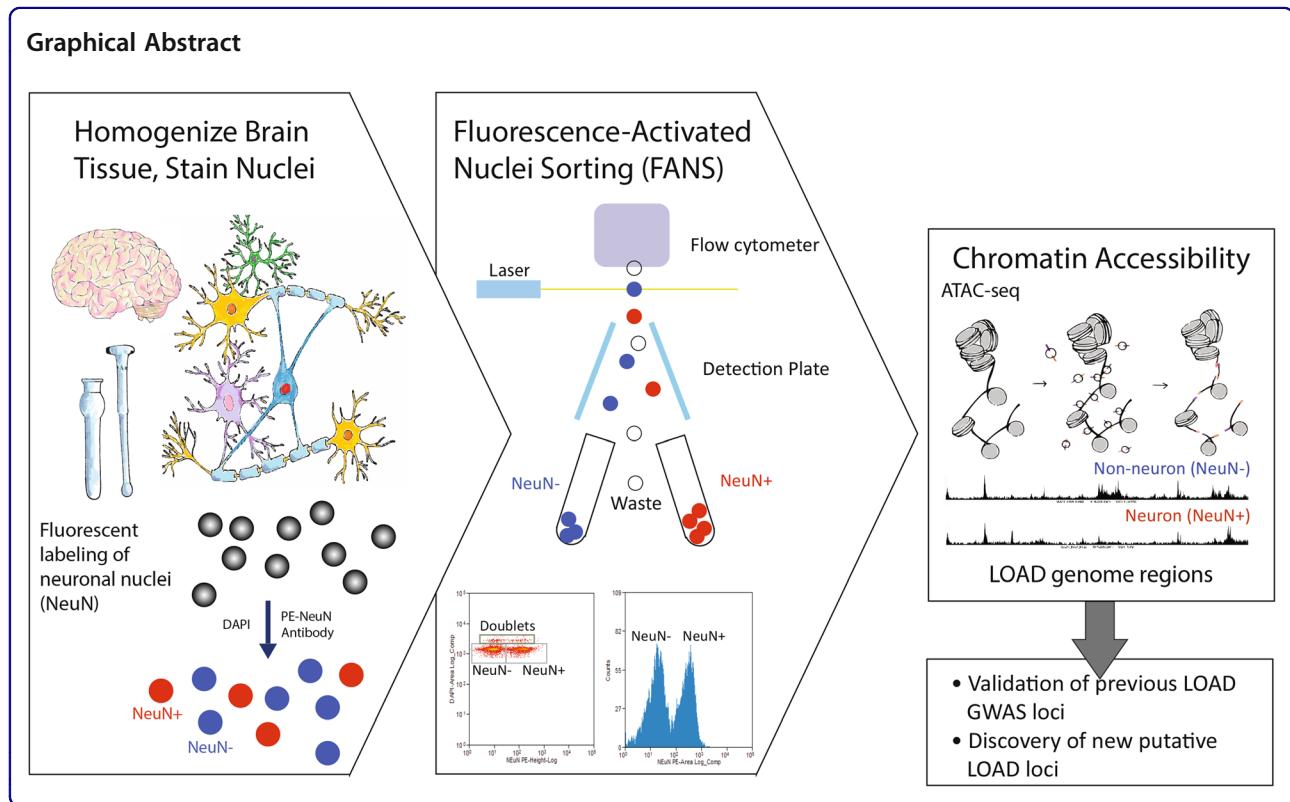
<sup>3</sup>Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC 27701, USA

<sup>2</sup>Center for Genomic and Computational Biology, Duke University Medical Center, Durham, NC 27708, USA

<sup>1</sup>Department of Neurology, Division of Translational Brain Sciences, Duke University Medical Center, DUMC, Box 2900, Durham, NC 27710, USA  
Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



## Background

Large multi-center genome-wide association studies (GWAS) have identified associations between numerous genomic loci and late-onset Alzheimer's disease (LOAD) [1–6]. The most recent GWAS meta-analysis reported a total of 25 LOAD-GWAS regions [7]. However, the precise disease-responsible genes, the specific causal genetic variants, and the molecular mechanisms of action that mediate their pathogenic effects are yet to be explained. Most LOAD-GWAS SNPs are in noncoding (intergenic and intragenic) genomic regions, and thus may have a gene regulatory function. Supporting this hypothesis, differential gene expression in brain tissues have been described between LOAD and healthy controls [8–10]. Moreover, several expression quantitative trait loci (eQTL) studies in brain tissues from cognitively normal [11] and LOAD [12–15] individuals reported overlaps with LOAD-GWAS loci. Collectively, these observations suggest that dysregulation of gene expression plays an important role in LOAD pathogenesis [16]. Noteworthy, integration of findings from LOAD epigenome wide association studies (EWAS) and GWAS also identified a number of shared loci [17–24], providing further support for the role of gene dysregulation in LOAD pathogenesis.

To date, most brain expression, DNA-methylation, and chromatin studies have used brain tissue homogenates that represent multiple cell-types, i.e. various types

of neurons and types of glia cells. The heterogeneity of brain tissue makes it difficult to determine the specific brain cell-type responsible for the changes in gene expression and in epigenome landscape. The mixture of cell-types could potentially mask signals corresponding to a particular cell-type, especially if the causal cell types comprise a small fraction of the entire sample. An additional shortcoming of studying bulk brain tissues is the bias related to sample-to-sample differences in the cell-type composition of the tissue. The obstacle of variability in the neuron:glia proportions across samples is even more pronounced when analyzing disease-affected brain tissues that underwent the neurodegeneration processes of neuronal loss and gliosis. Single cell experimental approaches can circumvent these limitations; however, these methods are costly for studying larger sample sizes and have been underutilized in the field of LOAD genetics. Frozen tissues pose additional technical challenges as it is difficult to isolate intact cell bodies. Fluorescence Activated Nuclei Sorting (FANS) was developed to extract, purify and sort nuclei (vs. cells) from frozen brain tissues [25] using nuclei neuronal markers, such as NeuN, to greatly reduce cellular heterogeneity found in bulk tissues, and allow characterization of neuronal (NeuN+) and non-neuronal (NeuN-) cell populations. Recently, two studies used FANS to analyze LOAD-specific changes in DNA-methylation on a whole-

genome level [26] and across the *APOE* locus [27]. Furthermore, two new studies used single-cell (sc)RNA-seq from cortex of LOAD patients. The first found that the strongest LOAD-associated changes appeared early in pathological progression and were highly cell-type specific [28], and the second identified LOAD-associated gene dysregulation in specific cell subpopulations, particularly for *APOE* and transcription factors [29]. These results further highlight the importance of cell-type specific analysis of human brain tissues to understanding of the molecular underpinnings and cellular basis of LOAD.

In this study, we performed NeuN-FANS from 40 archived frozen human brain samples (19 LOAD, 21 normal), and used the assay for transposase-accessible chromatin using sequencing (ATAC-seq) to characterize the chromatin accessibility landscape in neuronal and non-neuronal nuclei. We identified over 170,000 chromatin accessibility differences between neuronal and non-neuronal nuclei. We also report LOAD-specific differences in chromatin accessibility in both neurons and non-neurons. Interestingly, while the neuronal changes appeared to be independent of sex, in the non-neuronal cells LOAD differences in chromatin accessibility were detected only in females. LOAD-specific differences in chromatin accessibility significantly overlap with known LOAD GWAS regions, and also point to new candidate LOAD loci. These results provide new insights into the mechanistic and sex-specific pathogenesis of this disorder.

## Methods

### Human brain tissue samples

Following quality-control filtering, the final dataset was generated using fresh-frozen temporal cortex from neurologically healthy controls ( $n = 21$ ), mild LOAD ( $n = 16$ ), and severe LOAD ( $n = 3$ ) patients. These samples were obtained from the Kathleen Price Bryan Brain Bank (KPBBB) at Duke University, and the demographics for this cohort are included in Table 1 and detailed in Supplementary Table 2. Clinical diagnosis of LOAD was pathologically confirmed using Braak staging (AT8 immunostaining) and amyloid deposition assessment

(4G8 immunostaining) for all LOAD samples. Braak staging was used to define mild (stages IV and below) vs. severe (stages V and VI) LOAD. All donors are Caucasians with APOE 33; post mortem interval (PMI) averaged 7.15 h (standard error of the mean (SEM) 0.81). The archived frozen tissues have high-quality DNA as required for genomic analyses, and RNA (RIN  $\geq 8$ ) suitable for quantitative RNA analyses. The project was approved by the Duke Institutional Review Board (IRB). The methods described were carried out in accordance with the relevant guidelines and regulations. Tissue samples were processed in random pairs of one normal and one LOAD patient. Tissue homogenization, nuclei extraction, FANS, and fragmentation were performed on each pair on the same day. Library preparation and sequencing was performed blinded to age, sex, and pathology. Out of this group of samples four female donors were analyzed by snRNA-seq, two of which were neurologically healthy controls and two age matched mild LOAD patients (Braak Stage III). These four samples were ages 79–90 with PMI averaged 8.02 h (SEM = 1.99) (Supplementary Table 2, marked “\*”). Tissue samples were processed for snRNA-seq on the same day and in the same 10X Genomics microfluidics chip.

### Tissue dissociation and nuclei extraction

Methods were performed according to established protocols [25, 30, 31] with some modifications. Briefly, 50 mg of frozen temporal cortex (gray matter) was thawed for 10 min on ice in lysis buffer (0.32 M Sucrose, 5 mM CaCl<sub>2</sub>, 3 mM magnesium acetate, 0.1 mM EDTA, 10 mM Tris-HCl pH 8, 1 mM DTT, 0.1% Triton X-100). The tissue was gently dissociated and homogenized in a 7 ml dounce tissue homogenizer (Corning) with approximately 25 strokes of pestle A in 45 s, then filtered through a 100  $\mu$ m cell strainer. The filtered lysate was transferred to a 14  $\times$  89 mm polypropylene ultracentrifuge tube, carefully underlaid with sucrose solution (1.8 M sucrose, 3 mM magnesium acetate, 1 mM DTT, 10 mM Tris-HCl, pH 8) and subjected to ultracentrifugation at 107,000 RCF for approximately 30 min at 4 °C. Supernatant and the debris interphase were carefully

**Table 1** Demographic description of study cohort. Abbreviations: mild Alzheimer’s disease (mAD), severe Alzheimer’s disease (sAD)

	Pathology	Individuals	Males (%)	Age (y) $\pm$ SD	PMI (h) $\pm$ SD	NeuN+ (% $\pm$ SD)
40 Paired (neuronal and non-neuronal)	Normal	21	11 (52.4)	81.10 $\pm$ 9.70	7.91 $\pm$ 5.54	35.27 $\pm$ 13.26
	mAD	16	10 (62.5)	79.75 $\pm$ 7.23	7.09 $\pm$ 5.58	38.59 $\pm$ 14.70
	sAD	3	1 (33.3)	76.67 $\pm$ 7.77	5.43 $\pm$ 1.65	19.16 $\pm$ 7.49
	Total	40	22 (55)	80.23 $\pm$ 8.49	7.41 $\pm$ 5.26	35.39 $\pm$ 14.16
9 Unpaired (non-neuronal only)	Normal	3	0	84.67 $\pm$ 2.52	13.03 $\pm$ 4.95	N/A
	mAD	6	0	76.50 $\pm$ 8.48	16.95 $\pm$ 16.09	N/A
	Total	9	0	79.22 $\pm$ 8.73	15.64 $\pm$ 14.64	N/A

aspirated, and 100 ul PBS ( $-Mg^{2+}$ ,  $-Ca^{2+}$ ) was added to the nuclei pellet. After a 5-min incubation on ice, nuclei were gently resuspended and transferred to a 1.5 ml polypropylene microcentrifuge tube for staining.

For snRNA-seq downstream experiment, the above protocol was modified to optimize sample preparation for single-nuclei sorting. Briefly, 50 mg frozen temporal cortex (gray matter) was thawed for 20 min on ice in lysis buffer and ultracentrifuged at 107,000 RCF for approximately 10 min at 4 °C. Supernatant and the debris interphase were carefully aspirated, and 500 ul wash and resuspension buffer (1X PBS, 1% BSA, 0.2 U/ml RNase Inhibitor) was added to the nuclei pellet. After a 5-min incubation on ice, nuclei were gently resuspended and centrifuged at 300 RCF for 5 min at 4 °C. The supernatant was again aspirated and 500 ul wash and resuspension buffer without RNase Inhibitor (1X PBS, 1% BSA) was added to the nuclei pellet. After a 1-min incubation on ice, the nuclei were gently resuspended and filtered through a 35 um strainer. Ten microliters nuclei were taken for quality assessment and counting prior to library preparation.

#### Immunostaining of nuclei

Nuclei were stained in 0.05% BSA, 1% normal goat serum, DAPI (1 µg/ml), and PE-conjugated anti-NeuN antibody (1:125, Millipore FCMB317PE) in PBS ( $-Mg^{2+}$ ,  $-Ca^{2+}$ ), in the dark for 30 min at 4 °C. A DAPI-only control was prepared to set gates for sorting. After staining, nuclei were filtered through a 40 um cell strainer into a polypropylene round-bottom 5 ml tube and sorted.

#### Immunofluorescence microscopy

After homogenization and sucrose gradient ultracentrifugation, a portion of the nuclei was counted, resuspended in 4% PFA, stained, plated on 12 mm coverslips at 10,000 nuclei per coverslip, incubated 20 min at room temperature, mounted, and imaged on a confocal microscope.

#### Fluorescence-activated nuclei sorting (FANS) of neuronal and non-neuronal nuclei

Sorting was performed using a MoFlo Astrios flow cytometer (Beckman Coulter) equipped with a 70 µm nozzle, operating at 35 psi. Standard gating procedures were used. Briefly, the first gate allowed separation of intact nuclei from debris. The second gate allowed us to identify individual nuclei, and exclude doublets and other aggregates. The third and fourth gates distinguish between PE+ and PE- nuclei and allowed us to sort and separate NeuN+ nuclei from NeuN- nuclei. Nuclei were sorted into 1 ml PBS ( $-Mg^{2+}$ ,  $-Ca^{2+}$ ) in a 2 ml polypropylene

tube pre-coated with 200 ul of 5% BSA and rotated at 20 rpm at 4°C.

#### Omni-ATAC on FANS nuclei

Approximately 100,000–700,000 sorted nuclei (Supplementary Table 2) were used for ATAC-seq library preparation as described in the Omni-ATAC protocol [32]. Libraries were quantified by Qubit, and size distribution was inspected by Bioanalyzer (Agilent Genomic DNA chip, Agilent Technologies). Barcoded ATAC-seq libraries were combined into pools of 6 libraries and sequenced on an Illumina HiSeq 4000 sequencer (50 bp, single read) at the Duke Sequencing and Genomic Technologies shared resource.

#### Omni-ATAC-seq on bulk tissue

In addition to performing ATAC-seq on NeuN+/- sorted nuclei, we also compared to ATAC-seq performed on total nuclei isolated from frozen tissue. Approximately 50,000 nuclei from 25 mg of pulverized frozen tissue were used for the transposition reaction applying the Omni-ATAC protocol as described previously [32].

#### Data processing pipeline

ATAC-seq libraries made from 83 glia samples and 80 neuron samples were sequenced on Hi-seq 4000. Raw fastq sequencing files were first processed through cutadapt (v 1.9.1) to remove adaptors and bases with quality scores < 30. Filtered reads were aligned to hg19 by Bowtie2 (v 2.1.0) using default parameters [33]. Bam files were sorted by samtools (v 0.1.18) and duplicates were removed by Picard MarkDuplicates. Sequences that overlapped ENCODE “blacklist” regions (<https://sites.google.com/site/anshulkundaje/projects/blacklists>) were removed, and narrow open chromatin peaks were called by Model-based Analysis of ChIP-seq (MACS v 2.1), with parameters --nomodel --shift -100 --ext 200, using a FDR cutoff of  $q < 0.01$  or 0.05 [34]. For visualization on the UCSC browser, BigWig files were generated using wigToBigWig (v 4).

#### Data quality control

We characterized the quality of all ATAC-seq datasets by the following metrics, based on ENCODE suggestions [35] <https://genome.ucsc.edu/ENCODE/analysis.html> (Supplementary Table 1): (1) total numbers of reads, (2) numbers of reads trimmed by cutadapt, (3) total bases entering cutadapt, (4) quality-trimmed bases by cutadapt, (5) total numbers of reads processed by Bowtie2, (6) uniquely mappable reads, (7) percentages of alignment, (8) read aligned to the mitochondrial chromosome, (9) peak calls at FDR  $q < 0.05$ , (10) peak calls normalized by sequencing depth, (11) GC content of

sequences, (12) Non-Redundant Fraction (NRF), which is equal to the number of distinct uniquely mapped reads divided by total reads, (13) PCR Bottleneck Coefficient 1 (PBC1, 14) PCR Bottleneck Coefficient 2 (PBC2). To reduce the impact of samples with lower quality, we removed datasets with normalized peak calls at FDR  $q < 0.01 < 100$ , and any additional replicates that displayed lower signal-to-noise ratios. The 90 remaining datasets were analyzed by PCA and hierarchical clustering (`hclust` function, `method = "ward. D"`), of which one NeuN+ sample was identified as an outlier (Supplementary Fig. 10). After excluding this outlier, we obtained a final set of 89 datasets from 49 donors (40 donors with matching NeuN+ and NeuN- data, and 9 donors that only contained NeuN- data). Sequencing and QC metrics for all 89 samples are described in Supplementary Table 2. The library complexity for these 89 samples are comparable to ENCODE Data Quality Metrics spreadsheet published on 2012-04-25 (<https://genome.ucsc.edu/ENCODE>).

#### Evaluation of variables affecting chromatin peaks

Prior to the differential analyses described below, we considered the effect of 37 variables that could impact the quality of the ATAC-seq results for each sample. These variables were collected from key features of the ATAC-seq processing pipeline, as well as individual sample characteristics such as case-control status, sex, and age. For example, the metadata for each sample included transposase batch, nuclei sorting date, mass, mean GC percentage of sequenced reads, mean mapped read length, alignment quality metrics, subject age at death, sex, diagnosis, and postmortem interval (PMI). Two subjects were missing PMI, therefore we used the R package MICE [36] to impute missing values using the classification and regression trees methodology.

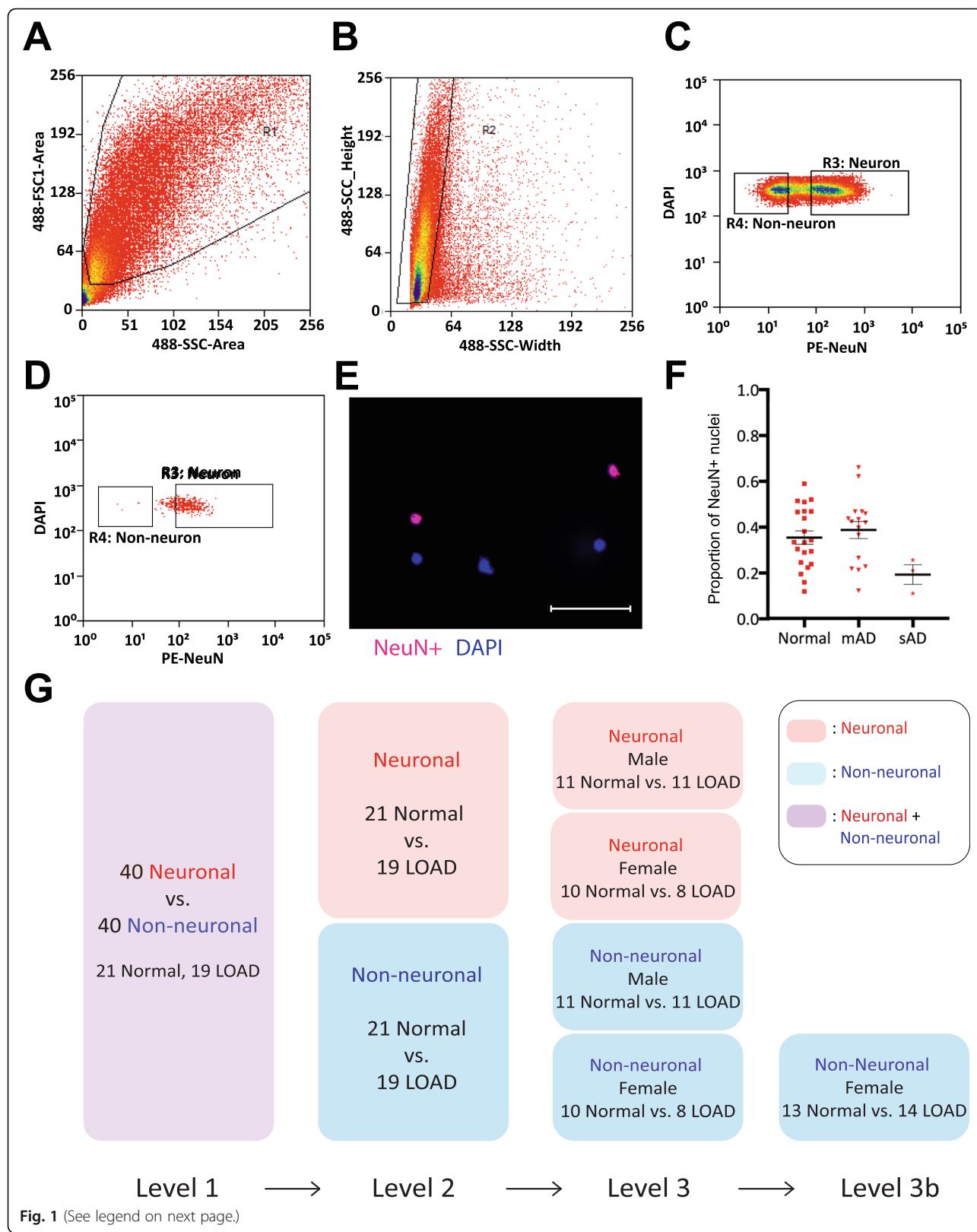
#### Covariate selection

For differential peak analyses, selection of covariates for adjustment was carefully tailored to each comparative analysis to account for the variable number of peaks between groups and to minimize false positives in peak calling. We performed a linear regression of all metadata variables against the first 10 principal components (PCs) of the TMM normalized peak quantification in an effort to identify covariates to be utilized in the differential expression analyses. In an iterative process, we selected one variable (preferentially a variable directly related to the ATAC-seq experiment, explaining one of the largest proportions of variance and with few parameters), regressed its effect on the peak quantifications and performed a new principal component analysis independent of the selected variable(s). We repeated this procedure until there were no more Bonferroni significant ( $q <$

0.05) variables associated with the peak PCs. For the comparison of neuronal and non-neuronal nuclei, we selected the following variables for our differential chromatin analysis: NRF, number of nuclei, normalized peak calls, nuclei sort date, alignment percentage, passing base pairs, age, and donor. For the comparison of LOAD vs. control neuronal nuclei, we selected normalized peak calls, alignment percentage, nuclei sort date, and number of nuclei as covariates. For the comparison of LOAD vs. control non-neuronal nuclei, we included the number of nuclei, NRF, nuclei sort date, and alignment percentage as covariates in the differential chromatin analysis. When performing a sex-stratified analysis of LOAD vs. control samples in the NeuN+ nuclei, we selected alignment percentage in the female-only subset and nuclei sort date in the male-only subset. Finally, in the sex-stratified differential chromatin analysis of LOAD vs. control NeuN- nuclei, we included normalized peak calls as a covariate in the female-only subset, and normalized peak calls and percent GC content as covariates in the male-only subset. In the supplemented comparison of LOAD vs. control NeuN- nuclei in females only (nine additional non-neuronal samples), percent GC content was included as a covariate.

#### Differential chromatin accessibility analyses

Differential ATAC-seq peaks were detected using EdgeR package (version 3.22.3), which models counts using a negative binomial distribution [37]. ATAC-seq reads corresponding to chrX and chrY were excluded due to unequal numbers of female and male donors. For each comparison, counts of reads within peaks which were merged from all of the replicates were extracted from BigWig files and normalized by weighted trimmed mean of M-values [38] (TMM). Quasi-likelihood F-tests (QLF) was performed to determine differential sites at cut off adjusted  $p$  values  $< 0.05$ . As an approximate error model, QLF works more robustly and gives more reliable Type I error rate control than the other options, especially when there are smaller numbers of replicates (EdgeR User Guides, Bioconductor package vignettes). Three levels of case-to-control chromatin accessibility comparisons were performed (Fig. 1g): *Level\_1*: 40 NeuN- vs 40 NeuN+, all samples are from matched donors; *Level\_2*: (A) NeuN-, 19 cases vs 21 controls (B) NeuN+, 19 cases vs 21 controls; *Level\_3*: (A) NeuN- females, 11 cases vs 11 controls; (B) NeuN- males 8 cases vs 10 controls; (C) NeuN+ females, 11 cases vs 11 controls; and (D) NeuN+ males, 8 cases vs 10 controls. Since the number of female samples was less than that of male, we included 9 additional NeuN- datasets (those that did not have matching NeuN+ data from the same donor but met QC criteria) into Level 3 and processed an additional comparison: (E) NeuN- females, 14 cases vs 13 control. For



(See figure on previous page.)

**Fig. 1** Isolation of nuclei from frozen brain samples and analysis of ATAC-seq data. Human postmortem frontal cortex was dissociated, nuclei were isolated and stained with the nuclear stain DAPI and a monoclonal NeuN antibody conjugated to PE. **a** Nuclei were first sorted based on their forward and side scatter from all possible events (R1 gate). **b** Single nuclei were further sorted based on their size from the doublets or larger clumps of nuclei (R2 gate). **c** DAPI positive single cells were gated as either NeuN-PE positive (neurons, R3 gate) or NeuN-PE negative (glia, R4 gate). **d** Post-sort data showing the purity of the separation between neuronal and non-neuronal nuclei. **e** Fluorescence image showing unsorted nuclei stained for NeuN (red) and DAPI (blue). The scale bar represents 100 μm. **f** Proportion of neuronal nuclei from each sample. Error bars show standard error of the mean. **g** Overview schematic of Levels 1, 2, 3, and 3b of differential analysis. Level 1 compares neuronal vs. non-neuronal for 21 normal and 19 LOAD samples. Level 2 compares normal vs. LOAD for each neuronal and non-neuronal subpopulation. Level 3 compares LOAD samples separated by female and male. Level 3b is the same comparison done after adding 9 female non-neuronal samples (3 normal and 6 LOAD)

all these comparisons we included covariates (described in “covariate selection”) in the EdgeR analysis with model design as open chromatin accessibility ~ Disease (LOAD state) + covariates. Importantly, for each analysis, we remerged and recalled the chromatin peaks. Thus, the separate analyses were not direct subsets of each other.

#### Genomic distribution analysis

ATAC-seq peaks were categorized as promoters (transcription start site to upstream 1 kb), 1st exon, intragenic (excluding 1st exon), 5' UTR, 3' UTR, and intergenic, based on human gene hg19 NCBI RefSeq gene information from the UCSC genome browser.

#### Gene ontology analysis

Differential open chromatin regions were characterized by GREAT (v 3.0.0) (<http://bejerano.stanford.edu/great/public/html/>), using the hg19 genome as background regions. Genes associated to open chromatin regions were determined by default “basal plus extension” settings (i.e., 5 kb upstream, 1 kb downstream, plus distal up to 1000 kb).

#### Overlap with LOAD GWAS loci

The tag SNP for 25 LOAD loci were obtained from the literature [7]. Genome coordinates +/- 100 kb surrounding each GWAS tag SNP were used to identify any differential open chromatin region that mapped within the region. Permutation analyses were performed by randomly selecting the same numbers of sites from the union set of ATAC-seq peak calls. For each comparison, we performed 10,000 permutations to estimate the empirical *P*-value.

#### Motif search

Transcription factor motif enrichments were evaluated using HOMER [39] (v4.10.3) and MEME Suite [40]. For each comparison, we used the centered 300 bp of open chromatin regions as input, and the union peak calls as background. GC matching was applied to the background peaks to ensure that this did not lead to spurious results. We primarily used all peaks as the background,

but to check for accuracy, we also randomly selected 10,000 peaks as the background from all peaks. For this analysis, we focused on the known motif search rather than search for de novo motifs.

#### Differential expression analysis of ROSMAP data

Gene expression data from AMP-AD was obtained from the Religious Orders Study and Memory and Aging Project (ROSMAP) [41, 42]. This study includes RNA-seq data from the dorsolateral prefrontal cortex of 724 subjects [43, 44]. limma was used to perform the differential analysis in R on normalized FPKM values obtained from RNA-seq of ROSMAP bulk tissue samples. Samples with cogdx of 1 (no cognitive impairment, *n* = 201) and 4 (Alzheimer's dementia and no other case of cognitive impairment, *n* = 222) were included in the analysis.

#### Single-nuclei library preparation and sequencing

Libraries were prepared for snRNA-seq using the Chromium Single Cell 3' Reagent Kits v3 (10X Genomics) according to the manufacturer's instructions. Nuclei were diluted in nuclease-free water to a concentration of 1 million nuclei / ml in a final volume of 100 μl, and transferred to an 8-strip tube on ice. Sixteen thousand nuclei were added to the Master Mix for a targeted nuclei recovery of 10,000. Libraries were pooled onto a single S1 flow cell, and sequenced using the Illumina NovaSeq 6000 system to obtain paired-end 2 × 100 bp reads. Sequencing saturations ranged from 59.3 to 81.2%.

#### snRNA-seq data analysis

*CellRanger* software version 3.1.0 (10X Genomics) was used to demultiplex raw Illumina base call files into FASTQ files. A pre-mRNA GTF file was generated with the pre-built GRCh38 3.0.0 human reference using the Linux utility *awk*, in order to be compatible with *CellRanger* and to include intronic reads from nuclear RNA in UMI counts for each gene and barcode. The *CellRanger* count pipeline was run to align reads to the pre-mRNA GRCh38 reference and gene expression matrices generated separately for each of the four samples were merged together into a single matrix.

For quality control and filtering, the *quickPerCellQC* function from the R package *scater* v1.14.6 [45] was used to identify low-quality cells based on QC metrics (UMI count, number of genes detected, percentage of UMIs that mapped to the mitochondrial genome). Cells in which 229 or more genes were detected and less than 17.4% of UMIs mapped to the mitochondrial genome were used in downstream analyses. Out of 21,262 nuclei in the initial dataset 18,032 remained after filtering with a median number of 2332 detected genes per nucleus. The R package Seurat v3.1.1 standard workflow was used for integration of multiple samples to combine the four samples into a unique dataset [46]. Prior to integration, gene expression was normalized for each sample by scaling by the total number of transcripts, multiplying by 10,000, and then log transforming (log-normalization). We then identified the 2000 genes that were most variable across each sample, controlling for the relationship between mean expression and variance. Next, we identified anchor genes between pairs of samples using the *FindIntegrationAnchors* function that were then passed to the *IntegrateData* function to harmonize the four samples.

We scaled the integrated dataset before running a Principal Component Analysis (PCA). To distinguish principal components (PCs) for further analysis, we used the *JackStraw* method to determine statistically significant PCs and found that up to 30 PCs were enriched in genes with a PC score that was unlikely to have been observed by chance. We then utilized the shared nearest neighbor (SNN) modularity optimization-based clustering algorithm implemented in Seurat for identifying clusters of cells. This was performed using the *FindNeighbors* function with 30 PCs, followed by the *FindClusters* function with the Louvain algorithm using a 0.4 resolution. This allowed us to assign cells into a total of 21 clusters. We applied the uniform manifold approximation and projection (UMAP) method on the cell loadings of the previously selected 30 PCs to visualize the cells in two dimensions and to separate nuclei into clusters.

Differential expression to identify cluster markers that are conserved between the samples was performed using the Seurat function *FindConservedMarkers* for each cluster on the normalized gene expression before integration. R package SingleR v1.0.1 [47] was used to annotate cell types based on correlation profiles with 713 bulk microarray samples from the Human Primary Cell Atlas [48] (HPCA) as reference expression data. Four major cell types were detected by the SingleR method using HPCA: macrophages, astrocytes, neurons, and endothelial cells. Because of the specific expression of microglial markers PTPRC and CSF1R in the macrophage cluster, as well as the differences in biological systems used in

the HPCA reference, we manually refined the HPCA annotation of this specific cluster to microglia. Identification of neuronal vs. non-neuronal clusters for some differential expression analyses was performed by determining the proportion of cells expressing the neuronal marker NeuN/RBFOX3. Clusters with more than 50% of cells expressing NeuN/RBFOX3 were defined as neuronal clusters. Control samples showed an average of 2417 (SD 169) neuronal and 2319 (SD 1149) non-neuronal nuclei, whereas LOAD samples showed 1329 (SD 294) neuronal and 2952 (SD 1783) non-neuronal nuclei. NeuN neuronal clusters and neuronal clusters identified by SingleR using the HPCA reference were very consistent with only one additional cluster in NeuN neuronal clusters (out of a total of 16 clusters). Differential expression analyses between female LOAD nuclei and control nuclei within specific groups of nuclei were performed using the Wilcoxon Rank Sum test as implemented in the *FindMarkers* function in the Seurat package. ‘min.pct’ and ‘logfc threshold’ arguments were set to 0 to allow for the testing of a majority of genes in each analysis.

## Results

### Study sample and quality control analyses

Nuclei isolated from archived frozen temporal cortex of 51 individuals with LOAD ( $n = 26$ ) or controls ( $n = 25$ ) (Supplementary Table 1) were stained and sorted using a PE-conjugated monoclonal NeuN antibody (Fig. 1a-d). Staining of pre-sorted nuclei with nuclear membrane markers was confirmed by immunofluorescence (Fig. 1e). We observed a smaller proportion of sorted neuronal (NeuN+) nuclei from total isolated nuclei in the severe LOAD cases (Table 1, Fig. 1f), as expected due to neuronal cell loss and gliosis, hallmarks of LOAD pathological progression [49, 50]. We next performed ATAC-seq using the NeuN+ and NeuN- sorted nuclei populations. A total of 90 neuronal and non-neuronal datasets passed quality control criteria (see Methods), which were derived from 40 individuals (19 LOAD cases and 21 controls) that had matched neuronal and non-neuronal data, and an additional 9 individuals (3 LOAD cases and 6 controls) that had only non-neuronal data. The final cohort of 40 consisted of 22 males and 18 females with similar post-mortem intervals (PMI); all donors were Caucasians and homozygous for *APOE* e3 (Supplementary Table 2). Subsequent analyses compared different sub-groups (Fig. 1g). Correlations of all potential numerical covariates showed expected patterns of co-linearity (Supplementary Fig. 1). The samples displayed a range of QC metrics, with some samples displaying a higher signal to noise (Supplementary Table 1). Repeated experiments on a subset of samples demonstrated that signal to noise was reproducible and thus a sample-specific

characteristic (Supplementary Table 1). In addition to randomizing sample preparation (see [Methods](#)), we found that no metadata variables were significantly associated with case-control status (Supplementary Fig. 2), indicating an absence of batch effects.

#### Differential analysis of ATAC-seq data between neuronal and non-neuronal nuclei

To demonstrate that we can robustly identify chromatin accessibility differences between major cell types of the brain, we first compared the ATAC-seq profiles between neuronal and non-neuronal nuclei groups from the entire study sample (Level 1, Fig. 1g). PCA of all samples showed that 37.5% of the total variance was explained by the cell type (neuronal versus non-neuronal, Supplementary Fig. 3, Supplementary Fig. 11). Performing quantitative differential chromatin accessibility analysis using EdgeR (FDR  $q < 0.05$ ), we found that 87,570 regions were more accessible in the neuronal population, 83,171 regions were more accessible in the non-neuronal population, and 54,484 regions were not detected as differential (Fig. 2a, Supplementary Table 10). Representative screenshots show differential ATAC-seq peaks around genes known to be expressed specifically in neuronal (Fig. 2b) or non-neuronal (Fig. 2c) cell types. A higher percentage of sites more accessible in non-neuronal population mapped to promoters, which could be explained by the more distal regulation of neuronal cell types or the greater diversity and heterogeneity of the cell types composing the non-neuronal population (Supplementary Fig. 4). Using GREAT gene ontology analysis [51], neuronal-specific regions were enriched with genes associated with neuronal function, while non-neuronal specific regions were enriched with genes implicated in glial function (Supplementary Table 3).

Motif analysis for regions with increased chromatin accessibility in neuronal nuclei showed enrichment for the transcription factor recognition sites of Early Growth Response 2 (EGR2), Regulatory Factor X1 (RXF1), and Myocyte Enhancer Factor 2C (MEF2C) [52–54], while motif analysis for regions with increased chromatin accessibility in non-neuronal nuclei showed enrichment for the SRY-related HMG-box (SOX) transcription factor family [55] (Supplementary Table 4). When compared to a previous NeuN sorted-nuclei ATAC-seq study [56] using a smaller number ( $n = 8$ ) of healthy brain samples, we found a substantial degree of overlap (Supplementary Fig. 5a) with similar peak length distribution (Supplementary Fig. 5b). However, the 5x larger number of samples used in our study identified  $> 80,000$  additional significantly differential chromatin sites.

#### Comparison of sorted nuclei vs. bulk brain tissue homogenate

To our knowledge, no study has directly compared chromatin accessibility differences between bulk brain tissue,

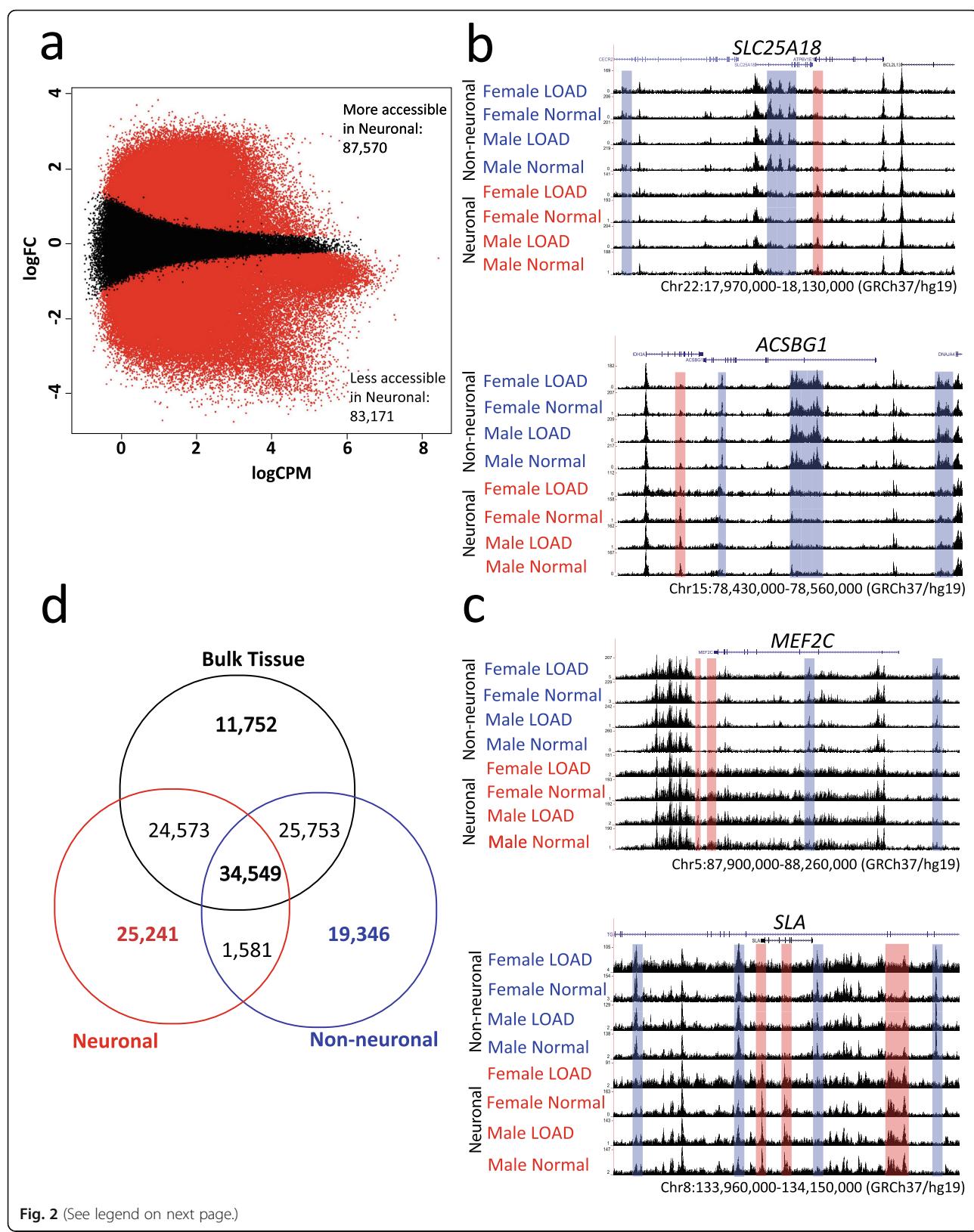
neuronal, and non-neuronal fractions. Bulk brain ATAC-seq data was generated using pulverized frozen whole tissue homogenate samples from six individuals for whom high-quality ATAC-seq data was collected from neuronal and non-neuronal populations (Supplementary Table 5). Binary peak overlap analysis shows that while there are some regions that are only accessible in bulk tissue, a substantially larger number of sites are uniquely detected in the nuclei populations of neuronal and non-neuronal only (Fig. 2d). This observation justifies the performance of ATAC-seq analysis by brain cell-type to reduce cellular heterogeneity as it allows the identification of more cell type specific signals. Quantitative differential chromatin accessibility (EdgeR comparisons (Supplementary Fig. 6) and GO annotations (Supplementary Table 6–8) showed that chromatin accessibility sites specifically identified in bulk tissue, neuronal and non-neuronal populations displayed many biologically relevant pathways.

#### Identification of LOAD-associated differences in chromatin accessibility

To determine the association of LOAD status with changes in chromatin accessibility, we performed differential ATAC-seq analysis of 19 LOAD compared to 21 healthy controls stratified by brain cell-type (neuronal and non-neuronal only populations; Level 2, Fig. 1g). The comparison using the neuronal nuclei data resulted in 211 neuronal chromatin differences (FDR  $q < 0.05$ ) between LOAD and control (Fig. 3a, Supplementary Table 11), while the analysis of the non-neuronal nuclei detected no differential chromatin between LOAD and control (Fig. 3b). Motif analysis using the 141 sites that showed decreased chromatin accessibility in LOAD neuronal nuclei were enriched for transcription factor motifs including Wilms Tumor protein (WT1), Early Growth Response 1 (Egr1), Retinoic acid receptor gamma (RAR $\gamma$ ), and Kruppel like factor 14 (KLF14) (Fig. 3c) that have been reported relevant to brain function [57–60]. No significant motif enrichment was detected from 70 sites (FDR  $q < 0.05$ ) with increased chromatin accessibility in neuronal LOAD samples.

#### Identification of sex-dependent chromatin accessibility differences between LOAD and control samples

Since sex has an effect on LOAD onset and progression [61], we performed a sex-stratified differential analysis of LOAD compared to control (Level 3, Fig. 1g) to examine whether the effect of sex on LOAD risk is mediated, at least in part, by chromatin remodeling. The differential analysis of the neuronal groups did not yield significant differences between LOAD and normal when stratified by sex for either females or males (FDR  $q > 0.05$ ) (Fig. 4a-b). Analysis of the non-neuronal nuclei identified 24



(See figure on previous page.)

**Fig. 2** Level 1 comparison of ATAC-seq data from neuronal vs non-neuronal nuclei. **a** MA plot showing differential ATAC-seq sites between neuronal (blue) vs. non-neuronal regions (red). Red dots represent ATAC-seq peaks that are significantly different between groups (FDR  $q < 0.05$ ). **b** ATAC-seq data around non-neuronal-specific genes *SLC25A18* (upper panel) and *ACSBG1* (lower panel). Boxes highlight peaks that are more accessible in neuronal (red) or non-neuronal (blue) nuclei. **c** ATAC-seq data around neuron-specific genes *MEF2C* (upper panel) and *SLA* (lower panel). All regions indicate hg19 coordinates. **d** Venn diagram of ATAC-seq peaks detected in whole tissue, sorted neuron and sorted non-neuron nuclei from 6 donor-matching samples

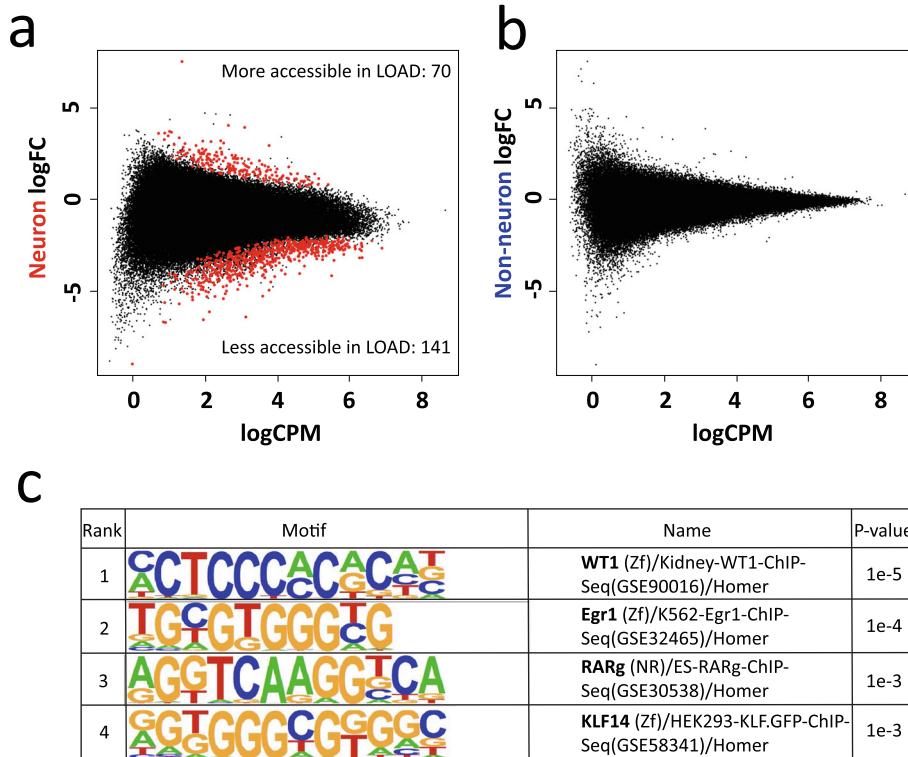
chromatin accessibility differences in female LOAD compared to control (Fig. 4c, Supplementary Table 12), while no significant differences were identified when the analysis was performed in males (Fig. 4d). To further investigate this trend, we increased the sample size with 9 additional non-neuronal samples (6 female controls and 3 female LOAD cases). Differential analysis using ATAC-seq data from the larger female non-neuronal dataset resulted in 842 differential sites between LOAD and control (FDR  $q < 0.05$ ) (Fig. 4e, Supplementary Table 13).

Comparison of these results with the LOAD associated chromatin accessibility sites obtained from the differential analysis of female neuronal nuclei showed that while there are differences detected in both female neuronal and glial cells in LOAD cases versus controls, we detected larger effect sizes in the female glia cells (Supplementary Fig. 12a). In addition, several differential

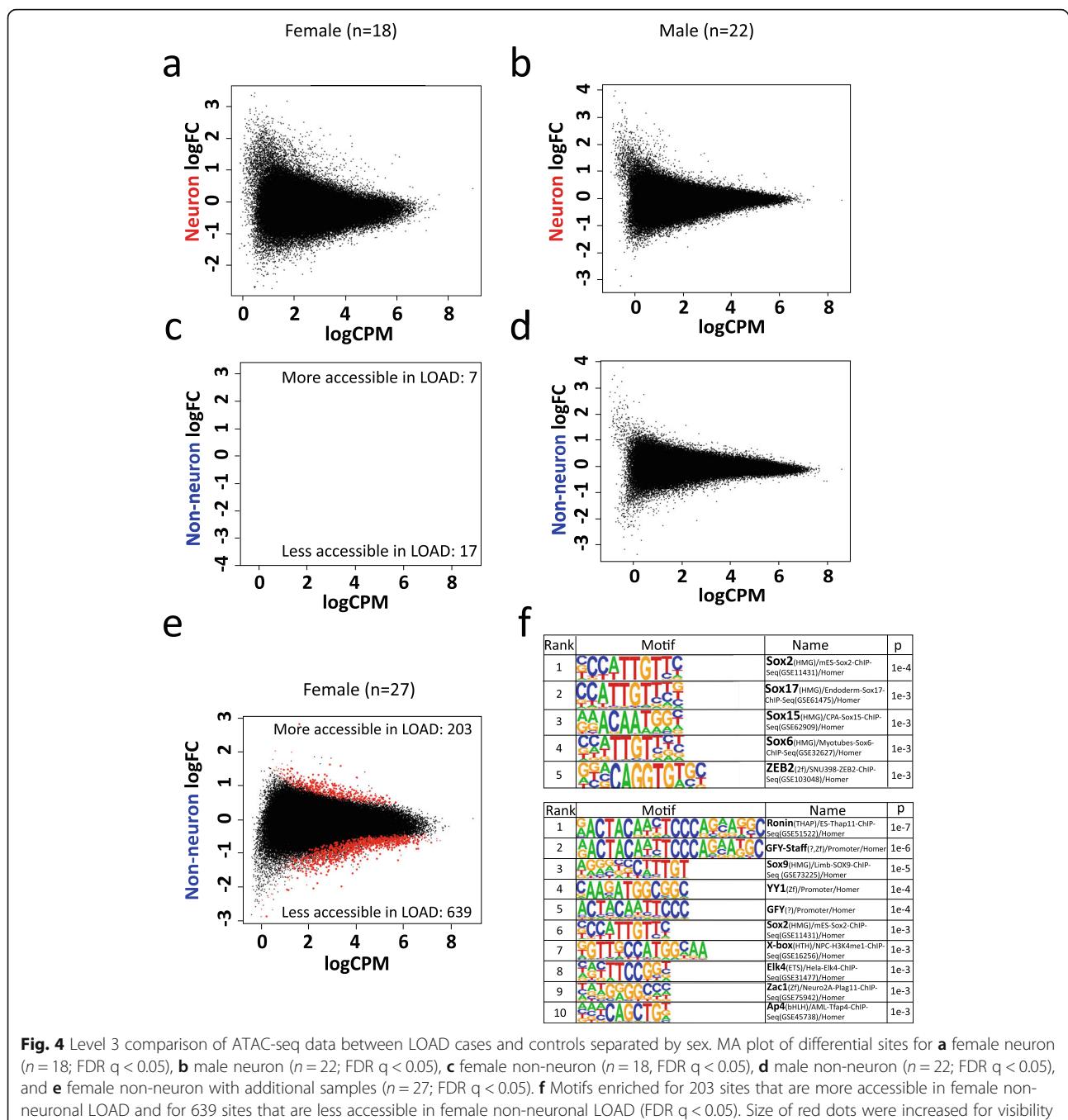
regions identified in female non-neuronal nuclei were also observed in male non-neuronal nuclei with similar trends, however, these did not reach statistical significance (Supplementary Fig. 12b). Nonetheless, these associations in the female group showed a stronger effect size (Supplementary Fig. 12b).

We performed motif enrichment using the 203 sites that were more accessible in non-neuronal female LOAD and found significantly enriched motifs for the SOX family (Fig. 4f). The analysis of 639 sites that displayed decreased accessibility in non-neuronal female LOAD was enriched for transcription factors known to be highly associated with glia or neuron functions, such as RONIN, SOX9, YY1 and ELK4 [62–65] (Fig. 4f).

To determine whether binding of transcription factors identified in female non-neuronal cells (Level 3, Fig. 1g) has downstream effects on expression of



**Fig. 3** Level 2 comparison of ATAC-seq data between LOAD cases and controls. MA plots of differential chromatin sites for **a** neuronal and **b** non-neuronal nuclei. Red dots represent differential ATAC-seq sites with FDR  $q < 0.05$ . **c** Motifs that are enriched in neuronal ATAC-seq sites that are less accessible in LOAD samples. Size of red dots were increased for visibility



**Fig. 4** Level 3 comparison of ATAC-seq data between LOAD cases and controls separated by sex. MA plot of differential sites for **a** female neuron ( $n=18$ ; FDR  $q < 0.05$ ), **b** male neuron ( $n=22$ ; FDR  $q < 0.05$ ), **c** female non-neuron ( $n=18$ , FDR  $q < 0.05$ ), **d** male non-neuron ( $n=22$ ; FDR  $q < 0.05$ ), and **e** female non-neuron with additional samples ( $n=27$ ; FDR  $q < 0.05$ ). **f** Motifs enriched for 203 sites that are more accessible in female non-neuronal LOAD and for 639 sites that are less accessible in female non-neuronal LOAD (FDR  $q < 0.05$ ). Size of red dots were increased for visibility

target genes, we used the publicly available ROSEMAP bulk RNA-seq data to perform differential expression analyses of those genes for LOAD vs. Normal. We found that several genes exhibited changes in the expected direction, including *ZAC1* (*PLAGL1*), *YY1*, and *SOX2* (Supplementary Table 15). This provides further mechanistic insights into gene dysregulation underlying LOAD pathogenesis in females. Last, we also used the chromatin accessibility differential sites

identified in female non-neuronal LOAD (Level 3) for GO analysis and found pathways involved in immune response and myelination (Supplementary Table 9).

#### Overlap of LOAD-specific differential chromatin accessibility sites with LOAD GWAS regions

To determine the relationship of the LOAD-specific chromatin accessibility sites we compared these data

with LOAD GWAS regions [7]. We defined LOAD GWAS regions by anchoring on the top 25 associated SNPs  $\pm$  100 kb, and cataloged LOAD-GWAS regions of 200 kb each. Importantly, we were comparing genomic regions of chromatin accessibility to the GWAS regions, not chromatin QTLs, to GWAS loci. Of the 211 LOAD-specific differential chromatin accessibility sites in neuronal samples (FDR  $q < 0.05$ , Fig. 3a), we identified five sites that overlap four of the 25 LOAD GWAS regions (Table 2). We show representative examples of overlapping LOAD-specific differential chromatin accessibility sites surrounding the *PTK2B* and *CLU* GWAS loci (Fig. 5a, Supplementary Fig. 7). Using permutation testing, we did not detect a significant enrichment with LOAD GWAS regions compared to regions selected at random (Supplementary Fig. 8a-b).

Of the 842 differential sites in non-neuronal female LOAD (FDR  $q < 0.05$ , Fig. 4e), we detected nine differential sites that overlap seven of the 25 LOAD-GWAS regions (Table 2). Using permutation analysis (see Methods), we found significant enrichment in the amount of overlap with sites less accessible in non-neuronal female LOAD ( $P < 0.05$ , Supplementary Fig. 8c-d). Representative examples of differential chromatin accessibility sites between LOAD cases and controls

overlapping representative LOAD-GWAS regions are shown for the loci *APOE* (Fig. 5b, Supplementary Fig. 7b), and *IQCK* (Fig. 5c, Supplementary Fig. 7c). While differences around the *APOE* and *IQCK* locus are more pronounced in females, we also detect subtle differences in the same direction in males (Fig. 5b-c, Supplementary Fig. 7b-c). The identification of LOAD differences in chromatin accessibility in the *APOE* region, suggests a possible regulatory mechanism independent of the e4 allele effect at this locus.

#### Overlap of LOAD-specific differential chromatin accessibility sites in female non-neurons with single nuclei RNA-seq data

Next, we performed functional validation of key findings from the chromatin accessibility differential sites in female LOAD non-neuronal cells that overlapped LOAD-GWAS regions (Table 3, Supplementary Fig. 9). We focused the analysis on two representative LOAD differential chromatin accessibility sites in female LOAD non-neuronal cells that overlapped LOAD-GWAS loci: (1) LOAD more accessible sites surrounding the *ECHDC3* locus (annotated by the proximate gene to the associated SNP) and (2) LOAD less accessible sites surrounding the *ABCA7* locus. The loci were defined by anchoring on

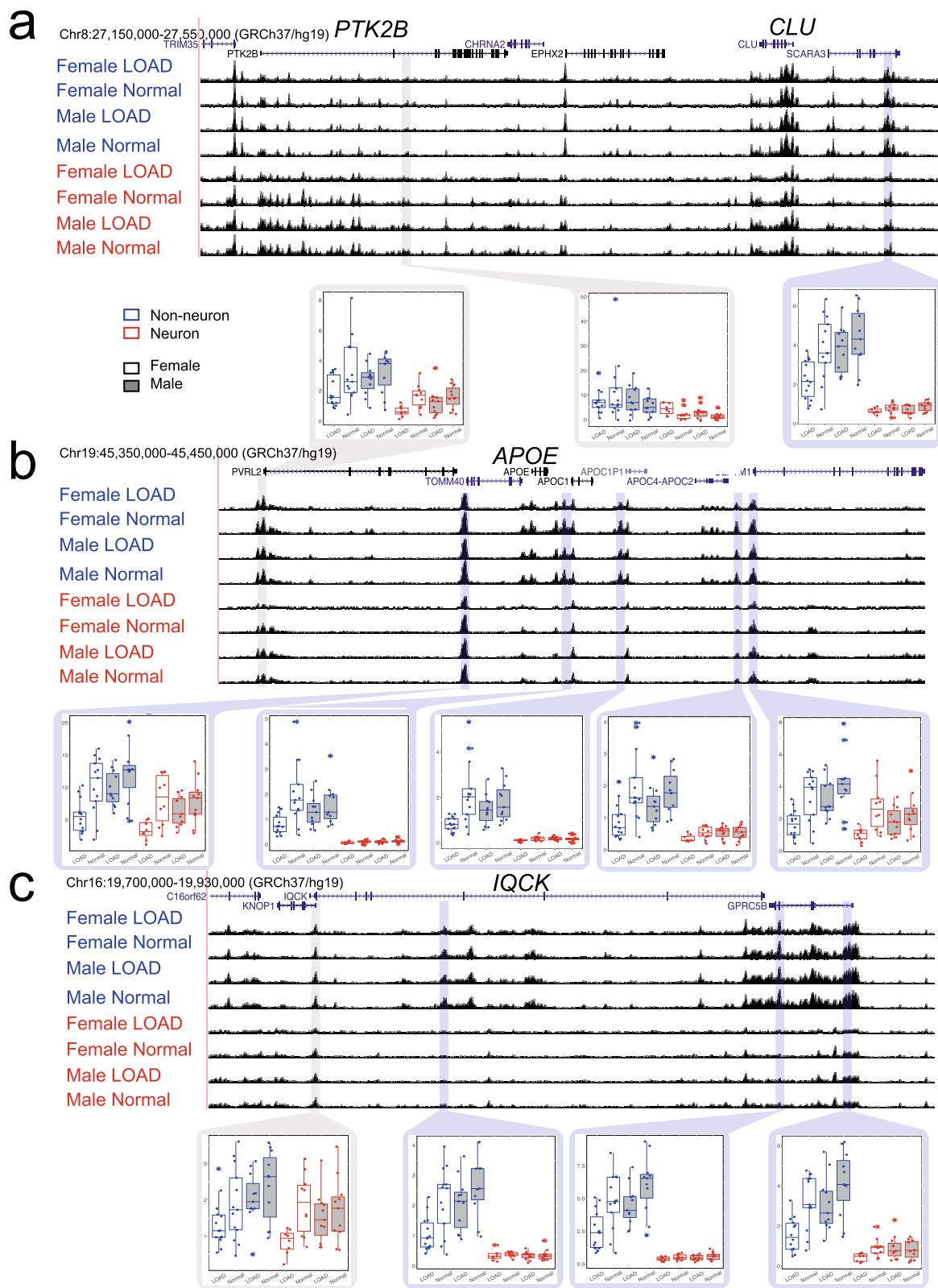
**Table 2** Neuron control vs LOAD differential peaks overlapped with 25 LOAD-GWAS regions as well as female glia control vs LOAD differential peaks overlapped with 25 LOAD-GWAS regions. For each panel, upper: more open in LOAD; lower: less open in LOAD

#### Neuron control vs. LOAD differential peaks overlapped with 25 LOAD-GWAS regions

Differential sites with FDR < 0.05	Peak call	GWAS $\pm$ 100 kb	Gene	SNP	SNP Position
	More accessible in LOAD				
	chr6:32551592-32552707	chr6:32475406-32675406	<i>HLA-DRB1</i>	Rs9271058	chr6:32575406
	chr6:32489394-32490069	chr6:32475406-32675406	<i>HLA-DRB1</i>	Rs9271058	chr6:32575406
	Less accessible in LOAD				
	chr20:54994099-54994641	chr20:54897568-55097568	<i>CASS4</i>	Rs6024870	chr20:54997568
	chr8:27209401-27210247	chr8:27119987-27319987	<i>PTK2B</i>	Rs73223431	chr8:27219987
	chr8:27563212-27563671	chr8:27367686-27567686	<i>CLU</i>	Rs9331896	chr8:27467686

#### Female glia control vs. LOAD differential peaks overlapped with 25 LOAD-GWAS regions

Differential sites with FDR < 0.05	Peak call	GWAS $\pm$ 100 kb	Gene	SNP	SNP Position
	More accessible in LOAD				
	chr10:11784165-11784736	chr10:11620308-11820308	<i>ECHDC3</i>	Rs7920721	chr10:11720308
	Less accessible in LOAD				
	chr14:92966426-92967442	chr14:92832828-93032828	<i>SLC24A4</i>	Rs12881735	chr14:92932828
	chr16:19894588-19895301	chr16:19708163-19908163	<i>IQCK</i>	Rs7185636	chr16:19808163
	chr17:61627113-61628372	chr17:61438148-61638148	<i>ACE</i>	Rs138190086	chr17:61538148
	chr19:1101756-1102160	chr19:956492-1156492	<i>ABCA7</i>	Rs3752246	chr19:1056492
	chr19:45416036-45416558	chr19:45311941-45511941	<i>APOE</i>	Rs429358	chr19:45411941
	chr19:45428555-45429362	chr19:45311941-45511941	<i>APOE</i>	Rs429358	chr19:45411941
	chr19:45454352-45455071	chr19:45311941-45511941	<i>APOE</i>	Rs429358	chr19:45411941
	chr7:100025783-100028055	chr7:99991795-100191795	<i>NYAP1</i> ( <i>ZCWPW1</i> )	Rs12539172	chr7:100091795



**Fig. 5** (See legend on next page.)

(See figure on previous page.)

**Fig. 5** Differential LOAD specific ATAC-seq peaks around LOAD-GWAS regions. Screenshots of ATAC-seq data around **a** *PTK2B*, *CLU*, **b** *APOE*, and **c** *IQCK* loci. Box plots show ATAC-seq read counts for individual ATAC-seq peaks (blue frames highlight significant differential peaks for cases vs. controls, gray frames show control peaks that are not differential between cases and controls). Box plots are color coded for non-neuronal (blue) neuronal (red), female (no fill), and male (gray fill). All tracks show hg19 coordinates and all y-axes on tracks range from 0 to 250. For box plots, the line within each box represents the median, and the top and bottom borders of the box represent the 25th and 75th percentiles, respectively. The top and bottom whiskers of the box plots represent the 75th percentile plus 1.5 times the interquartile range and the 25th percentile minus 1.5 times the interquartile range, respectively

**Table 3** Differential snRNA-seq expression from level 3 female glia controls vs cases that are within 500 kb of LOAD-GWAS SNPs at loci found to be more accessible (rs7920721) or less accessible (rs3752246) in LOAD by ATAC-seq

SNP (coordinates, proximate gene)	Genes	NeuN- non-neuronal		HPCA-annotated astrocytes		HPCA-annotated microglia	
		Mean log (FC)	Adjusted p value	Mean log (FC)	Adjusted p value	Mean log (FC)	Adjusted p value
<b>rs7920721 (chr10:11,678,309, <i>ECHDC3</i>)</b>	<i>CELF2</i>	<u>0.087</u>	1	-0.090	1	<u>0.47</u>	<b><math>4.90 \times 10^{-16}</math></b>
	<i>USP6NL</i>	<u>0.12</u>	0.32	-0.022	1	<u>0.29</u>	1
	<i>ECHDC3</i>	<u>0.067</u>	1	<u>0.016</u>	1	<u>0.31</u>	1
	<i>UPF2</i>	-0.34	<b>0.00042</b>	-0.31	1	-0.48	1
	<i>DHTKD1</i>	-0.16	1	-0.23	1	<u>0.011</u>	1
	<i>SEC61A2</i>	<u>0.028</u>	1	-0.028	1	<u>0.15</u>	1
	<i>NUDTS5</i>	-0.032	1	-0.097	1	<u>0.21</u>	1
<b>rs3752246 (chr19:1,056,493, <i>ABCA7</i>)</b>	<i>BSG</i>	0.035	1	0.10	0.32	<u>-0.11</u>	1
	<i>POLRMT</i>	<u>-0.047</u>	1	<u>-0.030</u>	1	0.061	1
	<i>FSTL3</i>	<u>-0.12</u>	0.065	<u>-0.13</u>	1	0.047	1
	<i>PALM</i>	<u>-0.071</u>	1	0.010	1	<u>-0.25</u>	1
	<i>PTBP1</i>	<u>-0.034</u>	1	<u>-0.053</u>	1	0.065	1
	<i>AZU1</i>	<u>-0.012</u>	1	<u>-0.015</u>	1	<u>-0.039</u>	1
	<i>MED16</i>	<u>-0.036</u>	1	<u>-0.051</u>	1	<u>-0.19</u>	1
	<i>ARID3A</i>	0.089	0.97	0.047	1	0.12	1
	<i>TMEM259</i>	0.050	1	0.089	1	<u>-0.11</u>	1
	<i>ABCA7</i>	<u>-0.013</u>	1	<u>-0.022</u>	1	0.075	1
	<i>STK11</i>	<u>-0.024</u>	1	<u>-0.041</u>	1	<u>-0.13</u>	1
	<i>CBARP</i>	0.013	1	0.017	1	0.026	1
	<i>ATP5F1D</i>	0.040	1	0.11	1	<u>-0.032</u>	1
	<i>MIDN</i>	0.042	1	<u>-0.073</u>	1	0.14	1
	<i>CIRBP</i>	<u>-0.71</u>	<b><math>1.52 \times 10^{-57}</math></b>	<u>-0.80</u>	<b><math>1.27 \times 10^{-42}</math></b>	<u>-0.51</u>	1
	<i>C19orf24</i>	0.033	1	0.014	1	0.12	1
	<i>NDUFS7</i>	0.083	<b>0.016</b>	0.11	<b>0.0014</b>	<u>-0.014</u>	1
	<i>GAMT</i>	<u>-0.10</u>	1	<u>-0.065</u>	1	<u>-0.28</u>	1
	<i>RPS15</i>	<u>-0.14</u>	1	<u>-0.19</u>	0.66	0.088	1
	<i>APC2</i>	<u>-0.027</u>	1	0.022	1	<u>-0.14</u>	1
	<i>PCSK4</i>	<u>-0.033</u>	1	<u>-0.025</u>	1	<u>-0.10</u>	1

Negative mean log (FC) values indicate down-regulation in LOAD and positive values indicate up-regulation in LOAD. Significant adjusted p values are bolded and mean log (FC) values consistent with ATAC-seq data are underlined. FC = fold change. Genes that were not detected or that had mean log (FC) values below 0.01 for any of the three groups were excluded from the table. Refer to detailed genomic structure presented in Supplemental Fig. 9

the associated SNPs +/- 500 kb, rs7920721 (*ECHDC3*) and rs3752246 (*ABCA7*) using the UCSC Genome Browser [66] (<http://genome.ucsc.edu/>) GRCh38/hg38 assembly released December 2013 (Table 3, Supplementary Fig. 9). Single-nuclei (sn)RNA-seq data was collected from female brains using a small subset group of two LOAD and two control (Supplementary Table 2, marked with '\*'), and differential expression analysis of female LOAD nuclei compared to control nuclei was performed for all genes mapped with these regions in three cell-type specific groups: 1) NeuN- non-neuronal, 2) Human Primary Cell Atlas (HPCA)-annotated astrocyte, and 3) HPCA-annotated microglia. Pseudogenes, RNA genes, and novel transcripts were excluded from the analysis. Out of 54 total genes within 1 Mb of the two SNPs, 8 within the *ECHDC3* locus and 46 within the *ABCA7* locus (Supplementary Fig. 9), many were up- or down-regulated as predicted by chromatin accessibility profiles (i.e. increased chromatin accessibility overlapping promoters and enhancers was associated with upregulation and vice versa). Overall 28 (51.9%), 23 (42.6%), and 24 (44.4%) genes of the NeuN-, astrocyte, and microglia clusters, respectively, showed differential expression with trends corresponding to the changes in chromatin accessibility. Of those genes, some showed nominal significance (6 (21.4%) for NeuN-, 3 (13.0%) for astrocyte, and 5 (20.8%) for microglia clusters), but did not reach adjusted statistical significance, while other trends did not reach nominal significance. This could be explained by the small sample size and/or number of cells in the clusters. Furthermore, we examined the consistency between our dataset and a similar recently reported dataset [28]. To address this question we conducted differential expression analysis for HPCA-annotated neurons, astrocytes, and microglia and compared the results for the top 15 significant genes, i.e. 5 from each cell type identified by the previous study [28]. We found that out of these 15 genes, our results show the same directionality for 13 genes, 10 of which also had significant adjusted *p*-values (Supplementary Table 14), providing further validation of our findings.

Table 3 catalogues the genes surrounding the *ECHDC3* and *ABCA7* loci, excluding 26 genes that were not detected or had low fold change values. Out of seven genes positioned within the 1 Mb region of the LOAD associated *ECHDC3* SNP, we found that *CELF2* was significantly upregulated in LOAD HPCA-annotated microglia clusters (mean log (fold change) = 0.47, adjusted *p* =  $4.90 \times 10^{-16}$ , Table 3). In addition, out of 21 genes mapped within the 1 Mb region of the LOAD associated *ABCA7* SNP, *CIRBP* was significantly downregulated in LOAD NeuN- non-neuronal clusters (mean log (fold change) = -0.71, adjusted *p* =  $1.52 \times 10^{-57}$ , Table 3) and,

HPCA-annotated astrocyte clusters (mean log (fold change) = -0.80, adjusted *p* =  $1.27 \times 10^{-42}$ , Table 3). These results demonstrated that alteration in chromatin accessibility in LOAD-GWAS loci correlates with changes in gene expression. Furthermore, the results suggested that the target genes within LOAD-GWAS loci affected by LOAD specific changes in chromatin state may not be simply interpreted as the most proximate gene to the associated SNP.

Our pilot analysis to identify the intersection of LOAD- and glia-specific ATAC-seq sites and snRNA-seq patterns discovered novel LOAD regulatory elements that lead to gene expression changes in LOAD. However, this pilot analysis was limited by the small number of samples (2 cases and 2 control) and therefore validation of the findings warrant further analysis using a larger RNA-seq dataset from LOAD patients and controls. Furthermore, intersecting these data sets provided a functional validation for the top significant findings from the ATAC-seq experiments showing sex-dependent LOAD-specific open and closed chromatin accessibility sites in non-neuronal cells. Collectively, these results suggest that integrating brain cell-type specific 'omics data is a powerful mechanistic strategy to discover regulatory elements that impact expression of disease genes in a cell-type specific manner.

## Discussion

Decoding the genetic and genomic mechanisms of LOAD is a major challenge in the post-GWAS era, since the majority of the LOAD associated SNPs are in non-coding regions [1–7]. Noncoding disease-associated loci have been shown to be enriched for regulatory elements in tissues and cells relevant to the disease [56, 67–69]. Thus, post-GWAS research requires an in-depth characterization of cell-type specific DNA regulatory elements. Mapping chromatin accessibility has been widely used to identify the location of active DNA regulatory elements, including promoters, enhancers, and insulators [70–73]. We performed the first systematic interrogation of the chromatin accessibility landscape in neuronal and non-neuronal sorted nuclei from LOAD and healthy brains.

To our knowledge, this study represents the first and the most comprehensive dataset to date of chromatin accessibility in LOAD brains and matched controls. Furthermore, this study was performed with brain cell-type specific resolution, i.e. neuronal and non-neuronal cells. The study has four major findings for the field of LOAD epigenetics. First, we have generated a map of LOAD-associated cell-type specific chromatin accessibility sites. Second, we provide a catalogue of female-specific disease-associated chromatin accessibility sites in non-neuronal cells. Third, we suggest a non-coding

regulatory mechanism, namely chromatin accessibility, by which ~25% of LOAD-GWAS loci may exert their effect. Fourth, we have demonstrated that LOAD associated changes in chromatin accessibility can result in gene dysregulation through their overlap with the transcriptome profile of LOAD-GWAS loci. Overall, these results suggest that the *cis* interactions between regulatory elements and key genes contribute, at least in part, to the development and/or progression of LOAD. Of note, 87% of the differentially accessible chromatin peaks in the female NeuN- analysis and 92% of the differentially accessible chromatin peaks in the NeuN+ analysis fall within known regions of topologically associating domains (TADs) from fetal brain [74]. This suggests that chromatin regions associated with risk for LOAD belong to highly interacting regulatory domains. Furthermore, several LOAD loci may exert their pathogenic effects in a cell-type specific manner while others act in multiple cell types to drive LOAD pathogenesis. Alternatively, these cell-type specific and common changes in chromatin structure may represent secondary effects as consequences of disease-related processes such as neurodegeneration and gliosis.

A growing number of epigenome-wide association studies (EWAS) in LOAD have profiled DNA methylation, hydroxymethylation, and histone acetylation marks (H4K16ac), and assessed associations with LOAD risk and other related endophenotypes including the burden of pathology [17–24]. These studies have been a powerful approach to validate known LOAD loci, discover new candidate genes, and identify disease-related pathways [17–24]. However, the majority of the LOAD EWAS datasets were generated using bulk brain tissues, and the heterogeneous cell-type populations of brain tissue samples pose technical and biological limitations. Our results showed that a substantially larger number of ATAC-seq sites are uniquely detected in the neuronal and non-neuronal cells, compared to bulk tissues obtained from the same donors. This observation demonstrates the importance of cell-type specific epigenomic studies relative to bulk tissues, as it reduces cellular heterogeneity allowing the identification of more cell-type specific signals. Collectively, our outcomes open a new window for the exploration of the particular cell-types that contribute to LOAD pathogenesis, and the genes and pathways that mediate the cell-type specific pathogenic effects.

These data advance the mechanistic understanding of LOAD, and moreover, uncover new candidate LOAD loci. To date, in addition to this study only five others (two transcriptome, two DNA-methylation, and one histone acetylation) have compared genomic signatures stratified by different cell types in the LOAD brain using sorted- and single- nuclei based methodologies [27–29,

70, 75]. Altogether, our and other studies demonstrated the importance of applying these cell-type specific approaches in molecular analyses of brain tissues and highlight the impact of transitioning into single-cell based ‘omics studies in LOAD functional genomic research. In addition, we show that integrating our cell-type specific ATAC-seq data with our scRNA-seq LOAD data corroborate the interpretation of the results and provide functional validation. By aligning the two datasets, we identified LOAD-specific female-dependent non-neuronal regulatory elements around two differentially expressed genes in LOAD glia cells. Moreover, analysis of chromatin accessibility differential sites in female LOAD non-neuronal cells that overlapped LOAD-GWAS regions not only provided functional validation and established the link between LOAD -GWAS -ATAC-seq and -snRNA-seq but also demonstrated that genes other than the most proximate to the associated SNP may play a role in LOAD pathogenesis. These results exemplify the potential of integration of cell-type specific datasets to validate known LOAD loci and also to identify new candidate genes. In future studies, a larger sample size may allow conducting a chromatin accessibility QTL study to determine colocalization with GWAS loci, and may also provide additional power to detect differential sites in non-neuronal nuclei as in a recent study [76]. In addition, investigating the relationship between neuropathology severity and cell type-specific chromatin accessibility signatures in a larger cohort will be crucial to understanding how gene regulation in specific cell types is impacted by the progression of the disease.

Several bulk tissue ChIP-seq studies have used functional genomics and integrative systems biology approaches to infer cell types. Consistent with our findings, these epigenomic studies strongly suggest that non-neuronal cell types contribute to LOAD-specific histone marks associated with active regulatory elements (promoters and enhancers). It was reported that LOAD GWAS loci were enriched in enhancer elements specific to immune cells [77] and tangle-associated H3K9ac signals located in both promoters and enhancers were significantly associated with modules classified as non-neuronal [78]. Recently, two studies used FANS-sorted nuclei followed by ChIP-seq and demonstrated that microglia were the non-neuronal cell type contributing to LOAD epigenomic signatures. Nott et al. found that LOAD SNP heritability was most significant in microglial enhancers [79], and Ramamurthy et al. showed that hyper-acetylated peaks in microglia colocalize more with LOAD SNPs than the histone acetylome of other cell types [75]. Collectively, ours and others’ studies point to non-neuronal epigenomic dysregulation, likely microglial, as a major contributing factor to LOAD pathogenesis.

Sex is an important factor in LOAD etiology and there are sex differences in disease risk, progression and clinicopathological phenotypes [80]. Foremost, there is a sex-dependent difference in LOAD prevalence and almost two-thirds of LOAD patients are female [81]. Historically, it has been attributed to the longer average life expectancy in women [82], however, recent research suggests that other factors, such as the sudden drop in the level of sex hormones (estrogens) in women at menopause, contribute to the differences in susceptibility between males and females [61]. It was also reported that women manifest faster disease progression and cognitive decline, increased brain atrophy and pathological burden largely driven by neurofibrillary tangles [61], and a more advanced disease stage as indicated by CSF biomarkers, especially higher concentrations of total tau and phosphorylated tau. Although several conflicting reports suggested opposite trends [83], the effect of sex on LOAD has been widely accepted. Nonetheless, the molecular mechanisms underlying the role of sex as a risk factor in LOAD are understudied. Our study provides new insights into these gaps in knowledge showing sex-dependent changes in chromatin structure between LOAD and control brains. We identified hundreds of LOAD differential chromatin accessibility sites specific to females, which overlap nearly one-third of all LOAD GWAS regions. Since the majority of differential chromatin accessibility sites do not overlap LOAD-GWAS regions, these represent novel candidate LOAD loci. Moreover, a female-specific effect on LOAD-associated changes in chromatin accessibility appeared exclusively in glial cells and resulted in nearly three-fold overrepresentation of sites that were more closed in female LOAD patient samples. However, we cannot rule out the possibility that LOAD-associated changes in chromatin accessibility also occur in glia from male LOAD patients, but because of the plausible much smaller effect size in males, we could be underpowered to detect significant associations in our male cohort. These results warrant further investigation to determine the effects of sex-dependent chromatin remodeling on dysregulation of gene expression in the context of LOAD. In this respect, the recent scRNA-seq study [28] also reported sex-dependent LOAD effects and specifically observed a sex-specific differential transcriptional response to LOAD pathology, enrichment of females cells in LOAD-associated cell subpopulation, and higher expression in females for the marker genes of LOAD-associated cellular subpopulations [28]. Future integration of diverse ‘omics datasets stratified by sex will decipher the underpinning mechanisms of sex differences in LOAD by establishing the cross interactions between sex-dependent chromatin structure and function, transcriptome and LOAD phenotypic outcomes.

One of the loci that showed sex-dependent effects in our study is the *APOE* linkage disequilibrium (LD) region. The e4 allele of the *APOE* gene is the first identified, most highly replicated, and the strongest genetic risk factor for LOAD [84, 85]. Furthermore, LOAD GWAS have confirmed strong associations with the *APOE* LD genomic region, and no other LOAD-association remotely approached the same level of significance [1, 3, 86]. Interestingly, female carriers of *APOE* e4 have an increased risk of LOAD versus male and the adverse effect of *APOE* e4 on LOAD biomarkers was generally stronger in women versus men [87–92]. Overall the *APOE* LD region displayed more open chromatin in glia versus neurons as expected since this gene is much more highly expressed in glia. Interestingly, we found decreased chromatin accessibility at multiple sites in female LOAD glia across the *APOE* region. This result provides molecular clues to the observations that *APOE* e4 allele confers a greater risk for LOAD in women than in men [87–91]. In addition, significant downregulation of *APOE* in astrocytes from LOAD brains, although not sex-dependent, was reported by scRNA-seq analysis [28, 29]. This evidence is consistent with the trend we found of more closed chromatin in LOAD samples, providing functional validation to our result\_ENREF\_28 [28]. In summary, we proposed molecular insights based on chromatin structure that may explain, at least in part, some aspects related to the role of *APOE* in LOAD.

In conclusion, this LOAD genomic research pioneers the approach of brain cell-type specific chromatin accessibility profiling and lays the foundation for additional sorted- and single-nuclei ‘omics analyses in LOAD. Our outcomes warrant further investigations using a larger sample size to enhance the discovery smaller LOAD-associated effects on chromatin accessibility and to allow the utilization of a single peakset and set of covariates to perform multiple testing. Future and ongoing studies using even more advanced single-cell technologies will generate complementary ‘omics datasets with finer cell-type resolution from larger well-characterized LOAD cohorts. Data sharing via publicly available portals, such as Accelerating Medicines Partnership-Alzheimer’s Disease (AMP-AD), will facilitate integrative single-cells ‘omics towards moving forward our understanding of the underpinning genetic drivers and molecular mechanisms of LOAD.

## Conclusions

This study shows that LOAD-associated changes in chromatin accessibility in the brain occur on a cell type-specific level and may explain up to 25% of GWAS loci conferring risk for LOAD. We found that changes in accessibility in non-neuronal cells were sex-dependent and specific to female samples, thus providing a plausible molecular basis for the increased risk for and accelerated

progression of LOAD in females. We also identified correlations between LOAD-associated accessibility changes and dysregulation of gene expression within the same cell type. Lastly, we provide evidence suggesting that in some LOAD loci the precise disease-causing gene is not merely the nearest gene to LOAD SNPs.

## Abbreviations

LOAD: Late-onset Alzheimer's disease; GWAS: Genome-wide association study; eQTL: Expression quantitative trait loci; EWAS: Epigenome-wide association study; FANS: Fluorescence activated nuclei sorting; ATAC-seq: Assay for transposase-accessible chromatin using sequencing; KPBBB: Kathleen Price Bryan Brain Bank; SEM: Standard error of the mean; PMI: Post mortem interval; IRB: Institutional Review Board; PC: Principal component; TMM: Trimmed mean of M-values; QLF: Quasi-likelihood F-test; ROSMAP: Religious Orders Study and Memory and Aging Project; PCA: Principal Component Analysis; UMAP: Uniform manifold approximation and projection; HPCA: Human Primary Cell Atlas; EGR2: Early Growth Response 2; RFX1: Regulatory Factor X1; MEF2C: Myocyte Enhancer Factor 2C; SOX: SRY-related HMG-box; WT1: Wilms Tumor protein; Egr1: Early Growth Response 1; RAR: Retinoic acid receptor gamma; KLF14: Kruppel like factor 14

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13024-021-00481-0>.

**Additional file 1 : Supplementary Table 1.** Demographic and sample information of all 51 patients analyzed. **Supplementary Table 2.** Demographic and sample information of group of 40 paired, 9 unpaired. \*Donor tissue used for snRNA-seq. **Supplementary Table 3.** GREAT analysis of Top 5000 differential sites for Level 1 analysis. **Supplementary Table 4.** HOMER Motif search result of top 5000 sites more accessible in neuron and non-neuron. **Supplementary Table 5.** Demographic information for bulk ATAC-seq tissue samples. **Supplementary Table 6.** Differential analysis of Whole tissue vs neuron/non-neuron mixed. Terms in blue represent association with glial function and the ones in red represent association with neuronal function. Terms in green represent association with general brain function. GO terms enriched in DE sites of non-neuron vs whole tissue. **Supplementary Table 7.** GO terms enriched in DE sites of neuron vs whole tissue. **Supplementary Table 8.** GO terms enriched in sites more accessible in mixed neuron/non-neuron vs whole tissue. **Supplementary Table 9.** GO term enrichment of LOAD down differential sites in female glia controls vs. cases, FDR < 0.05. Terms in blue are associated with glial function. **Supplementary Table 10.** Level 1 (neuronal vs. non-neuronal) differentially accessible peaks. **Supplementary Table 11.** Level 2 (LOAD vs. Normal neuronal) differentially accessible peaks. **Supplementary Table 12.** Level 3 (LOAD vs. Normal non-neuronal female) differentially accessible peaks. **Supplementary Table 13.** Level 3b (LOAD vs. Normal non-neuronal female) differentially accessible peaks. **Supplementary Table 14.** Consistency between our snRNA-seq LOAD vs. Normal differential expression results and those of a recent publication. Our results are derived from HPCA-annotated neurons, astrocytes, and microglia. Dark orange = significantly up-regulated in LOAD, light orange = non-significantly up-regulated in LOAD, dark blue = significantly down-regulated in LOAD, light blue = non-significantly down-regulated in LOAD. **Supplementary Table 15.** Differential expression analysis of ROSMAP data for Level 3 female non-neuronal cells with and without covariates.

**Additional file 2 : Supplementary Fig. 1:** Pearson correlations of all potential numerical covariates and the first ten principal components of the potential numerical covariates. **Supplementary Fig. 2.** Association of meta variables with case-control status at a Bonferroni significance level of  $q < 0.05$ . The vertical line indicates Bonferroni significance. Linear regression was performed for numerical variables ( $n = 27$ ) and chi-square tests were conducted for categorical variables ( $n = 2$ , excluding diagnosis). **Supplementary Fig. 3.** Scree plot depicting the proportion of total

variance explained by each principal component of peaks in all samples passing QC. **Supplementary Fig. 4.** Level 1 comparison: Genomic distribution of Top 5000 differential sites of neuron/non-neuron. **Supplementary Fig. 5.** Fullard et al. vs. control samples from this study. (a) Venn diagram of differential sites from Fullard et al. publication and this study. Differential sites were generated by DESeq2 at fdr 0.01. Upper panels were analyzed by choosing the same numbers of samples from this study as those in Fullard et al. publication. (b) Peak length distribution of Fullard et al. 8 + 8, This study 8 + 8 and This study 26 + 21. **Supplementary**

**Fig. 6.** MA plots (FDR < 0.05) of Diff analysis of Whole Tissue vs neuron/non-neuron mixed ( $n = 6$  vs 6). (a) Comparison of sorted neuron vs whole issue. (b) Comparison of non-neuron vs whole brain tissue. (c) Comparison of 1:1 in silico mixed neuron/non-neuron cells control vs cases (no differential sites, dispersion = 0.23). (d) Overlap between DE sites. **Supplementary Fig. 7.** Box plots showing accessibility of peaks not differential between cases and controls for (a) CLU, PTK2B, (b) APOE, and (c) IQCK. The green transparent box indicates the regions for the controls in order. Box plots show the median, 25th percentile, and 75th percentile. Box plot whiskers show the 75th and 25th percentiles plus and minus 1.5 times the interquartile range, respectively. **Supplementary Fig. 8.** Permutation ( $n = 10,000$ ) controls for differential sites overlap with GWA LOAD sites. Red lines represent LOAD sites numbers overlapped by differential sites.

**a.** neuron LOAD up, randomly chosen 537 peak calls (pvalue = 0.13); **b.** neuron LOAD down, randomly chosen 947 sites (pvalue = 0.06); **c.** female non-neuron LOAD up, randomly chosen 1000 sites (pva = 0.98); **d.** female non-neuron LOAD down, randomly chosen 2352 sites (pva = 0.05). **Supplementary Fig. 9.** All genes within 1 Mb of LOAD-GWAS loci found to be more (top) or less (bottom) accessible by ATAC-seq of FANS-sorted nuclei. SNPs (light blue lines) were anchored in the center of the region and red boxes indicate genes found to be significantly dysregulated by snRNA-seq. Pseudogenes, RNA genes, and novel transcripts are excluded. Figures generated using the UCSC Genome Browser (<http://genome.ucsc.edu/>) GRCh38/hg38 assembly released December 2013. **Supplementary Fig. 10.** Clustering of 90 replicates for data selection. (a) Hierarchical clustering. (b) K-mean clustering ( $K = 2$ ). **Supplementary Fig. 11.** Scatterplot of PC1 vs. PC2 peaks showing association between PC1 and cell type ( $p = 5.11 \times 10^{-45}$ ). **Supplementary Fig. 12.** (a) Differential sites found in the female glia normal vs LOAD comparison (blue and red circles) were mapped to MA plots of female neuron normal vs LOAD comparison (black dots). For the female glia normal vs LOAD comparison, blue circles represent differential sites which were more accessible in LOAD while red circles represent differential sites which were less accessible in LOAD. (b) Differential sites found in 27 replicates of female glia normal vs LOAD comparison were mapped to log fold changes plot of 18 replicates of female glia normal vs LOAD comparison vs. 22 replicates of male glia normal vs LOAD comparison. Black dots represent sites shared by the above two comparisons. Blue circles represent differential sites which were corresponding to those in female glia normal vs LOAD and more open in LOAD. Red circles represent differential sites which were corresponding to those in female glia normal vs LOAD and less open in LOAD. The dashed line represents  $y = x$ . The solid line represents linear regression of all of data points.

## Acknowledgments

We thank the Kathleen Price Bryan Brain Bank at Duke University (funded by NIA AG028377) for providing us with the brain tissues, the Duke Cancer Institute Flow Cytometry Shared Resource for sorting, and the Duke Sequencing and Genomic Technologies Shared Resource for sequencing. We also thank John Ervin for his assistance in obtaining the brain samples with neuropathological and clinical data required for the study, and Lynn Martinek for providing training and technical assistance in using the cell sorter and in analyzing the flow data. This work used a high-performance computing facility partially supported by grant 2016-IDG-1013 ("HARDAC+: Reproducible HPC for Next-generation Genomics") from the North Carolina Biotechnology Center. The results published here are in part based on data obtained from the AD Knowledge Portal (<https://adknowledgeportal.synapse.org>). Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG36836,

and the Illinois Department of Public Health. Additional phenotypic data can be requested at [www.radc.rush.edu](http://www.radc.rush.edu).

#### Authors' contributions

All authors read and approved the final manuscript. G.E.C. and O.C.F. conceived of the presented idea. J.B. and O.C.F. acquired brain samples, designed the study sample, and interpreted metadata. J.B. and Y.Y. performed brain tissue handling, nuclei extraction, labeling and sorting, and fluorescence microscopy evaluation to ensure nuclei were of high quality. I.P., D.S., and D.C. provided technical support in extraction of bulk nuclei from brain tissues. A.S. performed the ATAC-seq experiments. J.B., Y.Y. and J.G. performed the snRNA-seq experiments. M.E.G. and A.A.K. performed QC and covariate analysis. L.S. performed alignment, peak quantification, QC assessment, differential chromatin, and motif analysis. R.G. advised on the motif analysis. J.G. and H.F. performed the snRNA-seq data analysis and differential expression analysis. J.L. assisted with the QC metric analysis. K.S. assisted with whole genome genotyping. A.A.K., G.E.C. and O.C.F. planned and supervised the work. G.E.C. and O.C.F. obtained funding support. All authors discussed and interpreted the results. J.B., L.S., A.S., Y.Y., M.E.G., J.G., A.A.K., G.E.C., and O.C.F. designed the study and wrote the manuscript.

#### Funding

This work was funded in part by the National Institutes of Health/National Institute on Aging (NIH/NIA) [R01 AG057522 to O.C.F.] and a seed grant from The Duke Center for Genomic and Computational Biology (to O.C.F. and GEC).

#### Availability of data and materials

The ATAC-Sequencing and single-nucleus RNA-Sequencing data are available at Synapse (<https://www.synapse.org/#!Synapse:syn20755767>). The DOI for this dataset is <https://doi.org/10.7303/syn20755767>. The data are available under controlled use conditions.

#### Declaration

##### Competing interests

The authors declare no competing interests.

##### Author details

<sup>1</sup>Department of Neurology, Division of Translational Brain Sciences, Duke University Medical Center, DUMC, Box 2900, Durham, NC 27710, USA. <sup>2</sup>Center for Genomic and Computational Biology, Duke University Medical Center, Durham, NC 27708, USA. <sup>3</sup>Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC 27701, USA. <sup>4</sup>Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC 27705, USA. <sup>5</sup>Department of Computer Science, Duke University, Durham, NC 27705, USA. <sup>6</sup>Department of Medicine, Duke University Medical Center, DUMC, Box 104775, Durham, NC 27708, USA. <sup>7</sup>Department of Pediatrics, Division of Medical Genetics, Duke University Medical Center, DUMC, Box 3382, Durham, NC 27708, USA. <sup>8</sup>Center for Advanced Genomic Technologies, Duke University Medical Center, Durham, NC 27708, USA.

Received: 14 June 2021 Accepted: 11 August 2021

Published online: 24 August 2021

#### References

- Lambert J-C, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45(12):1452–8. <https://doi.org/10.1038/ng.2802>.
- Naj AC, Jun G, Beecham GW, Wang LS, Vardarajan BN, Buros J, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet*. 2011;43(5):436–41.
- Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet*. 2009;41(10):1088–93. <https://doi.org/10.1038/ng.440>.
- Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet*. 2011;43(5):429–35. <https://doi.org/10.1038/ng.803>.
- Jansen I, Savage J, Watanabe K, Bryois J, Williams D, Steinberg S, et al. Genetic meta-analysis identifies 9 novel loci and functional pathways for Alzheimer's disease risk. *bioRxiv*. 2018;258533. <https://doi.org/10.1101/258533>.
- Marioni RE, Harris SE, Zhang Q, McRae AF, Hagenaars SP, Hill WD, et al. GWAS on family history of Alzheimer's disease. *Transl Psychiatry*. 2018;8(1):99. <https://doi.org/10.1038/s41398-018-0150-6>.
- Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet*. 2019;51(3):414–30. <https://doi.org/10.1038/s41588-019-0358-2>.
- Linnertz C, Anderson L, Gottschalk W, Crenshaw D, Lutz MW, Allen J, et al. The cis-regulatory effect of an Alzheimer's disease-associated poly-T locus on expression of TOMM40 and apolipoprotein E genes. *Alzheimers Dement*. 2014;10(5):541–51. <https://doi.org/10.1016/j.jalz.2013.08.280>.
- Matsui T, Ingelsson M, Fukumoto H, Ramasamy K, Kowa H, Frosch MP, et al. Expression of APP pathway mRNAs and proteins in Alzheimer's disease. *Brain Res*. 2007;1161:116–23. <https://doi.org/10.1016/j.brainres.2007.05.050>.
- Zarow C, Victoroff J. Increased apolipoprotein E mRNA in the hippocampus in Alzheimer disease and in rats after entorhinal cortex lesioning. *Exp Neurol*. 1998;149(1):79–86. <https://doi.org/10.1006/exnr.1997.6709>.
- Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet*. 2010;6(5):e1000952.
- Allen M, Zou F, Chai HS, Younkin CS, Crook J, Pankratz VS, et al. Novel late-onset Alzheimer disease loci variants associate with brain gene expression. *Neurology*. 2012;79(3):221–8. <https://doi.org/10.1212/WNL.0b013e318605801>.
- Zou F, Chai HS, Younkin CS, Allen M, Crook J, Pankratz VS, et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet*. 2012;8(6):e1002707. <https://doi.org/10.1371/journal.pgen.1002707>.
- Karch CM, Jeng AT, Nowotny P, Cady J, Cruchaga C, Goate AM. Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains. *PLoS One*. 2012;7(11):e50976. <https://doi.org/10.1371/journal.pone.0050976>.
- Karch CM, Ezerskiy LA, Bertelsen S, Alzheimer's Disease Genetics C, Goate AM. Alzheimer's Disease Risk Polymorphisms Regulate Gene Expression in the ZCWPW1 and the CELF1 Loci. *PLoS One*. 2016;11(2):e0148717.
- Singleton A, Myers A, Hardy J. The law of mass action applied to neurodegenerative disease: a hypothesis concerning the etiology and pathogenesis of complex diseases. *Hum Mol Genet*. 2004;13 Spec No 1: R123–6.
- Smith AR, Smith RG, Condliffe D, Hannon E, Schalkwyk L, Mill J, et al. Increased DNA methylation near TREM2 is consistently seen in the superior temporal gyrus in Alzheimer's disease brain. *Neurobiol Aging*. 2016;47:35–40. <https://doi.org/10.1016/j.neurobiolaging.2016.07.008>.
- Zhao J, Zhu Y, Yang J, Li L, Wu H, De Jager PL, et al. A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's disease. *Alzheimers Dement*. 2017;13(6):674–88. <https://doi.org/10.1016/j.jalz.2016.10.004>.
- Lunnon K, Smith R, Hannon E, De Jager PL, Srivastava G, Volta M, et al. Methylocomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci*. 2014;17(9):1164–70. <https://doi.org/10.1038/nn.3782>.
- De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci*. 2014;17(9):1156–63. <https://doi.org/10.1038/nn.3786>.
- Chibnik LB, Yu L, Eaton ML, Srivastava G, Schneider JA, Kellis M, et al. Alzheimer's loci: epigenetic associations and interaction with genetic factors. *Ann Clin Transl Neurol*. 2015;2(6):636–47. <https://doi.org/10.1002/acn3.201>.
- Yu L, Chibnik LB, Srivastava GP, Pochet N, Yang J, Xu J, et al. Association of Brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease. *JAMA Neurol*. 2015;72(1):15–24. <https://doi.org/10.1001/jamaneurol.2014.3049>.
- Watson CT, Roussos P, Garg P, Ho DJ, Azam N, Katsel PL, et al. Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. *Genome Med*. 2016;8(1):5. <https://doi.org/10.1186/s13073-015-0258-8>.

24. Nativio R, Donahue G, Berson A, Lan Y, Amlie-Wolf A, Tuzer F, et al. Dysregulation of the epigenetic landscape of normal aging in Alzheimer's disease. *Nat Neurosci.* 2018;21(4):497–505. <https://doi.org/10.1038/s41593-018-0101-9>.
25. Matevossian A, Akbarian S. Neuronal nuclei isolation from human postmortem brain tissue. *J Vis Exp.* 2008;20.
26. Gasparoni G, Bultmann S, Lutsik P, Kraus TFJ, Sordon S, Vlcek J, et al. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. *Epigenetics Chromatin.* 2018;11(1):41. <https://doi.org/10.1186/s13072-018-0211-3>.
27. Tulloch J, Leong L, Thomson Z, Chen S, Lee EG, Keene CD, et al. Glia-specific APOE epigenetic changes in the Alzheimer's disease brain. *Brain Res.* 2018;1698:179–86.
28. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature.* 2019; 570(7761):332–7. <https://doi.org/10.1038/s41586-019-1195-2>.
29. Grubman A, Chew G, Ouyang JF, Sun G, Choo XY, McLean C, et al. A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat Neurosci.* 2019; 22(12):2087–97. <https://doi.org/10.1038/s41593-019-0539-4>.
30. Jiang Y, Matevossian A, Huang HS, Straubhaar J, Akbarian S. Isolation of neuronal chromatin from brain tissue. *BMC Neurosci.* 2008;9(1):42. <https://doi.org/10.1186/1471-2202-9-42>.
31. Marluff WF. Preparation of active nuclei. *Methods Enzymol.* 1990;181:30–6. [https://doi.org/10.1016/0076-6879\(90\)81109-8](https://doi.org/10.1016/0076-6879(90)81109-8).
32. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods.* 2017;14(10):959–62. <https://doi.org/10.1038/nmeth.4396>.
33. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
34. Fein FS, Cho S, Zola BE, Miller B, Factor SM. Cardiac pathology in the hypertensive diabetic rat. Biventricular damage with right ventricular predominance. *Am J Pathol.* 1989;134(5):1159–66.
35. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 2012;22(9):1813–31. <https://doi.org/10.1101/gr.136184.111>.
36. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw.* 2011;45(3):1–67.
37. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616>.
38. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010;11(3):R25. <https://doi.org/10.1186/gb-2010-11-3-r25>.
39. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38(4):576–89. <https://doi.org/10.1016/j.molcel.2010.05.004>.
40. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. *Nucleic Acids Res.* 2015;43(W1):W39–49. <https://doi.org/10.1093/nar/gkv416>.
41. Bennett DA, Schneider JA, Arvanitakis Z, Wilson RS. Overview and findings from the religious orders study. *Curr Alzheimer Res.* 2012;9(6):628–45. <https://doi.org/10.2174/156720512801322573>.
42. Bennett DA, Schneider JA, Buchman AS, Barnes LL, Boyle PA, Wilson RS. Overview and findings from the rush memory and aging project. *Curr Alzheimer Res.* 2012;9(6):646–63. <https://doi.org/10.2174/156720512801322663>.
43. Mostafavi S, Gaiteri C, Sullivan SE, White CC, Tasaki S, Xu J, et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci.* 2018;21(6):811–9. <https://doi.org/10.1038/s41593-018-0154-9>.
44. SageBionetworks. AMP-AD Knowledge Portal: ROSMAP RNA-Seq. Available at: <https://www.synapse.org/#Synapse:syn3388564>.
45. McCarthy DJ, Campbell KR, Lun AT, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics.* 2017;33(8):1179–86. <https://doi.org/10.1093/bioinformatics/btw777>.
46. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive integration of single-cell data. *Cell.* 2019;177(7):1888–902 e21. <https://doi.org/10.1016/j.cell.2019.05.031>.
47. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol.* 2019;20(2):163–72. <https://doi.org/10.1038/s41901-018-0276-y>.
48. Mabbott NA, Bailie JK, Brown H, Freeman TC, Hume DA. An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics.* 2013;14(1):632. <https://doi.org/10.1186/1471-2164-14-632>.
49. Garwood CJ, Ratcliffe LE, Simpson JE, Heath PR, Ince PG, Wharton SB. Review: astrocytes in Alzheimer's disease and other age-associated dementias: a supporting player with a central role. *Neuropathol Appl Neurobiol.* 2017;43(4):281–98. <https://doi.org/10.1111/nan.12338>.
50. Perl DP. Neuropathology of Alzheimer's disease. *Mt Sinai J Med.* 2010;77(1):32–42. <https://doi.org/10.1002/msj.20157>.
51. McLean CY, Bristol D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 2010;28(5):495–501. <https://doi.org/10.1038/nbt.1630>.
52. Nagarajan R, Svaren J, Le N, Araki T, Watson M, Milbrandt J. EGR2 mutations in inherited neuropathies dominant-negatively inhibit myelin gene expression. *Neuron.* 2001;30(2):355–68. [https://doi.org/10.1016/S0896-6273\(01\)00282-3](https://doi.org/10.1016/S0896-6273(01)00282-3).
53. Ma K, Zheng S, Zuo Z. The transcription factor regulatory factor X1 increases the expression of neuronal glutamate transporter type 3. *J Biol Chem.* 2006; 281(30):21250–5. <https://doi.org/10.1074/jbc.M600521200>.
54. Harrington AJ, Raissi A, Rajkovich K, Berto S, Kumar J, Molinaro G, et al. MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. *Elife.* 2016;5. <https://doi.org/10.7554/elife.20059>.
55. Wegner M. SOX after SOX: SOXession regulates neurogenesis. *Genes Dev.* 2011;25(23):2423–8. <https://doi.org/10.1101/gad.181487.111>.
56. Fullard JF, Giambartolomei C, Hauberg ME, Xu K, Voloudakis G, Shao Z, et al. Open chromatin profiling of human postmortem brain infers functional roles for non-coding schizophrenia loci. *Hum Mol Genet.* 2017;26(10):1942–51. <https://doi.org/10.1093/hmg/ddx103>.
57. Wagner N, Wagner KD, Schley G, Coupland SE, Heimann H, Grantyn R, et al. The Wilms' tumor suppressor Wt1 is associated with the differentiation of retinoblastoma cells. *Cell Growth Differ.* 2002;13(7):297–305.
58. Shi Z, Shen T, Liu Y, Huang Y, Jiao J. Retinoic acid receptor gamma (Rarg) and nuclear receptor subfamily 5, group A, member 2 (Nr5a2) promote conversion of fibroblasts to functional neurons. *J Biol Chem.* 2014;289(10): 6415–28. <https://doi.org/10.1074/jbc.M113.515601>.
59. Ductol F, Kabaj M. The role of early growth response 1 (Egr1) in brain plasticity and neuropsychiatric disorders. *Front Behav Neurosci.* 2017;11:35.
60. Parker-Katiraie L, Carson AR, Yamada T, Arnaud P, Feil R, Abu-Amero SN, et al. Identification of the imprinted KLF14 transcription factor undergoing human-specific accelerated evolution. *PLoS Genet.* 2007;3(5):e65. <https://doi.org/10.1371/journal.pgen.0030065>.
61. Riedel BC, Thompson PM, Brinton RD. Age, APOE and sex: triad of risk of Alzheimer's disease. *J Steroid Biochem Mol Biol.* 2016;160:134–47. <https://doi.org/10.1016/j.jsmb.2016.03.012>.
62. Poche RA, Zhang M, Rueda EM, Tong X, McElwee ML, Wong L, et al. RONIN is an essential transcriptional regulator of genes required for mitochondrial function in the developing retina. *Cell Rep.* 2016;14(7):1684–97. <https://doi.org/10.1016/j.celrep.2016.01.039>.
63. Vong Kl, Leung CK, Behringer RR, Kwan KM. Sox9 is critical for suppression of neurogenesis but not initiation of gliogenesis in the cerebellum. *Mol Brain.* 2015;8(1):25. <https://doi.org/10.1186/s13041-015-0115-0>.
64. Kolsch H, Jessen F, Wiltfang J, Lewczuk P, Dichgans M, Kornhuber J, et al. Influence of SORL1 gene variants: association with CSF amyloid-beta products in probable Alzheimer's disease. *Neurosci Lett.* 2008;440(1):68–71. <https://doi.org/10.1016/j.neulet.2008.05.049>.
65. Demir O, Aysit N, Onder Z, Turkel N, Ozturk G, Sharrocks AD, et al. ETS-domain transcription factor Elk-1 mediates neuronal survival: SMN as a potential target. *Biochim Biophys Acta.* 2011;1812(6):652–62. <https://doi.org/10.1016/j.bbadi.2011.02.012>.
66. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res.* 2002;12(6):996–1006. <https://doi.org/10.1101/gr.229102>.
67. Trynka G, Raychaudhuri S. Using chromatin marks to interpret and localize genetic associations to complex human traits and diseases. *Curr Opin Genet Dev.* 2013;23(6):635–41. <https://doi.org/10.1016/j.gde.2013.10.009>.

68. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet.* 2013;45(2):124–30. <https://doi.org/10.1038/ng.2504>.
69. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012;337(6099):1190–5. <https://doi.org/10.1126/science.1222794>.
70. Cockerill PN. Structure and function of active chromatin and DNase I hypersensitive sites. *FEBS J.* 2011;278(13):2182–210. <https://doi.org/10.1111/j.1742-4658.2011.08128.x>.
71. Gross DS, Garrard WT. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem.* 1988;57(1):159–97. <https://doi.org/10.1146/annurev.bi.57.070188.001111>.
72. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee BK, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* 2011;21(10):1757–67. <https://doi.org/10.1101/gr.121541.111>.
73. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature.* 2012;489(7414):75–82. <https://doi.org/10.1038/nature11232>.
74. Won H, de la Torre-Ubieta L, Stein JL, Parikhshak NN, Huang J, Opland CK, et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature.* 2016;538(7626):523–7. <https://doi.org/10.1038/nature19847>.
75. Ramamurthy E, Welch G, Cheng J, Yuan Y, Gunsalus L, Bennett DA, et al. Cell type-specific histone acetylation profiling of Alzheimer's Disease subjects and integration with genetics. *bioRxiv.* 2020;2020.03.26.010330. <https://doi.org/10.1101/2020.03.26.010330>.
76. Bendl J, Hauberg ME, Girdhar K, Im E, Vicari JM, Rahman S, et al. The three-dimensional landscape of chromatin accessibility in Alzheimer's disease. *bioRxiv.* 2021;2021.01.11.426303. <https://doi.org/10.1101/2021.01.11.426303>.
77. Gjoneska E, Pfenning AR, Mathys H, Quon G, Kundaje A, Tsai LH, et al. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature.* 2015;518(7539):365–9. <https://doi.org/10.1038/nature14252>.
78. Morabito S, Miyoshi E, Michael N, Swarup V. Hum Mol Genet: Integrative genomics approach identifies conserved transcriptomic networks in Alzheimer's disease; 2020.
79. Nott A, Holtzman IR, Coufal NG, Schlachetzki JCM, Yu M, Hu R, et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science.* 2019;366(6469):1134–9. <https://doi.org/10.1126/science.aay0793>.
80. Ferretti MT, Iulita MF, Cavedo E, Chiesa PA, Schumacher Dimech A, Santuccione Chadha A, et al. Sex differences in Alzheimer disease - the gateway to precision medicine. *Nat Rev Neurol.* 2018;14(8):457–69. <https://doi.org/10.1038/s41582-018-0032-9>.
81. Babapour Mofrad R, van der Flier WM. Nature and implications of sex differences in AD pathology. *Nat Rev Neurol.* 2019;15(1):6–8. <https://doi.org/10.1038/s41582-018-0115-7>.
82. Buckley RF, Waller M, Masters CL, Dobson A. To what extent does age at death account for sex differences in rates of mortality from Alzheimer disease? *Am J Epidemiol.* 2019;188(7):1213–23. <https://doi.org/10.1093/aje/kwz048>.
83. Jack CR Jr, Wiste HJ, Weigand SD, Knopman DS, Vemuri P, Mielke MM, et al. Age, sex, and APOE epsilon4 effects on memory, brain structure, and beta-amyloid across the adult life span. *JAMA Neurol.* 2015;72(5):511–9. <https://doi.org/10.1001/jamaneurol.2014.4821>.
84. Corder EH, Saunders AM, Strittmatter WJ, Schmeichel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science.* 1993;261(5123):921–3. <https://doi.org/10.1126/science.8346443>.
85. Saunders AM, Strittmatter WJ, Schmeichel D, George-Hyslop PH, Pericak-Vance MA, Joo SH, et al. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology.* 1993;43(8):1467–72. <https://doi.org/10.1212/WNL.43.8.1467>.
86. Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet.* 2009;41(10):1094–9. <https://doi.org/10.1038/ng.439>.
87. Bretsky PM, Buckwalter JG, Seeman TE, Miller CA, Poirier J, Schellenberg GD, et al. Evidence for an interaction between apolipoprotein E genotype, gender, and Alzheimer disease. *Alzheimer Dis Assoc Disord.* 1999;13(4):216–21. <https://doi.org/10.1097/00002093-199910000-00007>.
88. Payami H, Montee KR, Kaye JA, Bird TD, Yu CE, Wijsman EM, et al. Alzheimer's disease, apolipoprotein E4, and gender. *JAMA.* 1994;271(17):1316–7. <https://doi.org/10.1001/jama.1994.03510410028015>.
89. Poirier J, Davignon J, Bouthillier D, Kogan S, Bertrand P, Gauthier S. Apolipoprotein E polymorphism and Alzheimer's disease. *Lancet.* 1993;342(8873):697–9. [https://doi.org/10.1016/0140-6736\(93\)91705-Q](https://doi.org/10.1016/0140-6736(93)91705-Q).
90. Altmann A, Tian L, Henderson VW, Greicius MD. Alzheimer's disease neuroimaging initiative I. sex modifies the APOE-related risk of developing Alzheimer disease. *Ann Neurol.* 2014;75(4):563–73. <https://doi.org/10.1002/ana.24135>.
91. Neu SC, Pa J, Kukull W, Beekly D, Kuzma A, Gangadharan P, et al. Apolipoprotein E genotype and sex risk factors for Alzheimer disease: a Meta-analysis. *JAMA Neurol.* 2017;74(10):1178–89. <https://doi.org/10.1001/jamaneurol.2017.2188>.
92. Gamache J, Yun Y, Chiba-Falek O. Sex-dependent effect of APOE on Alzheimer's disease and other age-related neurodegenerative disorders. *Dis Model Mech.* 2020;13(8):dmm045211. <https://doi.org/10.1242/dmm.045211>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](http://biomedcentral.com/submissions)

