1　**Sex dependent glial-specific changes in the chromatin accessibility landscape in late-onset**
2　**Alzheimer's disease brains**
3

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

9　[1] Division of Translational Brain Sciences, Department of Neurology, Duke University Medical Center, Durham,
10　NC, 27710, USA.
11　[2] Center for Genomic and Computational Biology, Duke University Medical Center, Durham, NC, 27708, USA.
12　[3] Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC, 27701, USA.
13　[4] Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC, 27705, USA.
14　[5] Department of Computer Science, Duke University, Durham, NC, 27705, USA.
15　[6] Department of Medicine, Duke University Medical Center, Durham, NC, 27708, USA.
16　[7] Department of Pediatrics, Division of Medical Genetics, Duke University Medical Center, Durham, NC, 27708
17　[8] Center for Advanced Genomic Technologies, Duke University Medical Center, Durham, NC, 27708, USA.
18

19　[#] These authors contributed equally to this work
20

21　* To whom correspondence should be addressed:
22

23　Ornit Chiba-Falek
24　Dept of Neurology
25　DUMC Box 2900
26　Duke University Medical Center,
27　Durham, North Carolina 27710, USA
28　Tel: 919 681-8001
29　Fax: 919 684-6514
30　E-mail: o.chibafalek@duke.edu
31

32　Gregory E. Crawford
33　Dept of Pediatrics
34　DUMC Box 3382
35　Duke University Medical Center,
36　Durham, North Carolina 27708, USA.
37　Tel: 919 684-8196
38　Fax: 919 681-8973
39　E-mail: greg.crawford@duke.edu
40

41　Allison E. Ashley-Koch
42　Dept of Medicine
43　DUMC Box 104775
44　Duke University Medical Center,
45　Durham, North Carolina 27701, USA
46　Tel: 919 684-1805
47　E-mail: allison.ashleykoch@duke.edu
48

49 **Abstract:**

51 In the post-GWAS era, there is an unmet need to decode the underpinning genetic etiologies of

52 late-onset Alzheimer's disease (LOAD) and translate the associations to causation. Toward that

53 goal, we conducted ATAC-seq profiling using neuronal nuclear protein (NeuN) sorted-nuclei

54 from 40 frozen brain tissues to determine LOAD-specific changes in chromatin accessibility

55 landscape in a cell-type specific manner. We identified 211 LOAD-specific differential

56 chromatin accessibility sites in neuronal-nuclei, four of which overlapped with LOAD-GWAS

57 regions ($\pm$100kb of SNP). While the non-neuronal nuclei did not show LOAD-specific

58 differences, stratification by sex identified 842 LOAD-specific chromatin accessibility sites in

59 females. Seven of these sex-dependent sites in the non-neuronal samples overlapped LOAD-

60 GWAS regions including *APOE*. LOAD loci were functionally validated using single-nuclei

61 RNA-seq datasets. In conclusion, using brain sorted-nuclei enabled the identification of sex-

62 dependent cell type-specific LOAD alterations in chromatin structure. These findings

63 enhance the interpretation of LOAD-GWAS discoveries, provide potential pathomechanisms,

64 and suggest novel LOAD-loci. Furthermore, our results convey mechanistic insights into sex

65 differences in LOAD risk and clinicopathology.

66

67 Keywords: ATAC-seq; chromatin accessibility; nuclei sorting; late-onset Alzheimer's disease;

68 snRNA-seq; gene dysregulation

69 **GRAPHICAL ABSTRACT**

70

71

72

73 **Introduction:**

74
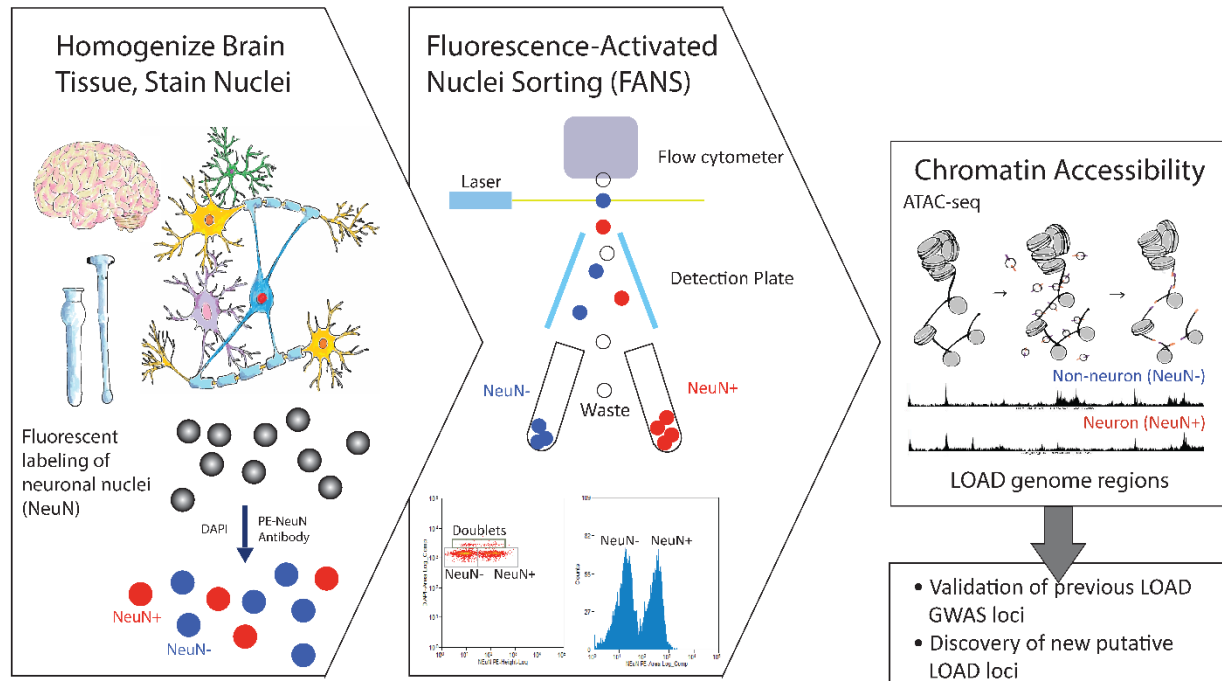
75 Large multi-center genome-wide association studies (GWAS) have identified associations

76 between numerous genomic loci and late-onset Alzheimer's disease (LOAD)[1-6]. The most recent

77 GWAS meta-analysis reported a total of 25 LOAD-GWAS regions[7]. However, the precise

78 disease-responsible genes, the specific causal genetic variants, and the molecular mechanisms of

79 action that mediate their pathogenic effects are yet to be explained. Most LOAD-GWAS SNPs

80 are in noncoding (intergenic and intragenic) genomic regions, and thus may have a gene

81 regulatory function. Supporting this hypothesis, differential gene expression in brain tissues have

82 been described between LOAD and healthy controls [8,9, 10]. Moreover, several expression

83 quantitative trait loci (eQTL) studies in brain tissues from cognitively normal[11] and LOAD[12-15]

84 individuals reported overlaps with LOAD-GWAS loci. Collectively, these observations suggest

85 that dysregulation of gene expression plays an important role in LOAD pathogenesis[16].

3

86   Noteworthy, integration of findings from LOAD epigenome wide association studies (EWAS)

87   and GWAS also identified a number of shared loci[17-24], providing further support for the role of

88   gene dysregulation in LOAD pathogenesis.

89   To date, most brain expression, DNA-methylation, and chromatin studies have used brain

90   tissue homogenates that represent multiple cell-types, *i.e.* various types of neurons and types of

91   glia cells. The heterogeneity of brain tissue makes it difficult to determine the specific brain cell-

92   type responsible for the changes in gene expression and in epigenome landscape. The mixture of

93   cell-types could potentially mask signals corresponding to a particular cell-type, especially if the

94   causal cell types comprise a small fraction of the entire sample. An additional shortcoming of

95   studying bulk brain tissues is the bias related to sample-to-sample differences in the cell-type

96   composition of the tissue. The obstacle of variability in the neuron:glia proportions across

97   samples is even more pronounced when analyzing disease-affected brain tissues that underwent

98   the neurodegeneration processes of neuronal loss and gliosis. Single cell experimental

99   approaches can circumvent these limitations; however, these methods are costly for studying

100   larger sample sizes and have been underutilized in the field of LOAD genetics. Frozen tissues

101   pose additional technical challenges as it is difficult to isolate intact cell bodies. Fluorescence

102   Activated Nuclei Sorting (FANS) was developed to extract, purify and sort nuclei (vs. cells)

103   from frozen brain tissues[25] using nuclei neuronal markers, such as NeuN, to greatly reduce

104   cellular heterogeneity found in bulk tissues, and allow characterization of neuronal (NeuN+) and

105   non-neuronal (NeuN-) cell populations. Recently, two studies used FANS to analyze LOAD-

106   specific changes in DNA-methylation on a whole-genome level[26] and across the *APOE* locus[27].

107   Furthermore, two new studies used single-cell (sc)RNA-seq from cortex of LOAD patients. The

108   first found that the strongest LOAD-associated changes appeared early in pathological

4

109 progression and were highly cell-type specific[28], and the second identified LOAD-associated

110 gene dysregulation in specific cell subpopulations, particularly for *APOE* and transcription

111 factors[29]. These results further highlight the importance of cell-type specific analysis of human

112 brain tissues to understanding of the molecular underpinnings and cellular basis of LOAD.

113   In this study, we performed NeuN-FANS from 40 archived frozen human brain samples

114 (19 LOAD, 21 normal), and used the assay for transposase-accessible chromatin using

115 sequencing (ATAC-seq) to characterize the chromatin accessibility landscape in neuronal and

116 non-neuronal nuclei. We identified over 170,000 chromatin accessibility differences between

117 neuronal and non-neuronal nuclei. We also report LOAD-specific differences in chromatin

118 accessibility in both neurons and non-neurons. Interestingly, while the neuronal changes

119 appeared to be independent of sex, in the non-neuronal cells LOAD differences in chromatin

120 accessibility were detected only in females. LOAD-specific differences in chromatin

121 accessibility significantly overlap with known LOAD GWAS regions, and also point to new

122 candidate LOAD loci. These results provide new insights into the mechanistic and sex-specific

123 pathogenesis of this disorder.

124

125 **Material and Methods:**

126 Human Brain Tissue Samples

127 Following quality-control filtering, the final dataset was generated using fresh-frozen temporal

128 cortex from neurologically healthy controls (n=21), mild LOAD (n=16), and severe LOAD

129 (n=3) patients. These samples were obtained from the Kathleen Price Bryan Brain Bank

130 (KPBBB) at Duke University, and the demographics for this cohort are included in **Table 1** and

131 detailed in **Table S2**. Clinical diagnosis of LOAD was pathologically confirmed using Braak

132   staging (AT8 immunostaining) and amyloid deposition assessment (4G8 immunostaining) for all

133   LOAD samples. Braak staging was used to define mild (stages IV and below) vs. severe (stages

134   V and VI) LOAD. All donors are Caucasians with APOE 33; post mortem interval (PMI)

135   averaged 7.15 hours (standard error of the mean (SEM) 0.81). The archived frozen tissues have

136   high-quality DNA as required for genomic analyses, and RNA (RIN≥8) suitable for quantitative

137   RNA analyses. The project was approved by the Duke Institutional Review Board (IRB). The

138   methods described were carried out in accordance with the relevant guidelines and regulations.

139   Tissue samples were processed in random pairs of one normal and one LOAD patient. Tissue

140   homogenization, nuclei extraction, FANS, and tagmentation were performed on each pair on the

141   same day. Library preparation and sequencing was performed blinded to age, sex, and pathology.

142   Out of this group of samples four female donors were analyzed by snRNA-seq, two of which

143   were neurologically healthy controls and two age matched mild LOAD patients (Braak Stage

144   III). These four samples were ages 79-90 with PMI averaged 8.02 hours (SEM=1.99) (**Table S2,**

145   **marked in *).** Tissue samples were processed for snRNA-seq on the same day and in the same

146   10X Genomics microfluidics chip.

147   <u>Tissue dissociation and nuclei extraction</u>

148   Methods were performed according to established protocols[25, 73, 74] with some modifications.

149   Briefly, 50 mg of frozen temporal cortex (gray matter) was thawed for 10 minutes on ice in lysis

150   buffer (0.32M Sucrose, 5 mM CaCl2, 3 mM magnesium acetate, 0.1 mM EDTA, 10 mM Tris-

151   HCl pH8, 1 mM DTT, 0.1% Triton X-100). The tissue was gently dissociated and homogenized

152   in a 7 ml dounce tissue homogenizer (Corning) with approximately 25 strokes of pestle A in 45

153   seconds, then filtered through a 100 μm cell strainer. The filtered lysate was transferred to a 14 x

154   89 mm polypropylene ultracentrifuge tube, carefully underlaid with sucrose solution (1.8 M

155   sucrose, 3 mM magnesium acetate, 1 mM DTT, 10 mM Tris-HCl, pH 8) and subjected to

156   ultracentrifugation at 107,000 RCF for approximately 30 minutes at 4°C. Supernatant and the

157   debris interphase were carefully aspirated, and 100 ul PBS (-Mg2+, -Ca2+) was added to the

158   nuclei pellet. After a 5-minute incubation on ice, nuclei were gently resuspended and transferred

159   to a 1.5 ml polypropylene microcentrifuge tube for staining.

160   For snRNA-seq downstream experiment, the above protocol was modified to optimize

161   sample preparation for single-nuclei sorting. Briefly, 50mg frozen temporal cortex (gray matter)

162   was thawed for 20 minutes on ice in lysis buffer and ultracentrifuged at 107,000 RCF for

163   approximately 10 minutes at 4°C. Supernatant and the debris interphase were carefully aspirated,

164   and 500 ul wash and resuspension buffer (1X PBS, 1% BSA, 0.2U/ul RNase Inhibitor) was

165   added to the nuclei pellet. After a 5-minute incubation on ice, nuclei were gently resuspended

166   and centrifuged at 300 RCF for 5 minutes at 4°C. The supernatant was again aspirated and 500 ul

167   wash and resuspension buffer without RNase Inhibitor (1X PBS, 1% BSA) was added to the

168   nuclei pellet. After a 1-minute incubation on ice, the nuclei were gently resuspended and filtered

169   through a 35 um strainer. 10 ul nuclei were taken for quality assessment and counting prior to

170   library preparation.

171   <u>Immunostaining of nuclei</u>

172   Nuclei were stained in 0.05% BSA, 1% normal goat serum, DAPI (1µg/ml), and PE-conjugated

173   anti-NeuN antibody (1:125, Millipore FCMAB317PE) in PBS (-Mg2+, -Ca2+), in the dark for

174   30 minutes at 4 °C. A DAPI-only control was prepared to set gates for sorting. After staining,

175   nuclei were filtered through a 40 um cell strainer into a polypropylene round-bottom 5ml tube

176   and sorted.

177   <u>Immunofluorescence microscopy</u>

178   After homogenization and sucrose gradient ultracentrifugation, a portion of the nuclei was

179   counted, resuspended in 4% PFA, stained, plated on 12 mm coverslips at 10,000 nuclei per

180   coverslip, incubated 20 minutes at room temperature, mounted, and imaged on a confocal

181   microscope.

182   <u>Fluorescence-activated nuclei sorting (FANS) of neuronal and non-neuronal nuclei</u>

183   Sorting was performed using a MoFlo Astrios flow cytometer (Beckman Coulter) equipped with

184   a 70μm nozzle, operating at 35 psi. Standard gating procedures were used. Briefly, the first gate

185   allowed separation of intact nuclei from debris. The second gate allowed us to identify individual

186   nuclei, and exclude doublets and other aggregates. The third and fourth gates distinguish

187   between PE+ and PE- nuclei and allowed us to sort and separate NeuN+ nuclei from NeuN-

188   nuclei. Nuclei were sorted into 1 ml PBS (-Mg2+, -Ca2+) in a 2 ml polypropylene tube pre-

189   coated with 200 ul of 5% BSA and rotated at 20 rpm at 4C.

190   <u>Omni-ATAC on FANS nuclei</u>

191   Approximately 100,000-700,000 sorted nuclei **(Table S2)** were used for ATAC-seq library

192   preparation as described in the Omni-ATAC protocol[75]. Libraries were quantified by Qubit, and

193   size distribution was inspected by Bioanalyzer (Agilent Genomic DNA chip, Agilent

194   Technologies). Barcoded ATAC-seq libraries were combined into pools of 6 libraries and

195   sequenced on an Illumina HiSeq 4000 sequencer (50 bp, single read) at the Duke Sequencing and

196   Genomic Technologies shared resource.

197   <u>Omni-ATAC-seq on bulk tissue</u>

198   In addition to performing ATAC-seq on NeuN+/- sorted nuclei, we also compared to ATAC-seq

199   performed on total nuclei isolated from frozen tissue. Approximately 50,000 nuclei from 25 mg

200   of pulverized frozen tissue were used for the transposition reaction applying the Omni-ATAC

201   protocol as described previously [75].

202   <u>Data processing pipeline</u>

203   ATAC-seq libraries made from 83 glia samples and 80 neuron samples were sequenced on Hi-

204   seq 4000. Raw fastq sequencing files were first processed through cutadapt (v 1.9.1) to remove

205   adaptors and bases with quality scores < 30. Filtered reads were aligned to hg19 by Bowtie2 (v

206   2.1.0) using default parameters[76]. Bam files were sorted by samtools (v 0.1.18) and duplicates

207   were removed by Picard MarkDuplicates. Sequences that overlapped ENCODE "blacklist"

208   regions (https://sites.google.com/site/anshulkundaje/projects/blacklists) were removed, and

209   narrow open chromatin peaks were called by Model-based Analysis of ChIP-seq (MACS v 2.1),

210   with parameters --nomodel --shift -100 --ext 200, using a FDR cutoff of q < 0.01 or 0.05[77]. For

211   visualization on the UCSC browser, BigWig files were generated using wigToBigWig (v 4).

212   <u>Data quality control</u>

213   We characterized the quality of all ATAC-seq datasets by the following metrics, based on

214   ENCODE suggestions[78] https://genome.ucsc.edu/ENCODE/analysis.html **(Table S1)**: (1) total

215   numbers of reads, (2) numbers of reads trimmed by cutadapt, (3) total bases entering cutadapt,

216   (4) quality-trimmed bases by cutadapt, (5) total numbers of reads processed by Bowtie2, (6)

217   uniquely mappable reads, (7) percentages of alignment, (8) read aligned to the mitochondrial

218   chromosome, (9) peak calls at FDR q < 0.05, (10) peak calls normalized by sequencing depth,

219   (11) GC content of sequences, (12) Non-Redundant Fraction (NRF), which is equal to the

220   number of distinct uniquely mapped reads divided by total reads, (13) PCR Bottleneck

221   Coefficient 1 (PBC1), and (14) PCR Bottleneck Coefficient 2 (PBC2). To reduce the impact of

222   samples with lower quality, we removed datasets with normalized peak calls at FDR q < 0.01 <

9

223   100, and any additional replicates that displayed lower signal-to-noise ratios. The 90 remaining

224   datasets were analyzed by PCA and hierarchical clustering (hclust function, method="ward.D"),

225   of which one NeuN- sample was identified as an outlier **(Fig. S10)**. After excluding this outlier,

226   we obtained a final set of 89 datasets from 49 donors (40 donors with matching NeuN+ and

227   NeuN- data, and 9 donors that only contained NeuN- data). Sequencing and QC metrics for all

228   89 samples are described in **Table S2**. The library complexity for these 89 samples are

229   comparable to ENCODE Data Quality Metrics spreadsheet published on 2012-04-25

230   (https://genome.ucsc.edu/ENCODE)

231   Evaluation of variables affecting chromatin peaks. Prior to the differential analyses described

232   below, we considered the effect of 37 variables that could impact the quality of the ATAC-seq

233   results for each sample. These variables were collected from key features of the ATAC-seq

234   processing pipeline, as well as individual sample characteristics such as case-control status, sex,

235   and age. For example, the metadata for each sample included transposase batch, nuclei sorting

236   date, mass, mean GC percentage of sequenced reads, mean mapped read length, alignment

237   quality metrics, subject age at death, sex, diagnosis, and postmortem interval (PMI). Two

238   subjects were missing PMI, therefore we used the R package MICE[79] to impute missing values

239   using the classification and regression trees methodology.

240   Covariate selection

241   For differential peak analyses, selection of covariates for adjustment was carefully tailored to

242   each comparative analysis to account for the variable number of peaks between groups and to

243   minimize false positives in peak calling. We performed a linear regression of all metadata

244   variables against the first 10 principal components (PCs) of the TMM normalized peak

245   quantification in an effort to identify covariates to be utilized in the differential expression

246   analyses. In an iterative process, we selected one variable (preferentially a variable directly

247   related to the ATAC-seq experiment, explaining one of the largest proportions of variance and

248   with few parameters), regressed its effect on the peak quantifications and performed a new

249   principal component analysis independent of the selected variable(s). We repeated this procedure

250   until there were no more Bonferroni significant (q<0.05) variables associated with the peak PCs.

251   For the comparison of neuronal and non-neuronal nuclei, we selected the following variables for

252   our differential chromatin analysis: NRF, number of nuclei, normalized peak calls, nuclei sort

253   date, alignment percentage, passing base pairs, age, and donor. For the comparison of LOAD vs.

254   control neuronal nuclei, we selected normalized peak calls, alignment percentage, nuclei sort

255   date, and number of nuclei as covariates. For the comparison of LOAD vs. control non-neuronal

256   nuclei, we included the number of nuclei, NRF, nuclei sort date, and alignment percentage as

257   covariates in the differential chromatin analysis. When performing a sex-stratified analysis of

258   LOAD vs. control samples in the NeuN+ nuclei, we selected alignment percentage in the female-

259   only subset and nuclei sort date in the male-only subset. Finally, in the sex-stratified differential

260   chromatin analysis of LOAD vs. control NeuN- nuclei, we included normalized peak calls as a

261   covariate in the female-only subset, and normalized peak calls and percent GC content as

262   covariates in the male-only subset. In the supplemented comparison of LOAD vs. control NeuN-

263   nuclei in females only (nine additional non-neuronal samples), percent GC content was included

264   as a covariate.

265   <u>Differential chromatin accessibility analyses</u>

266   Differential ATAC-seq peaks were detected using EdgeR package (version 3.22.3), which

267   models counts using a negative binomial distribution[80]. ATAC-seq reads corresponding to chrX

268   and chrY were excluded due to unequal numbers of female and male donors. For each

11

269 comparison, counts of reads within peaks which were merged from all of the replicates were

270 extracted from BigWig files and normalized by weighted trimmed mean of M-values[81] (TMM).

271 Quasi-likelihood F-tests (QLF) was performed to determine differential sites at cut off adjusted p

272 values < 0.05. As an approximate error model, QLF works more robustly and gives more reliable

273 Type I error rate control than the other options, especially when there are smaller numbers of

274 replicates (EdgeR User Guides, Bioconductor package vignettes). Three levels of case-to-

275 control chromatin accessibility comparisons were performed (**Fig. 1g**): *Level_1*: 40 NeuN- vs 40

276 NeuN+, all samples are from matched donors; *Level_2*: (A) NeuN-, 19 cases vs 21 controls (B)

277 NeuN+, 19 cases vs 21 controls; *Level_3*: (A) NeuN- females, 11 cases vs 11 controls; (B)

278 NeuN- males 8 cases vs 10 controls; (C) NeuN+ females, 11 cases vs 11 controls; and (D)

279 NeuN+ males, 8 cases vs 10 controls. Since the number of female samples was less than that of

280 male, we included 9 additional NeuN- datasets (those that did not have matching NeuN+ data

281 from the same donor but met QC criteria) into Level 3 and processed an additional comparison:

282 (E) NeuN- females,14 cases vs 13 control. For all these comparisons we included covariates

283 (described in "covariate selection") in the EdgeR analysis with model design as open chromatin

284 accessibility ~ Disease (LOAD state) + covariates. Importantly, for each analysis, we remerged

285 and recalled the chromatin peaks. Thus, the separate analyses were not direct subsets of each

286 other.

287 Genomic distribution analysis

288 ATAC-seq peaks were categorized as promoters (transcription start site to upstream 1kb), 1st

289 exon, intragenic (excluding 1st exon), 5' UTR, 3' UTR, and intergenic, based on human gene

290 hg19 NCBI RefSeq gene information from the UCSC genome browser.

291 Gene ontology analysis

292   Differential open chromatin regions were characterized by GREAT (v 3.0.0)

293   (http://bejerano.stanford.edu/great/public/html/), using the hg19 genome as background regions.

294   Genes associated to open chromatin regions were determined by default "basal plus extension"

295   settings (i.e., 5 kb upstream, 1 kb downstream, plus distal up to 1000 kb).

296   Overlap with LOAD GWAS loci

297   The tag SNP for 25 LOAD loci were obtained from the literature[7]. Genome coordinates +/-

298   100kb surrounding each GWAS tag SNP were used to identify any differential open chromatin

299   region that mapped within the region. Permutation analyses were performed by randomly

300   selecting the same numbers of sites from the union set of ATAC-seq peak calls. For each

301   comparison, we performed 10,000 permutations to estimate the empirical P-value.

302   Motif search

303   Transcription factor motif enrichments were evaluated using HOMER[82] (v4.10.3) and MEME

304   Suite[83]. For each comparison, we used the centered 300 bp of open chromatin regions as input,

305   and the union peak calls as background. GC matching was applied to the background peaks to

306   ensure that this did not lead to spurious results. We primarily used all peaks as the background,

307   but to check for accuracy, we also randomly selected 10,000 peaks as the background from all

308   peaks. For this analysis, we focused on the known motif search rather than search for *de novo*

309   motifs.

310   Differential expression analysis of ROSMAP data

311   Gene expression data from AMP-AD was obtained from the Religious Orders Study and

312   Memory and Aging Project (ROSMAP)[84, 85]. This study includes RNA-seq data from the

313   dorsolateral prefrontal cortex of 724 subjects[86, 87]. limma was used to perform the differential

314   analysis in R on normalized FPKM values obtained from RNA-seq of ROSMAP bulk tissue

13

315   samples. Samples with cogdx of 1 (no cognitive impairment, n=201) and 4 (Alzheimer's

316   dementia and no other case of cognitive impairment, n=222) were included in the analysis.

317   Single-nuclei library preparation and sequencing

318   Libraries were prepared for snRNA-seq using the Chromium Single Cell 3' Reagent Kits v3

319   (10X Genomics) according to the manufacturer's instructions. Nuclei were diluted in nuclease-

320   free water to a concentration of 1 million nuclei / ml in a final volume of 100 ul, and transferred

321   to an 8-strip tube on ice. 16,000 nuclei were added to the Master Mix for a targeted nuclei

322   recovery of 10,000. Libraries were pooled onto a single S1 flow cell, and sequenced using the

323   Illumina NovaSeq 6000 system to obtain paired-end 2 x 100bp reads. Sequencing saturations

324   ranged from 59.3% to 81.2%.

325   snRNA-seq data analysis

326   *CellRanger* software version 3.1.0 (10X Genomics) was used to demultiplex raw Illumina base

327   call files into FASTQ files. A pre-mRNA GTF file was generated with the pre-built GRCh38

328   3.0.0 human reference using the Linux utility *awk*, in order to be compatible with CellRanger

329   and to include intronic reads from nuclear RNA in UMI counts for each gene and barcode. The

330   CellRanger count pipeline was run to align reads to the pre-mRNA GRCh38 reference and gene

331   expression matrices generated separately for each of the four samples were merged together into

332   a single matrix.

333   For quality control and filtering, the *quickPerCellQC* function from the R package *scater*

334   v1.14.6[88] was used to identify low-quality cells based on QC metrics (UMI count, number of

335   genes detected, percentage of UMIs that mapped to the mitochondrial genome). Cells in which

336   229 or more genes were detected and less than 17.4% of UMIs mapped to the mitochondrial

337   genome were used in downstream analyses. Out of 21,262 nuclei in the initial dataset 18,032

14

338   remained after filtering with a median number of 2,332 detected genes per nucleus. The R

339   package Seurat v3.1.1 standard workflow was used for integration of multiple samples to

340   combine the four samples into a unique dataset[89]. Prior to integration, gene expression was

341   normalized for each sample by scaling by the total number of transcripts, multiplying by 10,000,

342   and then log transforming (log-normalization). We then identified the 2,000 genes that were

343   most variable across each sample, controlling for the relationship between mean expression and

344   variance. Next, we identified anchor genes between pairs of samples using the

345   *FindIntegrationAnchors* function that were then passed to the *IntegrateData* function to

346   harmonize the four samples.

347   We scaled the integrated dataset before running a Principal Component Analysis (PCA).

348   To distinguish principal components (PCs) for further analysis, we used the *JackStraw* method to

349   determine statistically significant PCs and found that up to 30 PCs were enriched in genes with a

350   PC score that was unlikely to have been observed by chance. We then utilized the shared nearest

351   neighbor (SNN) modularity optimization-based clustering algorithm implemented in Seurat for

352   identifying clusters of cells. This was performed using the *FindNeighbors* function with 30 PCs,

353   followed by the *FindClusters* function with the Louvain algorithm using a 0.4 resolution. This

354   allowed us to assign cells into a total of 21 clusters. We applied the uniform manifold

355   approximation and projection (UMAP) method on the cell loadings of the previously selected 30

356   PCs to visualize the cells in two dimensions and to separate nuclei into clusters.

357   Differential expression to identify cluster markers that are conserved between the

358   samples was performed using the Seurat function *FindConservedMarkers* for each cluster on the

359   normalized gene expression before integration. R package SingleR v1.0.1[90] was used to annotate

360   cell types based on correlation profiles with 713 bulk microarray samples from the Human

15

361   Primary Cell Atlas[91] (HPCA) as reference expression data. Four major cell types were detected

362   by the SingleR method using HPCA: macrophages, astrocytes, neurons, and endothelial cells.

363   Because of the specific expression of microglial markers PTPRC and CSF1R in the macrophage

364   cluster, as well as the differences in biological systems used in the HPCA reference, we

365   manually refined the HPCA annotation of this specific cluster to microglia. Identification of

366   neuronal vs. non-neuronal clusters for some differential expression analyses was performed by

367   determining the proportion of cells expressing the neuronal marker NeuN/RBFOX3. Clusters

368   with more than 50% of cells expressing NeuN/RBFOX3 were defined as neuronal clusters.

369   Control samples showed an average of 2,417 (SD 169) neuronal and 2,319 (SD 1,149) non-

370   neuronal nuclei, whereas LOAD samples showed 1,329 (SD 294) neuronal and 2,952 (SD 1,783)

371   non-neuronal nuclei. NeuN neuronal clusters and neuronal clusters identified by SingleR using

372   the HPCA reference were very consistent with only one additional cluster in NeuN neuronal

373   clusters (out of a total of 16 clusters). Differential expression analyses between female LOAD

374   nuclei and control nuclei within specific groups of nuclei were performed using the Wilcoxon

375   Rank Sum test as implemented in the *FindMarkers* function in the Seurat package. 'min.pct' and

376   'logfc threshold' arguments were set to 0 to allow for the testing of a majority of genes in each

377   analysis.

378
379   **Results:**
380
381   **Study sample and quality control analyses**

382   Nuclei isolated from archived frozen temporal cortex of 51 individuals with LOAD (n=26) or

383   controls (n=25) **(Table S1)** were stained and sorted using a PE-conjugated monoclonal NeuN

384   antibody **(Fig. 1a-d)**. Staining of pre-sorted nuclei with nuclear membrane markers was

385   confirmed by immunofluorescence **(Fig. 1e)**. We observed a smaller proportion of sorted
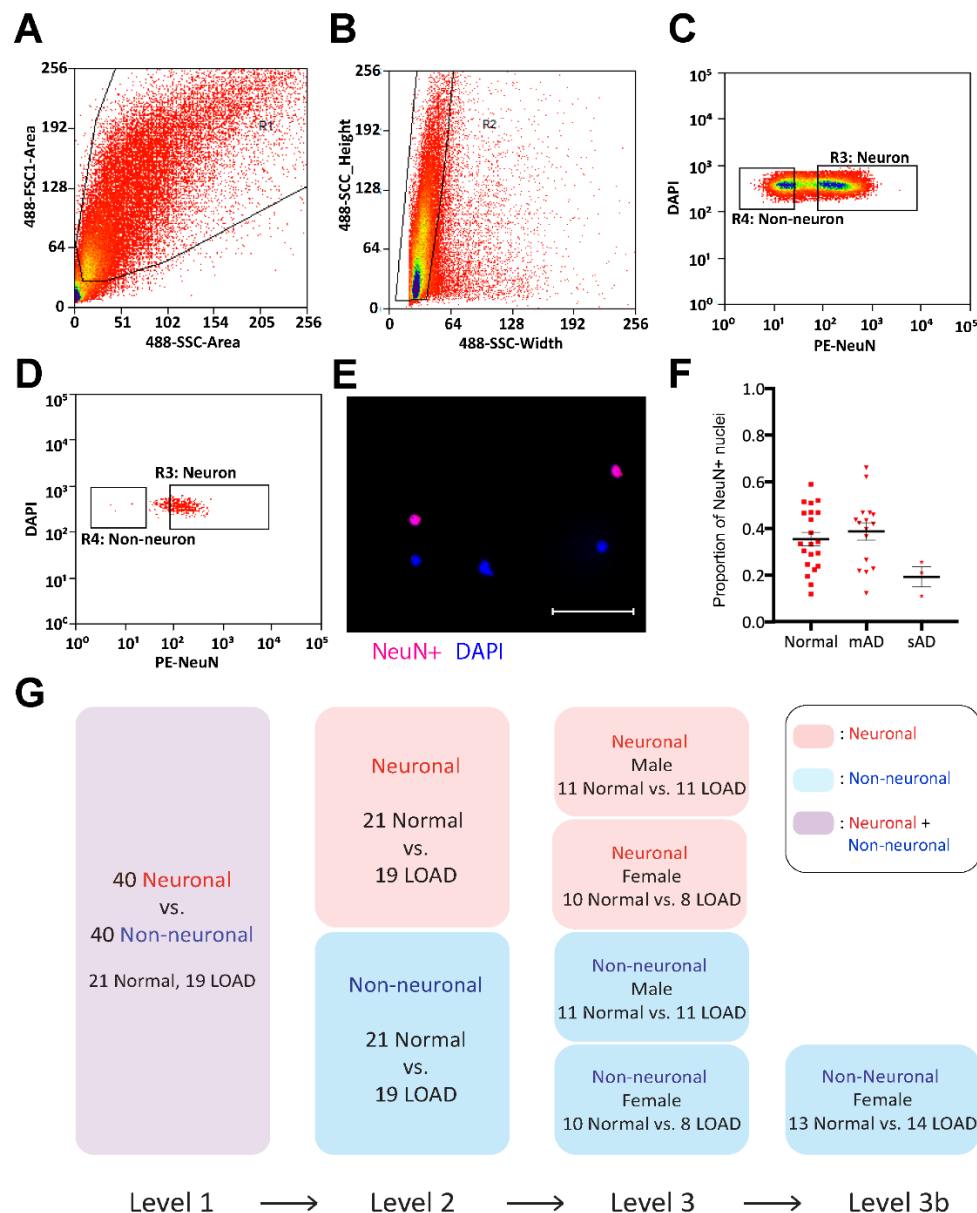
16

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 100um. **(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).

17

400   neuronal (NeuN+) nuclei from total isolated nuclei in the severe LOAD cases (**Table 1, Fig. 1f**),

401   as expected due to neuronal cell loss and gliosis, hallmarks of LOAD pathological progression[30,]

402   [31]. We next performed ATAC-seq using the NeuN+ and NeuN- sorted nuclei populations. A total

403   of 90 neuronal and non-neuronal datasets passed quality control criteria (see Methods), which

404   were derived from 40 individuals (19 LOAD cases and 21 controls) that had matched neuronal

405   and non-neuronal data, and an additional 9 individuals (3 LOAD cases and 6 controls) that had

406   only non-neuronal data. The final cohort of 40 consisted of 22 males and 18 females with similar

407   post-mortem intervals (PMI); all donors were Caucasians and homozygous for *APOE* e3 (**Table**

408   **S2**). Subsequent analyses compared different sub-groups (**Fig. 1g**). Correlations of all potential

409   numerical covariates showed expected patterns of co-linearity (**Fig. S1**). The samples displayed a

410   range of QC metrics, with some samples displaying a higher signal to noise (**Table S1**).

411   Repeated experiments on a subset of samples demonstrated that signal to noise was reproducible

412   and thus a sample-specific characteristic (**Table S1**). In addition to randomizing sample

413   preparation (see Methods), we found that no metadata variables were significantly associated

414   with case-control status (**Fig. S2**), indicating an absence of batch effects.

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

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

417

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

419   To demonstrate that we can robustly identify chromatin accessibility differences between major

420   cell types of the brain, we first compared the ATAC-seq profiles between neuronal and non-

421   neuronal nuclei groups from the entire study sample (Level 1, **Fig. 1g**). PCA of all samples

422   showed that 37.5% of the total variance was explained by the cell type (neuronal versus non-

423   neuronal, **Fig. S3, Fig. S11**). Performing quantitative differential chromatin accessibility analysis

424   using EdgeR (FDR q < 0.05), we found that 87,570 regions were more accessible in the neuronal

425   population, 83,171 regions were more accessible in the non-neuronal population, and 54,484

426   regions were not detected as differential **(Fig. 2a, Table S10)**. Representative screenshots show

427   differential ATAC-seq peaks around genes known to be expressed specifically in neuronal **(Fig.**

428   **2b)** or non-neuronal **(Fig. 2c)** cell types. A higher percentage of sites more accessible in non-

429   neuronal population mapped to promoters, which could be explained by the more distal

430   regulation of neuronal cell types or the greater diversity and heterogeneity of the cell types

431   composing the non-neuronal population **(Fig. S4)**. Using GREAT gene ontology analysis [32],

432   neuronal-specific regions were enriched with genes associated with neuronal function, while

433   non-neuronal specific regions were enriched with genes implicated in glial function **(Table S3)**.

434   Motif analysis for regions with increased chromatin accessibility in neuronal nuclei

435   showed enrichment for the transcription factor recognition sites of Early Growth Response 2

436   (EGR2), Regulatory Factor X1 (RFX1), and Myocyte Enhancer Factor 2C (MEF2C)[33-35], while

437   motif analysis for regions with increased chromatin accessibility in non-neuronal nuclei showed

438   enrichment for the SRY-related HMG-box (SOX) transcription factor family[36] **(Table S4)**. When

439   compared to a previous NeuN sorted-nuclei ATAC-seq study[37] using a smaller number (n = 8) of

440   healthy brain samples, we found a substantial degree of overlap **(Fig. S5a)** with similar peak
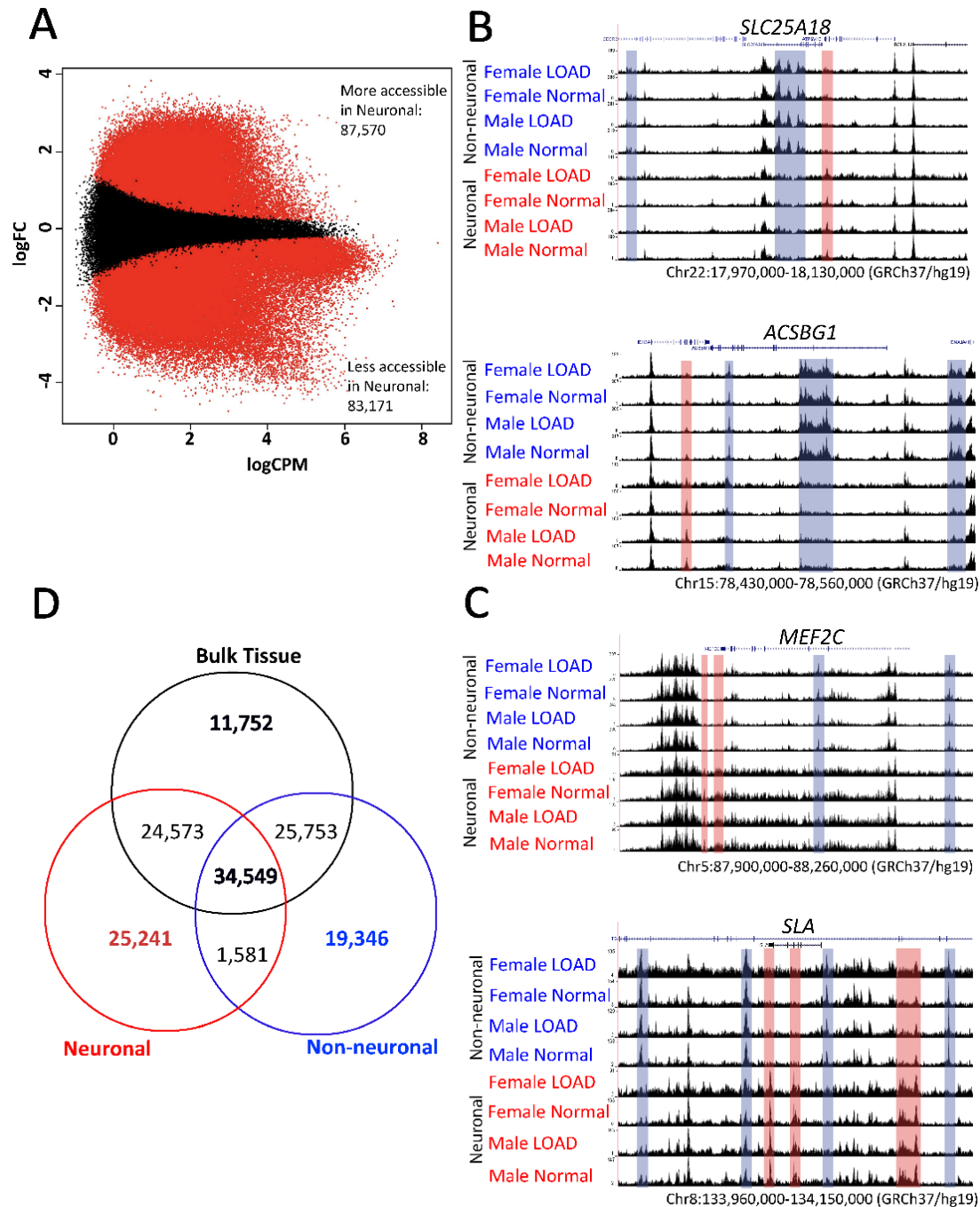
441
Fig. 2: Level 1 comparison of ATAC-seq data from neuronal vs non-neuronal nuclei. (a) MA plot
showing differential ATAC-seq sites between neuronal (blue) vs. non-neuronal regions (red). Red dots
represent ATAC-seq peaks that are significantly different between groups (FDR q < 0.05). **(B)** ATAC-
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.

450   length distribution **(Fig. S5b)**. However, the 5x larger number of samples used in our study

451   identified >80,000 additional significantly differential chromatin sites.

452   **Comparison of sorted nuclei vs. bulk brain tissue homogenate.**

453   To our knowledge, no study has directly compared chromatin accessibility differences between

454   bulk brain tissue, neuronal, and non-neuronal fractions. Bulk brain ATAC-seq data was

455   generated using pulverized frozen whole tissue homogenate samples from six individuals for

456   whom high-quality ATAC-seq data was collected from neuronal and non-neuronal populations

457   **(Table S5)**. Binary peak overlap analysis shows that while there are some regions that are only

458   accessible in bulk tissue, a substantially larger number of sites are uniquely detected in the nuclei

459   populations of neuronal and non-neuronal only **(Fig. 2d)**. This observation justifies the

460   performance of ATAC-seq analysis by brain cell-type to reduce cellular heterogeneity as it

461   allows the identification of more cell type specific signals. Quantitative differential chromatin

462   accessibility (EdgeR comparisons **(Fig. S6)** and GO annotations **(Table S6-8)** showed that

463   chromatin accessibility sites specifically identified in bulk tissue, neuronal and non-neuronal

464   populations displayed many biologically relevant pathways.

465   **Identification of LOAD-associated differences in chromatin accessibility**

466   To determine the association of LOAD status with changes in chromatin accessibility, we

467   performed differential ATAC-seq analysis of 19 LOAD compared to 21 healthy controls

468   stratified by brain cell-type (neuronal and non-neuronal only populations; Level 2, **Fig. 1g)**. The

469   comparison using the neuronal nuclei data resulted in 211 neuronal chromatin differences (FDR

470    $q < 0.05$) between LOAD and control **(Fig. 3a, Table S11)**, while the analysis of the non-

471   neuronal nuclei detected no differential chromatin between LOAD and control (**Fig. 3b**). Motif

472   analysis using the 141 sites that showed decreased chromatin accessibility in LOAD neuronal

473   nuclei were enriched for transcription factor motifs including Wilms Tumor protein (WT1),

474   Early Growth Response 1 (Egr1), Retinoic acid receptor gamma (RARg), and Kruppel like factor

475   14 (KLF14) **(Fig. 3c)** that have been reported relevant to brain function[38-41]. No significant motif

476   enrichment was detected from 70 sites (FDR q < 0.05) with increased chromatin accessibility in
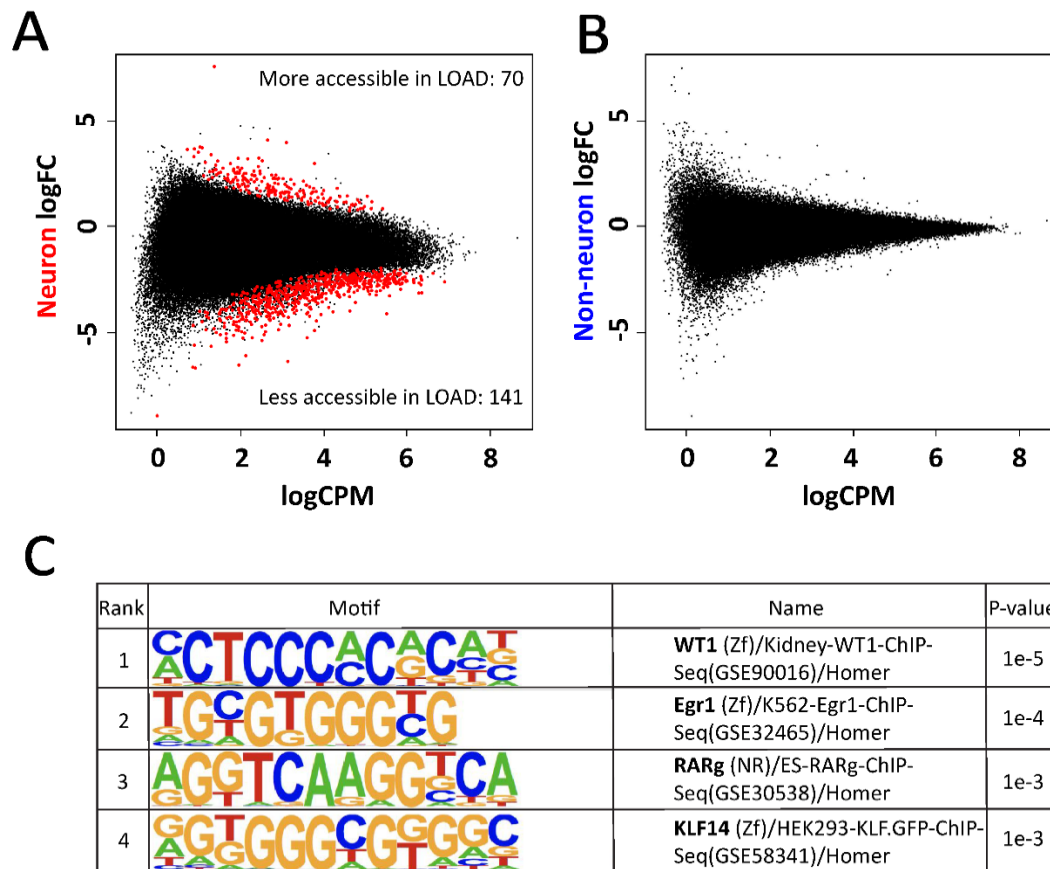
477   neuronal LOAD samples.

478
479   **Fig. 3: Level 2 comparison of ATAC-seq data between LOAD cases and controls.** MA plots
480   of differential chromatin sites for (A) neuronal and (B) non-neuronal nuclei. Red dots represent
481   differential ATAC-seq sites with FDR q < 0.05. (C) Motifs that are enriched in neuronal ATAC-
482   seq sites that are less accessible in LOAD samples. Size of red dots were increased for visibility.
483

484   **Identification of sex-dependent chromatin accessibility differences between LOAD and**

485   **control samples**

22

486   Since sex has an effect on LOAD onset and progression[42], we performed a sex-stratified

487   differential analysis of LOAD compared to control (Level 3, **Fig. 1g)** to examine whether the

488   effect of sex on LOAD risk is mediated, at least in part, by chromatin remodeling. The

489   differential analysis of the neuronal groups did not yield significant differences between LOAD

490   and normal when stratified by sex for either females or males (FDR q > 0.05) **(Fig. 4a-b)**.

491   Analysis of the non-neuronal nuclei identified 24 chromatin accessibility differences in female

492   LOAD compared to control **(Fig. 4c, Table S12)**, while no significant differences were identified

493   when the analysis was performed in males **(Fig. 4d)**. To further investigate this trend, we

494   increased the sample size with 9 additional non-neuronal samples (6 female controls and 3

495   female LOAD cases). Differential analysis using ATAC-seq data from the larger female non-

496   neuronal dataset resulted in 842 differential sites between LOAD and control (FDR q < 0.05)

497   **(Fig. 4e, Table S13)**.

498   Comparison of these results with the LOAD associated chromatin accessibility sites

499   obtained from the differential analysis of female neuronal nuclei showed that while there are

500   differences detected in both female neuronal and glial cells in LOAD cases versus controls, we

501   detected larger effect sizes in the female glia cells (**Fig. S12a**). In addition, several differential

502   regions identified in female non-neuronal nuclei were also observed in male non-neuronal nuclei

503   with similar trends, however, these did not reach statistical significance (**Fig. S12b**).

504   Nonetheless, these associations in the female group showed a stronger effect size (**Fig. S12b**).

505   We performed motif enrichment using the 203 sites that were more accessible in non-

506   neuronal female LOAD and found significantly enriched motifs for the SOX family **(Fig. 4f)**.

507   The analysis of 639 sites that displayed decreased accessibility in non-neuronal female LOAD
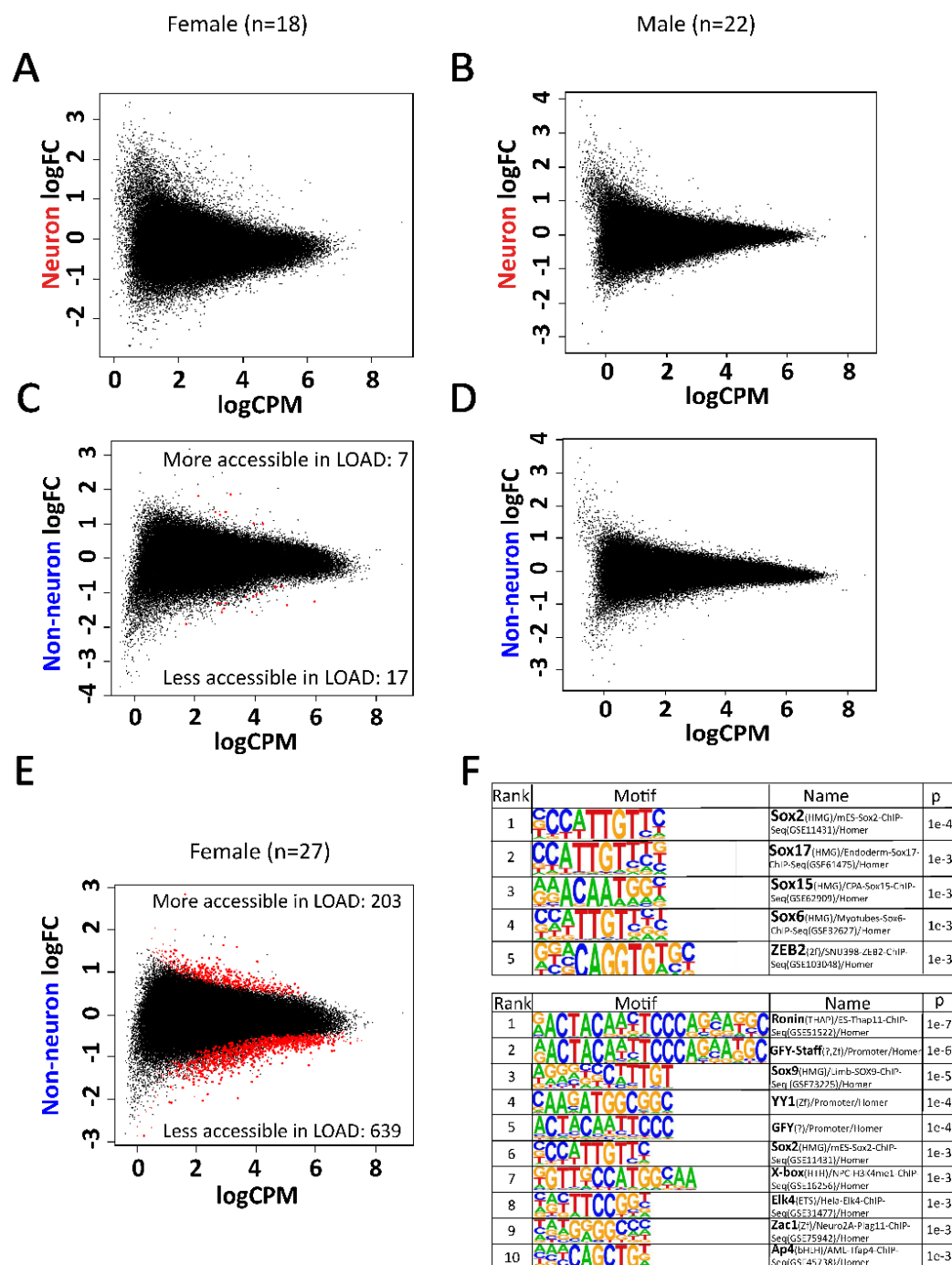
23

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.

516 was enriched for transcription factors known to be highly associated with glia or neuron

517 functions, such as RONIN, SOX9, YY1 and ELK4[43-46] **(Fig. 4f)**.

518 To determine whether binding of transcription factors identified in female non-neuronal

519 cells (Level 3, **Fig. 1g**) has downstream effects on expression of target genes, we used the

520 publicly available ROSMAP bulk RNA-seq data to perform differential expression analyses of

521 those genes for LOAD vs. Normal. We found that several genes exhibited changes in the

522 expected direction, including *ZAC1* (*PLAGL1*), *YY1*, and *SOX2* (**Table S15**). This provides

523 further mechanistic insights into gene dysregulation underlying LOAD pathogenesis in females.

524 Last, we also used the chromatin accessibility differential sites identified in female non-neuronal

525 LOAD (Level 3) for GO analysis and found pathways involved in immune response and

526 myelination **(Table S9)**.

527 **Overlap of LOAD-specific differential chromatin accessibility sites with LOAD GWAS**

528 **regions**

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

530 these data with LOAD GWAS regions[7]. We defined LOAD GWAS regions by anchoring on the

531 top 25 associated SNPs +/- 100kb, and cataloged LOAD-GWAS regions of 200kb each.

532 Importantly, we were comparing genomic regions of chromatin accessibility to the GWAS

533 regions, not chromatin QTLs, to GWAS loci. Of the 211 LOAD-specific differential chromatin

534 accessibility sites in neuronal samples (FDR q < 0.05**, Fig. 3a)**, we identified five sites that

535 overlap four of the 25 LOAD GWAS regions **(Table 2)**. We show representative examples of

536 overlapping LOAD-specific differential chromatin accessibility sites surrounding the *PTK2B* and

537 *CLU* GWAS loci **(Fig. 5a, Fig. S7)**. Using permutation testing, we did not detect a significant

538 enrichment with LOAD GWAS regions compared to regions selected at random **(Fig. S8a-b).**

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

Neuron control vs. LOAD differential peaks overlapped with 25 LOAD-GWAS regions					
	Peak call	**GWAS ± 100kb**	**Gene**	**SNP**	**SNP Position**
Differential sites with FDR < 0.05	More accessible in LOAD				
	chr6:32551592-32552707	chr6:32475406-32675406	*HLA-DRB1*	Rs9271058	chr6:32575406
	chr6:32489394-32490069	chr6:32475406-32675406	*HLA-DRB1*	Rs9271058	chr6:32575406
	Less accessible in LOAD				
	chr20:54994099-54994641	chr20:54897568-55097568	*CASS4*	Rs6024870	chr20:54997568
	chr8:27209401-27210247	chr8:27119987-27319987	*PTK2B*	Rs73223431	chr8:27219987
	chr8:27563212-27563671	chr8:27367686-27567686	*CLU*	Rs9331896	chr8:27467686
Female glia control vs. LOAD differential peaks overlapped with 25 LOAD-GWAS regions					
	Peak call	**GWAS ± 100kb**	**Gene**	**SNP**	**SNP Position**
Differential sites with FDR < 0.05	More accessible in LOAD				
	chr10:11784165-11784736	chr10:11620308-11820308	*ECHDC3*	Rs7920721	chr10:11720308
	Less accessible in LOAD				
	chr14:92966426-92967442	chr14:92832828-93032828	*SLC24A4*	Rs12881735	chr14:92932828
	chr16:19894588-19895301	chr16:19708163-19908163	*IQCK*	Rs7185636	chr16:19808163
	chr17:61627113-61628372	chr17:61438148-61638148	*ACE*	Rs138190086	chr17:61538148
	chr19:1101756-1102160	chr19:956492-1156492	*ABCA7*	Rs3752246	chr19:1056492
	chr19:45416036-45416558	chr19:45311941-45511941	*APOE*	Rs429358	chr19:45411941
	chr19:45428555-45429362	chr19:45311941-45511941	*APOE*	Rs429358	chr19:45411941
	chr19:45454352-45455071	chr19:45311941-45511941	*APOE*	Rs429358	chr19:45411941
	chr7:100025783-100028055	chr7:99991795-100191795	*NYAP1 (ZCWPW1)*	Rs12539172	chr7:100091795

543

544   Of the 842 differential sites in non-neuronal female LOAD (FDR q < 0.05**, Fig. 4e)**, we

545   detected nine differential sites that overlap seven of the 25 LOAD-GWAS regions **(Table 2)**.

546   Using permutation analysis (see Methods), we found significant enrichment in the amount of

547   overlap with sites less accessible in non-neuronal female LOAD **(P < 0.05, Fig. S8c-d)**.
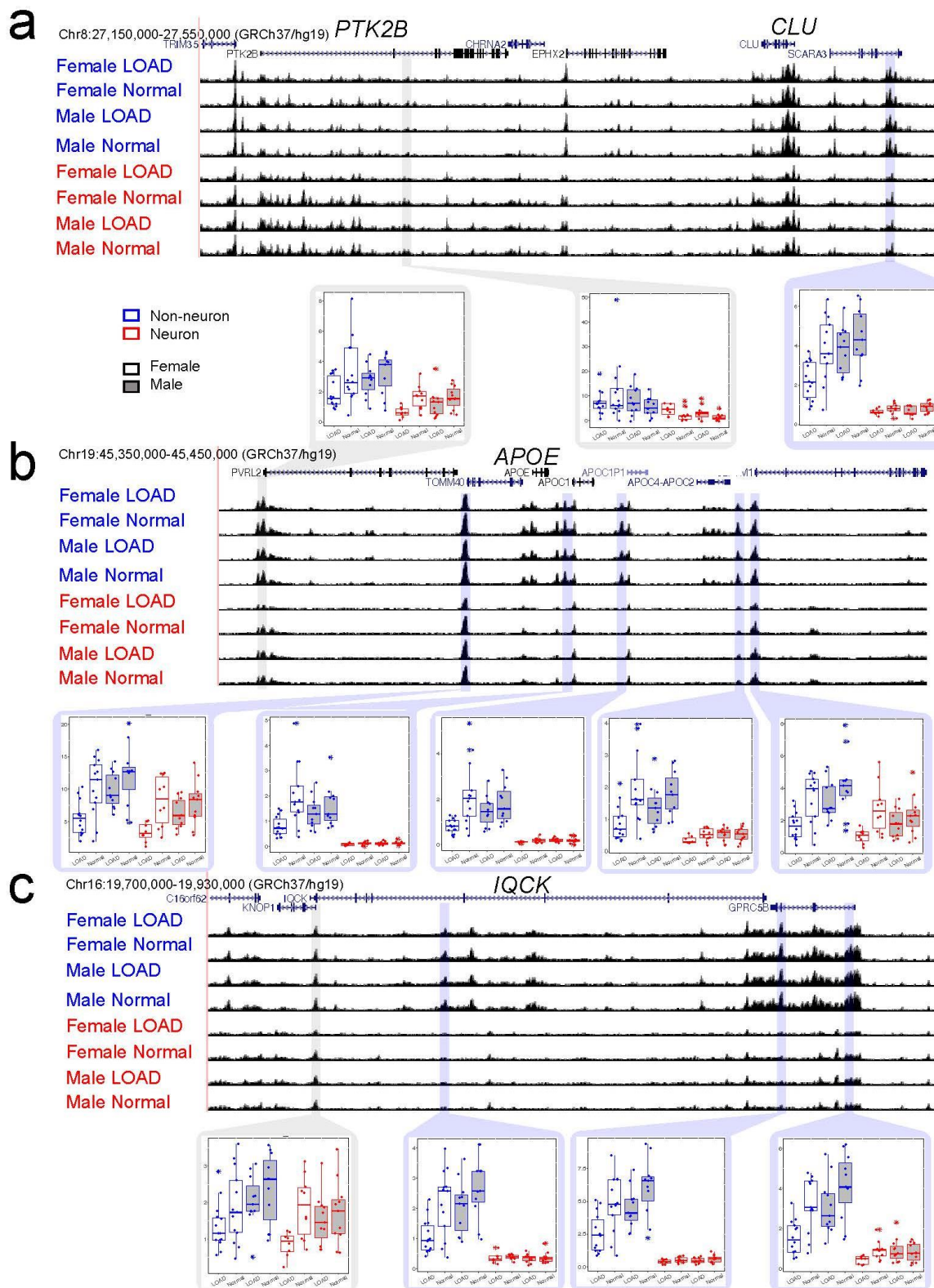
26

548
549   **Fig. 5** (Continued on the following page.)

550 **Fig. 5: Differential LOAD specific ATAC-seq peaks around LOAD-GWAS regions.**

551 Screenshots of ATAC-seq data around **(A)** *PTK2B, CLU*, **(B)** *APOE*, and **(C)** *IQCK* loci. Box

552 plots show ATAC-seq read counts for individual ATAC-seq peaks (blue frames highlight

553 significant differential peaks for cases vs. controls, gray frames show control peaks that are not

554 differential between cases and controls). Box plots are color coded for non-neuronal (blue)

555 neuronal (red), female (no fill), and male (gray fill). All tracks show hg19 coordinates and all y-

556 axes on tracks range from 0 to 250. For box plots, the line within each box represents the

557 median, and the top and bottom borders of the box represent the 25th and 75th percentiles,

558 respectively. The top and bottom whiskers of the box plots represent the 75th percentile plus 1.5

559 times the interquartile range and the 25th percentile minus 1.5 times the interquartile range,

560 respectively.

561

562 Representative examples of differential chromatin accessibility sites between LOAD cases and

563 controls overlapping representative LOAD-GWAS regions are shown for the loci *APOE* **(Fig.**

564 **5b, Fig. S7b)**, and *IQCK* **(Fig. 5c, Fig. S7c).** While differences around the *APOE* and *IQCK*

565 locus are more pronounced in females, we also detect subtle differences in the same direction in

566 males **(Fig. 5b-c, Fig. S7b-c).** The identification of LOAD differences in chromatin accessibility

567 in the *APOE* region, despite all cases and controls being of *APOE* e3/e3 genotype, underscores

568 our identification of a regulatory mechanism, rather than genetic mechanism, at this locus.

569 **Overlap of LOAD-specific differential chromatin accessibility sites in female non-neurons**

570 **with single nuclei RNA-seq data**

571 Next, we performed functional validation of key findings from the chromatin accessibility

572 differential sites in female LOAD non-neuronal cells that overlapped LOAD-GWAS regions

573 **(Table 3, Fig. S9)**. We focused the analysis on two representative LOAD differential chromatin

574 accessibility sites in female LOAD non-neuronal cells that overlapped LOAD-GWAS loci: (1)

575 LOAD more accessible sites surrounding the *ECHDC3* locus (annotated by the proximate gene

576 to the associated SNP) and (2) LOAD less accessible sites surrounding the *ABCA7* locus. The

577 loci were defined by anchoring on the associated SNPs +/- 500kb, rs7920721 (*ECHDC3*) and

578   rs3752246 (*ABCA7*) using the UCSC Genome Browser[47] (http://genome.ucsc.edu/)

579   GRCh38/hg38 assembly released December 2013 (**Table 3, Fig. S9**). Single-nuclei (sn)RNA-seq

580   data was collected from female brains using a small subset group of two LOAD and two control

581   (**Table S2,** marked with *****), and differential expression analysis of female LOAD nuclei

582   compared to control nuclei was performed for all genes mapped with these regions in three cell-

583   type specific groups: 1) NeuN- non-neuronal, 2) Human Primary Cell Atlas (HPCA)-annotated

584   astrocyte, and 3) HPCA-annotated microglia. Pseudogenes, RNA genes, and novel transcripts

585   were excluded from the analysis. Out of 54 total genes within 1Mb of the two SNPs, 8 within

586   *ECHDC3* locus and 46 within *ABCA7* locus (**Fig. S9**), many were up- or down- regulated as

587   predicted by chromatin accessibility profiles (*i.e.* increased chromatin accessibility overlapping

588   promoters and enhancers was associated with upregulation and vice versa). Overall 28 (51.9%),

589   23 (42.6%), and 24 (44.4%) genes of the NeuN-, astrocyte, and microglia clusters, respectively,

590   showed differential expression with trends corresponding to the changes in chromatin

591   accessibility. Of those genes, some showed nominal significance (6 (21.4%) for NeuN-, 3

592   (13.0%) for astrocyte, and 5 (20.8%) for microglia clusters), but did not reach adjusted statistical

593   significance, while other trends did not reach nominal significance. This could be explained by

594   the small sample size and/or number of cells in the clusters. Furthermore, we examined the

595   consistency between our dataset and a similar recently reported dataset[28]. To address this

596   question we conducted differential expression analysis for HPCA-annotated neurons, astrocytes,

597   and microglia and compared the results for the top 15 significant genes, *i.e.* 5 from each cell type

598   identified by the previous study[28]. We found that out of these 15 genes, our results show the

599   same directionality for 13 genes, 10 of which also had significant adjusted p-values (**Table S14**),

600   providing further validation of our findings.

601 **Table 3. Differential snRNA-seq expression from level 3 female glia controls vs cases that**
602 **are within 500 kb of LOAD-GWAS SNPs at loci found to be more accessible (rs7920721) or**
603 **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
rs7920721 (chr10:11,678,309, *ECHDC3*)	*CELF2*	0.087	1	-0.090	1	0.47	**4.90 x 10^{-16}**
	USP6NL	0.12	0.32	-0.022	1	0.29	1
	ECHDC3	0.067	1	0.016	1	0.31	1
	UPF2	-0.34	**0.00042**	-0.31	1	-0.48	1
	DHTKD1	-0.16	1	-0.23	1	0.011	1
	SEC61A2	0.028	1	-0.028	1	0.15	1
	NUDT5	-0.032	1	-0.097	1	0.21	1
rs3752246 (chr19:1,056,493, *ABCA7*)	*BSG*	0.035	1	0.10	0.32	-0.11	1
	POLRMT	-0.047	1	-0.030	1	0.061	1
	FSTL3	-0.12	0.065	-0.13	1	0.047	1
	PALM	-0.071	1	0.010	1	-0.25	1
	PTBP1	-0.034	1	-0.053	1	0.065	1
	AZU1	-0.012	1	-0.015	1	-0.039	1
	MED16	-0.036	1	-0.051	1	-0.19	1
	ARID3A	0.089	0.97	0.047	1	0.12	1
	TMEM259	0.050	1	0.089	1	-0.11	1
	ABCA7	-0.013	1	-0.022	1	0.075	1
	STK11	-0.024	1	-0.041	1	-0.13	1
	CBARP	0.013	1	0.017	1	0.026	1
	ATP5F1D	0.040	1	0.11	1	-0.032	1
	MIDN	0.042	1	-0.073	1	0.14	1
	CIRBP	-0.71	**1.52 x 10^{-57}**	-0.80	**1.27 x 10^{-42}**	-0.51	1
	C19orf24	0.033	1	0.014	1	0.12	1
	NDUFS7	0.083	**0.016**	0.11	**0.0014**	-0.014	1
	GAMT	-0.10	1	-0.065	1	-0.28	1
	RPS15	-0.14	1	-0.19	0.66	0.088	1
	APC2	-0.027	1	0.022	1	-0.14	1
	PCSK4	-0.033	1	-0.025	1	-0.10	1

604 Negative mean log(FC) values indicate down-regulation in LOAD and positive values indicate up-
605 regulation in LOAD. Significant adjusted p values are bolded and mean log(FC) values consistent with
606 ATAC-seq data are underlined. FC = fold change. Genes that were not detected or that had mean log(FC)
607 values below 0.01 for any of the three groups were excluded from the table. Refer to detailed genomic
608 structure presented in Supplemental Figure 9.
609

610 **Table 3** catalogues the genes surrounding the *ECHDC3* and *ABCA7* loci, excluding 26

611 genes that were not detected or had low fold change values. Out of seven genes positioned within

612 the 1Mb region of the LOAD associated *ECHDC3* SNP, we found that *CELF2* was significantly

613 upregulated in LOAD HPCA-annotated microglia clusters (mean log(fold change) = 0.47,

614 adjusted p = 4.90x10^{-16}, **Table 3**). In addition, out of 21 genes mapped within the 1Mb region of

615   the LOAD associated *ABCA7* SNP, *CIRBP* was significantly downregulated in LOAD NeuN-

616   non-neuronal clusters (mean log(fold change) = -0.71, adjusted p = 1.52x10^{-57}, **Table 3**) and

617   HPCA-annotated astrocyte clusters (mean log(fold change) = -0.80, adjusted p = 1.27x10^{-42},

618   **Table 3**). These results demonstrated that alteration in chromatin accessibility in LOAD-GWAS

619   loci correlates with changes in gene expression. Furthermore, the results suggested that the target

620   genes within LOAD-GWAS loci affected by LOAD specific changes in chromatin state may not

621   be simply interpreted as the most proximate gene to the associated SNP.

622   The intersection of LOAD- and glia- specific ATAC-seq sites and snRNA-seq patterns

623   discovered novel LOAD regulatory elements that lead to gene expression changes in LOAD.

624   Furthermore, intersecting these data sets provided a functional validation for the top significant

625   findings from the ATAC-seq experiments showing sex-dependent LOAD-specific open and

626   closed chromatin accessibility sites in non-neuronal cells. Collectively, these results suggest that

627   integrating brain cell-type specific 'omics data is a powerful mechanistic strategy to discover

628   regulatory elements that impact expression of disease genes in a cell-type specific manner.

629
630   **Discussion:**
631
632   Decoding the genetic and genomic mechanisms of LOAD is a major challenge in the post-

633   GWAS era, since the majority of the LOAD associated SNPs are in noncoding regions[1-7].

634   Noncoding disease-associated loci have been shown to be enriched for regulatory elements in

635   tissues and cells relevant to the disease[37, 48-50]. Thus, post-GWAS research requires an in-depth

636   characterization of cell-type specific DNA regulatory elements. Mapping chromatin accessibility

637   has been widely used to identify the location of active DNA regulatory elements, including

638   promoters, enhancers, and insulators[51-54]. We performed the first systematic interrogation of the

639   chromatin accessibility landscape in neuronal and non-neuronal sorted nuclei from LOAD and

640   healthy brains.

641   To our knowledge, this study represents the first and the most comprehensive dataset to

642   date of chromatin accessibility in LOAD brains and matched controls. Furthermore, this study

643   was performed with brain cell-type specific resolution, *i.e.* neuronal and non-neuronal cells. The

644   study has four major findings for the field of LOAD epigenetics. First, we have generated a map

645   of LOAD-associated cell-type specific chromatin accessibility sites. Second, we provide a

646   catalogue of female-specific disease-associated chromatin accessibility sites in non-neuronal

647   cells. Third, we suggest a non-coding regulatory mechanism, namely chromatin accessibility, by

648   which ~25% of LOAD-GWAS loci may exert their effect. Fourth, we have demonstrated that

649   LOAD associated changes in chromatin accessibility can result in gene dysregulation through

650   their overlap with the transcriptome profile of LOAD-GWAS loci. Overall, these results suggest

651   that the *cis* interactions between regulatory elements and key genes contribute, at least in part, to

652   the development and/or progression of LOAD. Of note, 87% of the differentially accessible

653   chromatin peaks in the female NeuN- analysis and 92% of the differentially accessible chromatin

654   peaks in the NeuN+ analysis fall within known regions of topologically associating domains

655   (TADs) from fetal brain[55]. This suggests that chromatin regions associated with risk for LOAD

656   belong to highly interacting regulatory domains. Furthermore, several LOAD loci may exert their

657   pathogenic effects in a cell-type specific manner while others act in multiple cell types to drive

658   LOAD pathogenesis. Alternatively, these cell-type specific and common changes in chromatin

659   structure may represent secondary effects as consequences of disease-related processes such as

660   neurodegeneration and gliosis.

661   A growing number of epigenome-wide association studies (EWAS) in LOAD have

662   profiled DNA methylation, hydroxymethylation, and histone acetylation marks (H4K16ac), and

663   assessed associations with LOAD risk and other related endophenotypes including the burden of

664   pathology[17-24]. These studies have been a powerful approach to validate known LOAD loci,

665   discover new candidate genes, and identify disease-related pathways[17-24]. However, the majority

666   of the LOAD EWAS datasets were generated using bulk brain tissues, and the heterogeneous

667   cell-type populations of brain tissue samples pose technical and biological limitations. Our

668   results showed that a substantially larger number of ATAC-seq sites are uniquely detected in the

669   neuronal and non-neuronal cells, compared to bulk tissues obtained from the same donors. This

670   observation demonstrates the importance of cell-type specific epigenomic studies relative to bulk

671   tissues, as it reduces cellular heterogeneity allowing the identification of more cell-type specific

672   signals.

673   Collectively, our outcomes open a new window for the exploration of the particular cell-

674   types that contribute to LOAD pathogenesis, and the genes and pathways that mediate the cell-

675   type specific pathogenic effects.

676   These data advance the mechanistic understanding of LOAD, and moreover, uncover

677   new candidate LOAD loci. To date, in addition to this study only five others (two transcriptome,

678   two DNA-methylation, and one histone acetylation) have compared genomic signatures stratified

679   by different cell types in the LOAD brain using sorted- and single- nuclei based methodologies[27]

680   [28 29 51 56]. Altogether, our and other studies demonstrated the importance of applying these cell-

681   type specific approaches in molecular analyses of brain tissues and highlight the impact of

682   transitioning into single-cell based 'omics studies in LOAD functional genomic research. In

683   addition, we show that integrating our cell-type specific ATAC-seq data with our scRNA-seq

33

684   LOAD data corroborate the interpretation of the results and provide functional validation. By

685   aligning the two datasets, we identified LOAD-specific female-dependent non-neuronal

686   regulatory elements around two differentially expressed genes in LOAD glia cells. Moreover,

687   analysis of chromatin accessibility differential sites in female LOAD non-neuronal cells that

688   overlapped LOAD-GWAS regions not only provided functional validation and established the

689   link between LOAD -GWAS -ATAC-seq and -snRNA-seq but also demonstrated that genes

690   other than the most proximate to the associated SNP may play a role in LOAD pathogenesis.

691   These results exemplify the potential of integration of cell-type specific datasets to validate

692   known LOAD loci and also to identify new candidate genes. In future studies, a larger sample

693   size may allow conducting a chromatin accessibility QTL study to determine colocalization with

694   GWAS loci.

695   Several bulk tissue ChIP-seq studies have used functional genomics and integrative

696   systems biology approaches to infer cell types. Consistent with our findings, these epigenomic

697   studies strongly suggest that non-neuronal cell types contribute to LOAD-specific histone marks

698   associated with active regulatory elements (promoters and enhancers). It was reported that

699   LOAD GWAS loci were enriched in enhancer elements specific to immune cells[57] and tangle-

700   associated H3K9ac signals located in both promoters and enhancers were significantly associated

701   with modules classified as non-neuronal[58]. Recently, two studies used FANS-sorted nuclei

702   followed by ChIP-seq and demonstrated that microglia were the non-neuronal cell type

703   contributing to LOAD epigenomic signatures. Nott *et al.* found that LOAD SNP heritability was

704   most significant in microglial enhancers[59], and Ramamurthy *et al.* showed that hyperacetylated

705   peaks in microglia colocalize more with LOAD SNPs than the histone acetylome of other cell

34

706   types[56]. Collectively, ours and others' studies point to non-neuronal epigenomic dysregulation,

707   likely microglial, as a major contributing factor to LOAD pathogenesis.

708   Sex is an important factor in LOAD etiology and there are sex differences in disease risk,

709   progression and clinicopathological phenotypes[60]. Foremost, there is a sex-dependent difference

710   in LOAD prevalence and almost two-thirds of LOAD patients are female[61]. Historically, it has

711   been attributed to the longer average life expectancy in women[62], however, recent research

712   suggests that other factors, such as the sudden drop in the level of sex hormones (estrogens) in

713   women at menopause, contribute to the differences in susceptibility between males and

714   females[42]. It was also reported that women manifest faster disease progression and cognitive

715   decline, increased brain atrophy and pathological burden largely driven by neurofibrillary

716   tangles[42], and a more advanced disease stage as indicated by CSF biomarkers, especially higher

717   concentrations of total tau and phosphorylated tau. Although several conflicting reports

718   suggested opposite trends[63], the effect of sex on LOAD has been widely accepted. Nonetheless,

719   the molecular mechanisms underlying the role of sex as a risk factor in LOAD are understudied.

720   Our study provides new insights into these gaps in knowledge showing sex-dependent changes in

721   chromatin structure between LOAD and control brains. We identified hundreds of LOAD

722   differential chromatin accessibility sites specific to females, which overlap nearly one-third of all

723   LOAD GWAS regions. Since the majority of differential chromatin accessibility sites do not

724   overlap LOAD-GWAS regions, these represent novel candidate LOAD loci. Moreover, a female-

725   specific effect on LOAD-associated changes in chromatin accessibility appeared exclusively in

726   glial cells and resulted in nearly three-fold overrepresentation of sites that were more closed in

727   female LOAD patient samples. However, we cannot rule out the possibility that LOAD-

728   associated changes in chromatin accessibility also occur in glia from male LOAD patients, but

729   because of the plausible much smaller effect size in males, we could be underpowered to detect

730   significant associations in our male cohort. These results warrant further investigation to

731   determine the effects of sex-dependent chromatin remodeling on dysregulation of gene

732   expression in the context of LOAD. In this respect, the recent scRNA-seq study[28] also reported

733   sex-dependent LOAD effects and specifically observed a sex-specific differential transcriptional

734   response to LOAD pathology, enrichment of females cells in LOAD-associated cell

735   subpopulation, and higher expression in females for the marker genes of LOAD-associated

736   cellular subpopulations[28]. Future integration of diverse 'omics datasets stratified by sex will

737   decipher the underpinning mechanisms of sex differences in LOAD by establishing the cross

738   interactions between sex-dependent chromatin structure and function, transcriptome and LOAD

739   phenotypic outcomes.

740   One of the loci that showed sex-dependent effects in our study is the *APOE* linkage

741   disequilibrium (LD) region. The e4 allele of the *APOE* gene is the first identified, most highly

742   replicated, and the strongest genetic risk factor for LOAD[64, 65]. Furthermore, LOAD GWAS have

743   confirmed strong associations with the *APOE* LD genomic region, and no other LOAD-

744   association remotely approached the same level of significance [3, 66, 67]. Interestingly, female

745   carriers of *APOE* e4 have an increased risk of LOAD versus male and the adverse effect of

746   *APOE* e4 on LOAD biomarkers was generally stronger in women versus men[68-72]. Overall the

747   *APOE* LD region displayed more open chromatin in glia versus neurons as expected since this

748   gene is much more highly expressed in glia. Interestingly, we found decreased chromatin

749   accessibility at multiple sites in female LOAD glia across the *APOE* region. This result provides

750   molecular clues to the observations that *APOE* e4 allele confers a greater risk for LOAD in

751   women than in men[68-72]. In addition, significant downregulation of *APOE* in astrocytes from

36

752   LOAD brains, although not sex-dependent, was reported by scRNA-seq analysis. This evidence

753   is consistent with the trend we found of more closed chromatin in LOAD samples, providing

754   functional validation to our result[28]. In summary, we proposed molecular insights based on

755   chromatin structure that may explain, at least in part, some aspects related to the role of *APOE* in

756   LOAD.

757   In conclusion, this LOAD genomic research pioneers the approach of brain cell-type

758   specific chromatin accessibility profiling and lays the foundation for additional sorted- and

759   single- nuclei 'omics analyses in LOAD. Our outcomes warrant further investigations using a

760   larger sample size to enhance the discovery smaller LOAD-associated effects on chromatin

761   accessibility and to allow the utilization of a single peakset and set of covariates to perform

762   multiple testing. Future and ongoing studies using even more advanced single-cell technologies

763   will generate complementary 'omics datasets with finer cell-type resolution from larger well-

764   characterized LOAD cohorts. Data sharing via publicly available portals, such as Accelerating

765   Medicines Partnership-Alzheimer's Disease (AMP-AD), will facilitate integrative single-cells

766   'omics towards moving forward our understanding of the underpinning genetic drivers and

767   molecular mechanisms of LOAD.

768

769   **Supplemental Data:**

770   Supplemental Data include 12 figures and 15 tables.

771

772   **Declaration of Interests:**

773   The authors declare no competing interests.

774

798   performed QC and covariate analysis. L.S. performed alignment, peak quantification, QC

799   assessment, differential chromatin, and motif analysis. R.G. advised on the motif analysis. J.G.

800   and H.F. performed the snRNA-seq data analysis and differential expression analysis. J.L.

801   assisted with the QC metric analysis. K.S. assisted with whole genome genotyping. A.A.K.,

802   G.E.C. and O.C-F. planned and supervised the work. G.E.C. and O.C-F. obtained funding

803   support. All authors discussed and interpreted the results. J.B., L.S., A.S., Y.Y., M.E.G., J.G.,

804   A.A.K., G.E.C., and O.C-F. designed the study and wrote the manuscript. All authors read and

805   approved the manuscript.

806

807   **Data Availabillity:**

808   The ATAC-Sequencing and single-nucleus RNA-Sequencing data are available at Synapse

809   (https://www.synapse.org/#!Synapse:syn20755767). The DOI for this

810   dataset is 10.7303/syn20755767. The data are available under controlled use conditions.

811

812   **References:**

813
814   1.   Lambert, J.-C.*, et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility
815   loci for Alzheimer's disease. *Nat Genet* **45**, 1452-1458 (2013).
816   2.   Naj, A.C.*, et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are
817   associated with late-onset Alzheimer's disease. *Nat Genet* **43**, 436-441.
818   3.   Harold, D.*, et al.* Genome-wide association study identifies variants at CLU and PICALM
819   associated with Alzheimer's disease. *Nat Genet* **41**, 1088-1093 (2009).
820   4.   Hollingworth, P.*, et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33
821   and CD2AP are associated with Alzheimer's disease. *Nat Genet* **43**, 429-435 (2011).
822   5.   Jansen, I.*, et al.* Genetic meta-analysis identifies 9 novel loci and functional pathways for
823   Alzheimers disease risk. *bioRxiv* (2018).
824   6.   Marioni, R.*, et al.* GWAS on family history of Alzheimer's disease. *bioRxiv* (2018).
825   7.   Kunkle, B.W.*, et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies
826   new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430
827   (2019).
828   8.   Linnertz, C.*, et al.* The cis-regulatory effect of an Alzheimer's disease-associated poly-T
829   locus on expression of TOMM40 and apolipoprotein E genes. *Alzheimers Dement* **10**, 541-551
830   (2014).

831 9.   Matsui, T.*, et al.* Expression of APP pathway mRNAs and proteins in Alzheimer's
832   disease. *Brain Res* **1161**, 116-123 (2007).
833 10.   Zarow, C. & Victoroff, J. Increased apolipoprotein E mRNA in the hippocampus in
834   Alzheimer disease and in rats after entorhinal cortex lesioning. *Exp Neurol* **149**, 79-86 (1998).
835 11.   Gibbs, J.R.*, et al.* Abundant quantitative trait loci exist for DNA methylation and gene
836   expression in human brain. *PLoS Genet* **6**, e1000952.
837 12.   Allen, M.*, et al.* Novel late-onset Alzheimer disease loci variants associate with brain
838   gene expression. *Neurology* **79**, 221-228 (2012).
839 13.   Zou, F.*, et al.* Brain expression genome-wide association study (eGWAS) identifies
840   human disease-associated variants. *PLoS Genet* **8**, e1002707 (2012).
841 14.   Karch, C.M.*, et al.* Expression of novel Alzheimer's disease risk genes in control and
842   Alzheimer's disease brains. *PLoS One* **7**, e50976 (2012).
843 15.   Karch, C.M., Ezerskiy, L.A., Bertelsen, S., Alzheimer's Disease Genetics, C. & Goate,
844   A.M. Alzheimer's Disease Risk Polymorphisms Regulate Gene Expression in the ZCWPW1 and
845   the CELF1 Loci. *PLoS One* **11**, e0148717 (2016).
846 16.   Singleton, A., Myers, A. & Hardy, J. The law of mass action applied to
847   neurodegenerative disease: a hypothesis concerning the etiology and pathogenesis of complex
848   diseases. *Hum Mol Genet* **13 Spec No 1**, R123-126 (2004).
849 17.   Smith, A.R.*, et al.* Increased DNA methylation near TREM2 is consistently seen in the
850   superior temporal gyrus in Alzheimer's disease brain. *Neurobiol Aging* **47**, 35-40 (2016).
851 18.   Zhao, J.*, et al.* A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's
852   disease. *Alzheimers Dement* **13**, 674-688 (2017).
853 19.   Lunnon, K.*, et al.* Methylomic profiling implicates cortical deregulation of ANK1 in
854   Alzheimer's disease. *Nat Neurosci* **17**, 1164-1170 (2014).
855 20.   De Jager, P.L.*, et al.* Alzheimer's disease: early alterations in brain DNA methylation at
856   ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci* **17**, 1156-1163 (2014).
857 21.   Chibnik, L.B.*, et al.* Alzheimer's loci: epigenetic associations and interaction with genetic
858   factors. *Ann Clin Transl Neurol* **2**, 636-647 (2015).
859 22.   Yu, L.*, et al.* Association of Brain DNA methylation in SORL1, ABCA7, HLA-DRB5,
860   SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease. *JAMA Neurol* **72**, 15-24
861   (2015).
862 23.   Watson, C.T.*, et al.* Genome-wide DNA methylation profiling in the superior temporal
863   gyrus reveals epigenetic signatures associated with Alzheimer's disease. *Genome Med* **8**, 5
864   (2016).
865 24.   Nativio, R.*, et al.* Dysregulation of the epigenetic landscape of normal aging in
866   Alzheimer's disease. *Nat Neurosci* **21**, 497-505 (2018).
867 25.   Matevossian, A. & Akbarian, S. Neuronal nuclei isolation from human postmortem brain
868   tissue. *J Vis Exp* (2008).
869 26.   Gasparoni, G.*, et al.* DNA methylation analysis on purified neurons and glia dissects age
870   and Alzheimer's disease-specific changes in the human cortex. *Epigenetics Chromatin* **11**, 41
871   (2018).
872 27.   Tulloch, J.*, et al.* Glia-specific APOE epigenetic changes in the Alzheimer's disease
873   brain. *Brain Res* **1698**, 179-186 (2018).
874 28.   Mathys, H.*, et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**,
875   332-337 (2019).
876 29.   Grubman, A.*, et al.* A single-cell atlas of entorhinal cortex from individuals with
877   Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat Neurosci* **22**,
878   2087-2097 (2019).
879 30.   Garwood, C.J.*, et al.* Review: Astrocytes in Alzheimer's disease and other age-
880   associated dementias: a supporting player with a central role. *Neuropathol Appl Neurobiol* **43**,
881   281-298 (2017).

882   31.   Perl, D.P. Neuropathology of Alzheimer's disease. *Mt Sinai J Med* **77**, 32-42 (2010).
883   32.   McLean, C.Y.*, et al.* GREAT improves functional interpretation of cis-regulatory regions.
884   *Nat Biotechnol* **28**, 495-501 (2010).
885   33.   Nagarajan, R.*, et al.* EGR2 mutations in inherited neuropathies dominant-negatively
886   inhibit myelin gene expression. *Neuron* **30**, 355-368 (2001).
887   34.   Ma, K., Zheng, S. & Zuo, Z. The transcription factor regulatory factor X1 increases the
888   expression of neuronal glutamate transporter type 3. *J Biol Chem* **281**, 21250-21255 (2006).
889   35.   Harrington, A.J.*, et al.* MEF2C regulates cortical inhibitory and excitatory synapses and
890   behaviors relevant to neurodevelopmental disorders. *Elife* **5** (2016).
891   36.   Wegner, M. SOX after SOX: SOXession regulates neurogenesis. *Genes Dev* **25**, 2423-
892   2428 (2011).
893   37.   Fullard, J.F.*, et al.* Open chromatin profiling of human postmortem brain infers functional
894   roles for non-coding schizophrenia loci. *Hum Mol Genet* **26**, 1942-1951 (2017).
895   38.   Wagner, N.*, et al.* The Wilms' tumor suppressor Wt1 is associated with the differentiation
896   of retinoblastoma cells. *Cell Growth Differ* **13**, 297-305 (2002).
897   39.   Shi, Z., Shen, T., Liu, Y., Huang, Y. & Jiao, J. Retinoic acid receptor gamma (Rarg) and
898   nuclear receptor subfamily 5, group A, member 2 (Nr5a2) promote conversion of fibroblasts to
899   functional neurons. *J Biol Chem* **289**, 6415-6428 (2014).
900   40.   Duclot, F. & Kabbaj, M. The Role of Early Growth Response 1 (EGR1) in Brain Plasticity
901   and Neuropsychiatric Disorders. *Front Behav Neurosci* **11**, 35 (2017).
902   41.   Parker-Katiraee, L.*, et al.* Identification of the imprinted KLF14 transcription factor
903   undergoing human-specific accelerated evolution. *PLoS Genet* **3**, e65 (2007).
904   42.   Riedel, B.C., Thompson, P.M. & Brinton, R.D. Age, APOE and sex: Triad of risk of
905   Alzheimer's disease. *J Steroid Biochem Mol Biol* **160**, 134-147 (2016).
906   43.   Poche, R.A.*, et al.* RONIN Is an Essential Transcriptional Regulator of Genes Required
907   for Mitochondrial Function in the Developing Retina. *Cell Rep* **14**, 1684-1697 (2016).
908   44.   Vong, K.I., Leung, C.K., Behringer, R.R. & Kwan, K.M. Sox9 is critical for suppression of
909   neurogenesis but not initiation of gliogenesis in the cerebellum. *Mol Brain* **8**, 25 (2015).
910   45.   Kolsch, H.*, et al.* Influence of SORL1 gene variants: association with CSF amyloid-beta
911   products in probable Alzheimer's disease. *Neuroscience letters* **440**, 68-71 (2008).
912   46.   Demir, O.*, et al.* ETS-domain transcription factor Elk-1 mediates neuronal survival: SMN
913   as a potential target. *Biochim Biophys Acta* **1812**, 652-662 (2011).
914   47.   Kent, W.J.*, et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006
915   (2002).
916   48.   Trynka, G. & Raychaudhuri, S. Using chromatin marks to interpret and localize genetic
917   associations to complex human traits and diseases. *Curr Opin Genet Dev* **23**, 635-641 (2013).
918   49.   Trynka, G.*, et al.* Chromatin marks identify critical cell types for fine mapping complex
919   trait variants. *Nat Genet* **45**, 124-130 (2013).
920   50.   Maurano, M.T.*, et al.* Systematic localization of common disease-associated variation in
921   regulatory DNA. *Science* **337**, 1190-1195 (2012).
922   51.   Cockerill, P.N. Structure and function of active chromatin and DNase I hypersensitive
923   sites. *FEBS J* **278**, 2182-2210 (2011).
924   52.   Gross, D.S. & Garrard, W.T. Nuclease hypersensitive sites in chromatin. *Annu Rev*
925   *Biochem* **57**, 159-197 (1988).
926   53.   Song, L.*, et al.* Open chromatin defined by DNaseI and FAIRE identifies regulatory
927   elements that shape cell-type identity. *Genome Res* **21**, 1757-1767 (2011).
928   54.   Thurman, R.E.*, et al.* The accessible chromatin landscape of the human genome. *Nature*
929   **489**, 75-82 (2012).
930   55.   Won, H.*, et al.* Chromosome conformation elucidates regulatory relationships in
931   developing human brain. *Nature* **538**, 523-527 (2016).

932   56.   Ramamurthy, E*., et al.* Cell type-specific histone acetylation profiling of Alzheimer's
933   Disease subjects and integration with genetics. *bioRxiv*, 2020.2003.2026.010330 (2020).
934   57.   Gjoneska, E*., et al.* Conserved epigenomic signals in mice and humans reveal immune
935   basis of Alzheimer's disease. *Nature* **518**, 365-369 (2015).
936   58.   Morabito, S., Miyoshi, E., Michael, N. & Swarup, V. Integrative genomics approach
937   identifies conserved transcriptomic networks in Alzheimer's disease. *Hum Mol Genet* (2020).
938   59.   Nott, A*., et al.* Brain cell type-specific enhancer-promoter interactome maps and
939   disease-risk association. *Science* **366**, 1134-1139 (2019).
940   60.   Ferretti, M.T*., et al.* Sex differences in Alzheimer disease - the gateway to precision
941   medicine. *Nat Rev Neurol* **14**, 457-469 (2018).
942   61.   Babapour Mofrad, R. & van der Flier, W.M. Nature and implications of sex differences in
943   AD pathology. *Nat Rev Neurol* **15**, 6-8 (2019).
944   62.   Buckley, R.F., Waller, M., Masters, C.L. & Dobson, A. To What Extent Does Age at
945   Death Account for Sex Differences in Rates of Mortality From Alzheimer Disease? *Am J*
946   *Epidemiol* **188**, 1213-1223 (2019).
947   63.   Jack, C.R., Jr*., et al.* Age, Sex, and APOE epsilon4 Effects on Memory, Brain Structure,
948   and beta-Amyloid Across the Adult Life Span. *JAMA Neurol* **72**, 511-519 (2015).
949   64.   Corder, E.H*., et al.* Gene dose of apolipoprotein E type 4 allele and the risk of
950   Alzheimer's disease in late onset families. *Science* **261**, 921-923 (1993).
951   65.   Saunders, A.M*., et al.* Association of apolipoprotein E allele epsilon 4 with late-onset
952   familial and sporadic Alzheimer's disease. *Neurology* **43**, 1467-1472 (1993).
953   66.   Lambert, J.C*., et al.* Genome-wide association study identifies variants at CLU and CR1
954   associated with Alzheimer's disease. *Nat Genet* **41**, 1094-1099 (2009).
955   67.   Lambert, J.C*., et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility
956   loci for Alzheimer's disease. *Nat Genet* **45**, 1452-1458 (2013).
957   68.   Bretsky, P.M*., et al.* Evidence for an interaction between apolipoprotein E genotype,
958   gender, and Alzheimer disease. *Alzheimer Dis Assoc Disord* **13**, 216-221 (1999).
959   69.   Payami, H*., et al.* Alzheimer's disease, apolipoprotein E4, and gender. *JAMA* **271**, 1316-
960   1317 (1994).
961   70.   Poirier, J*., et al.* Apolipoprotein E polymorphism and Alzheimer's disease. *Lancet* **342**,
962   697-699 (1993).
963   71.   Altmann, A., Tian, L., Henderson, V.W., Greicius, M.D. & Alzheimer's Disease
964   Neuroimaging Initiative, I. Sex modifies the APOE-related risk of developing Alzheimer disease.
965   *Ann Neurol* **75**, 563-573 (2014).
966   72.   Neu, S.C*., et al.* Apolipoprotein E Genotype and Sex Risk Factors for Alzheimer
967   Disease: A Meta-analysis. *JAMA Neurol* **74**, 1178-1189 (2017).
968   73.   Jiang, Y., Matevossian, A., Huang, H.S., Straubhaar, J. & Akbarian, S. Isolation of
969   neuronal chromatin from brain tissue. *BMC Neurosci* **9**, 42 (2008).
970   74.   Marzluff, W.F. Preparation of active nuclei. *Methods Enzymol* **181**, 30-36 (1990).
971   75.   Corces, M.R*., et al.* An improved ATAC-seq protocol reduces background and enables
972   interrogation of frozen tissues. *Nat Methods* **14**, 959-962 (2017).
973   76.   Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
974   **9**, 357-359 (2012).
975   77.   Fein, F.S., Cho, S., Zola, B.E., Miller, B. & Factor, S.M. Cardiac pathology in the
976   hypertensive diabetic rat. Biventricular damage with right ventricular predominance. *Am J Pathol*
977   **134**, 1159-1166 (1989).
978   78.   Landt, S.G*., et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE
979   consortia. *Genome Res* **22**, 1813-1831 (2012).
980   79.   van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained
981   Equations in R. *J Stat Softw* **45**, 1-67 (2011).

982   80.   Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for
983 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140
984 (2010).
985   81.   Robinson, M.D. & Oshlack, A. A scaling normalization method for differential expression
986 analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010).
987   82.   Heinz, S.*, et al.* Simple combinations of lineage-determining transcription factors prime
988 cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589
989 (2010).
990   83.   Bailey, T.L., Johnson, J., Grant, C.E. & Noble, W.S. The MEME Suite. *Nucleic Acids Res*
991 **43**, W39-49 (2015).
992   84.   Bennett, D.A., Schneider, J.A., Arvanitakis, Z. & Wilson, R.S. Overview and findings
993 from the religious orders study. *Curr Alzheimer Res* **9**, 628-645 (2012).
994   85.   Bennett, D.A.*, et al.* Overview and findings from the rush Memory and Aging Project.
995 *Curr Alzheimer Res* **9**, 646-663 (2012).
996   86.   Mostafavi, S.*, et al.* A molecular network of the aging human brain provides insights into
997 the pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci* **21**, 811-819 (2018).
998   87.   SageBionetworks. AMP-AD Knowledge Portal: ROSMAP RNA-Seq. Available at:
999 https://www.synapse.org/#!Synapse:syn3388564.
1000   88.   McCarthy, D.J., Campbell, K.R., Lun, A.T. & Wills, Q.F. Scater: pre-processing, quality
1001 control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**,
1002 1179-1186 (2017).
1003   89.   Stuart, T.*, et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902
1004 e1821 (2019).
1005   90.   Aran, D.*, et al.* Reference-based analysis of lung single-cell sequencing reveals a
1006 transitional profibrotic macrophage. *Nat Immunol* **20**, 163-172 (2019).
1007   91.   Mabbott, N.A., Baillie, J.K., Brown, H., Freeman, T.C. & Hume, D.A. An expression atlas
1008 of human primary cells: inference of gene function from coexpression networks. *BMC Genomics*
1009 **14**, 632 (2013).
1010