Differential transcript usage unravels transcriptional alterations in Alzheimer's disease human brains

Diego Marques-Coelho[1,2], Lukas Iohan da Cruz Carvalho[1,2], Jean-Charles Lambert[3], Marcos Romualdo Costa[1,3,*]

[1] Brain Institute, Federal University of Rio Grande do Norte, Av. Nascimento de Castro 2155 Natal, Brazil

[2] Bioinformatics Multidisciplinary Environment (BioME), Federal University of Rio Grande do Norte

[3] Unité INSERM 1167, RID-AGE-Risk Factors and Molecular Determinants of Aging-Related Diseases, Institut Pasteur de Lille, University of Lille, U1167-Excellence Laboratory LabEx DISTALZ, BP 245, 1 rue du Professeur Calmette, 59019 Lille Cedex, France

* Corresponding author: marcos.costa@pasteur-lille.fr

Abstract

Alzheimer's disease (AD) is the leading cause of dementia in aging individuals. However pathophysiological processes involved in the brain are still poorly understood. Among numerous strategies, a comprehensive overview of gene expression alterations in the diseased brain has been proposed to help for a better knowledge of the disease processes. In this work, we compared the differential expression of genes in different regions (temporal and frontal lobes) of the brain of healthy and AD adult subjects using data from three large studies: MAYO Clinic; Mount Sinai Brain Bank (MSBB) and ROSMAP. We show that gene expression is dramatically altered in the temporal lobe (TL), but not in the frontal lobe (FL), suggesting regional specificities in transcriptional alterations. Moreover, we show that a large set of genes undetected in the classical differential expression (DEG) analysis show alterations in the transcript usage ratio in AD, i.e. differential isoform fractions. Interestingly, the number of genes with differential transcript usage (gDTUs) is highest at advanced Braak stages and correlates with altered expression of genes associated with the splicing machinery. Using single-cell RNA sequencing (scRNAseq) data, we further assigned DEGs and gDTUs to individual cell types of the adult brain and observed similar signatures per cell type as to other two scRNAseq studies in AD. Lastly, we show that 52 out of 116 genes located in the loci associated with AD risk and expressed in the adult human brain present DEG and/or gDTUs in the TL, whereas only 13 are altered in the FL of AD patients. Among those altered genes, we detect an increased expression of APP isoforms containing the Kunitz protease inhibitor (KPI) domain, which is associated with plaque deposition. Altogether, our work proposes a novel integrative strategy to analyze transcriptional data in AD based on both gene/transcript expression and cell-type specificities.

Keywords: Alzheimer disease; transcriptomics; gene expression; isoform switch; alternative splicing; APP

Introduction

Alzheimer's disease (AD) is the most prevalent type of dementia and causes a progressive cognitive decline, for which there is no effective treatment or cure. Changes in gene expression characterize diverse human diseases and have been successfully used to predict molecular and cellular mechanisms associated with pathological processes in numerous diseases (Masters et al., 2015). Although expression analyses in brain diseases are generally limited by tissue availability, RNA sequencing (RNAseq) data have been generated from *postmortem* brain samples of healthy and AD individuals (Allen et al., 2016; De Jager et al., 2018; Wang et al., 2018). However, a comprehensive description of the transcriptional alterations in the AD brain remains elusive.

Recent work has started to address this important gap in the study of AD pathology using bulk brain tissue RNA sequencing (RNAseq) (T. Raj et al., 2018) or single-cells RNA sequencing (scRNAseq) (Grubman et al., 2019; Mathys et al., 2019). However, these studies have focused on samples obtained from different brain regions, namely the frontal cortex (Mathys et al., 2019; T. Raj et al., 2018) and entorhinal cortex (Grubman et al., 2019), which could lead to important discrepancies in the results. Indeed, the AD pathology shows a progressive affection of different brain regions, characterized at early stages by the presence of TAU protein inclusions in the locus coeruleus, the transentorhinal and entorhinal regions (stages I and II). This is followed by the presence of tau inclusions in the hippocampal formation and some parts of the neocortex (stages III and IV), followed by large parts of the neocortex (stages V and VI)(Braak & Braak, 1991). Accordingly, a recent study has shown that changes in protein expression are much more prominent in areas affected at early and intermediate stages, such as the hippocampus, entorhinal cortex and cingulate cortex in the temporal lobe, compared to other brain regions affected only at late stages of AD pathology, such as sensory cortex, motor cortex and cerebellum (Xu et al., 2019).

Another important aspect to consider is the power of transcriptional analysis based solely on the identification of differentially expressed genes (DEG), which fails to detect dynamics in the expression of multiple related transcripts (Yi et al., 2018). Recently, new approaches using transcripts-level analysis, so called differential transcript usage (DTU), enables identification of alternative splicing and isoform switches with prediction of functional consequences (Anders et al., 2012; Vitting-Seerup & Sandelin, 2019). Therefore, important

transcriptional alterations in the AD brain could occur at the transcript level and be overlooked in classical DEG analyses.

Here, we took advantage of three available RNAseq datasets, generated using samples from different brain regions, to systematically compare transcriptional alterations (DEG and DTU) in AD. In the Mayo's clinic study, both the temporal cortex and cerebellum were used to obtain bulk RNAseq (Allen et al., 2016). In the Religious Orders Study (ROS) and Memory and Aging Project (MAP), henceforth called ROSMAP, the dorsolateral prefrontal cortex was used (De Jager et al., 2018). Finally, in the Mount Sinai/JJ Peters VA Medical Center Brain Bank (MSBB), 4 different Brodmann areas of the brain were studied: areas 22 and 36 from the temporal lobe, areas 10 and 44 in the frontal lobe (Wang et al., 2018). We also added another level of complexity using an indirect approach to assign DEGs and gDTUs to unique cell types in order to identify AD transcriptional signatures for neural cells, microglia and endothelial cells. Finally, we linked these transcriptional alterations with putative AD risk genes.

Materials and Methods

Bulk RNAseq data from human control and disease banks

RNAseq datasets obtained from different brain regions were used (Mayo: Allen et al, 2016; MSBB: Wang et al, 2018; ROSMAP: De Jager et al, 2018). Metadata obtained from each study was used to classify patients into Control and Alzheimer's disease groups. Datasets were downloaded from AMP-AD Knowledge Portal (https://www.synapse.org) following all terms and conditions to use the data. The area of the brain analyzed and the number of individuals per condition were the following: Mayo - Temporal cortex, which neuroanatomically subdivides into the inferior, middle, and superior temporal gyri (STG), and cytoarchitectonically can be subdivided into Brodmann areas (BM, instead of BA) 20/21/22/41/42 (Strotzer, 2009), N=160 subjects (82 AD and 78 controls); MSBB - BM22, which is part of the Wernicke's area in the STG, N=159 subjects (98 AD and 61 controls); MSBB BM36, corresponding to the lateral perirhinal cortex, N=154 subjects (88 AD and 64 controls); MSBB BM10, corresponding to the anterior prefrontal cortex, N=176 subjects (105 AD and 71 controls); MSBB BM44, corresponding to the inferior frontal gyrus, N=153 subjects (90 AD and 63 controls); and ROSMAP - Dorsolateral prefrontal cortex (DLPFC), containing BM46 and part of BM9, N=423 subjects (222 AD and 201 controls). Unless stated otherwise, data obtained from different analyses were grouped in "temporal lobe" (TL) - Mayo, MSBB BM22 and MSBB BM26; or "frontal lobe" (FL) - ROSMAP, MSBB BM10 and MSBB BM44. For single-cell RNA sequencing (scRNAseq) analyses, we used processed data obtained from the middle temporal gyrus (MTG), available at the Allen Brain Atlas consortium (https://celltypes.brain-map.org/rnaseq).

Differential gene expression analyses

Using human GRCh38 cDNA release 94 (ftp://ftp.ensembl.org/pub/release-94) as reference, we built an index to align all our fastq files. Next, we used pseudoaligner Kallisto (Bray et al., 2016); version 0.43.1) with our pre-built index to align fastq files.

Differentially expressed genes (DEGs) were identified using differential gene expression at transcript-level using DESeq2 R library (Love et al., 2014, 2019). To facilitate kallisto output import, transcript-level estimated counts, length and abundance was extracted using *tximport* function (Soneson et al., 2016). As described by Michael Love group, transcript-

level differential gene expression enhances analysis resolution (Love et al., 2019). Using *DESeqDataSetFromTximport*, a DESeq2 object was created and filtered using rows with sum of all counts bigger than 10. Next, *DESeq* function was used with default values. Using the *results* function, we selected all genes with False Discovery Rate (FDR) < 0.01 and Fold Change (FC) > 1.3.

Differential Transcript Usage (DTU) analysis was performed using the R library IsoformSwitchAnalyzeR (Vitting-Seerup & Sandelin, 2019). Following pipeline instructions, kallisto abundance tables were imported using *importIsoformExpression* and *importRdata* functions to create a *switchAnalyzeRlist* object. Same cDNA release used in kallisto alignment and correspondent annotation (ftp://ftp.ensembl.org/pub/release-94/gtf/homo_sapiens/Homo_sapiens.GRCh38.94.chr_patch_hapl_scaff.gtf.gz) were applied as input. We filtered data using a gene expression cut off = 10, isoform expression cut off = 3, differential isoform fraction (dIF) cut off = 0.05 and removed single isoform genes. Although DEXSeq is recommended to test differential isoform usage, it does not work efficiently for large datasets (more than 100 samples; (Anders et al., 2012)). For that reason, we chose *isoformSwitchAnalysisPart1* function using DRIMSeq (Soneson et al., 2016) to test differential transcript usage. Using part1 fasta files, all external analysis was performed and used as input to *isoformSwitchAnalysisPart2* function. We used CPC2, Pfam, SignalIP and Netsurfp2 as indicated in the pipeline. Next, we performed a confirmation stage using stageR (Van den Berge et al., 2017) to generate isoforms overall false discovery rate (OFDR). We selected all isoforms with OFDR < 0.01 and dIF > 0.05.

To evaluate if intersections among different datasets were statistically relevant, we calculated the significance of overlap between every two sets using the hypergeometric test (phyper). Our universe set consists in genes with FDR < 0.1 within each dataset.

Single-cell RNAseq

Using R library *seurat*, we created a seurat object (*CreateSeuratObject*), normalized data (*NormalizeData*), found variable genes (*FindVariableFeatures*) and rescaled data using a linear model (*ScaleData, use.umi = F*). After that, we generated 50 PC's (*RunPCA*) but only used 35 of them based on the PC's visualization distribution (*ElbowPlot*). Since Allen data was already annotated, we only used tSNE (*RunTSNE*) to facilitate visualization. A group

classified as "None" by Allen metadata were removed from our analysis. This strategy generated 7 main different cell types: Astrocytes, Endothelial cells, Glutamatergic Neurons, GABAergic Neurons, Microglia, Oligodendrocytes and oligodendrocyte precursor cells (OPCs). To assign genes to specific cell-types, we used the AverageExpression function. Using pct.exp bigger than 0.1, we created a list of genes that were expressed by each cell type.

Gene-set enrichment analysis (GSEA)

For gene ontology analysis, R library *gprofiler2* was used. Using *gost* function, correction_method="fdr" and significant=TRUE. As a threshold, we choose prediction > 0.03, intersection_size > 3, p_value < 0.01. We used Gene Ontology (GO or by branch GO:MF, GO:BP, GO:CC), Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome (REAC), WikiPathways (WP), TRANSFAC (TF), miRTarBase (MIRNA), Human Protein Atlas (HPA), CORUM (CORUM), Human phenotype ontology (HP) as sources. For improved visualization, we plotted results only for GO:BP, GO:CC and KEGG .

Selection of splicing-associated genes

To select splicing related genes, we searched for terms containing the words "splicing" or "spliceosome" in gProfiler bank (https://biit.cs.ut.ee/gprofiler/gost). Taking only GO and WP datasets, 25 terms and 441 genes related to those terms were selected (Supplementary table 8).

Selection of AD risk genes

Late-onset AD risk genes were curated by Dourlen et al. (2019). Briefly, AD-associated loci were selected from genetic studies, mainly GWAS and WES [1–5, 7–25]. Genes in these loci were either indicated in the publications or determined based on the regional association plots provided in the publications assuming that the functional risk variants are located in the vicinity of the SNP producing the top signal and taking into account the linkage disequilibrium patterns and the recombination peaks within the loci of interest. Early-onset AD risk genes were APP, PSEN1 and PSEN2. The complete list of genes is described in the Supplementary Table 9.

Results

Regional transcriptional alterations in the AD brain correlates with pathological progression

Several consortia have generated data on the transcriptional profile of the brain of individuals with a clinical and/or pathological diagnostic of AD. However, different brain regions were used to obtain the major datasets available, namely the Mayo Clinic RNAseq (Allen et al., 2016), the Mount Sinai/JJ Peters VA Medical Center Brain Bank (MSBB) (Wang et al., 2018) and the Religious Orders Study and the Rush Memory and Aging Project (ROSMAP) (De Jager et al., 2018), that may cause strong discrepancies in the results considering the regional progression of AD pathology (Braak and Braak, 1991). We hypothesize that the transcriptional profiles of the temporal lobe (TL), e.g. hippocampus and surrounding structures including the perirhinal, parahippocampal and entorhinal neocortical regions, are likely more relevant than those observed in frontal lobe (FL) to detect changes in gene expression associated with AD pathology. To test this possibility, we performed an evaluation of differentially expressed genes (DEGs) in the FL (ROSMAP and MSBB_FL - BM10 and BM22) or TL (Mayo and MSBB_TL - BM36 and BM44) of AD and control subjects (Figure 1).

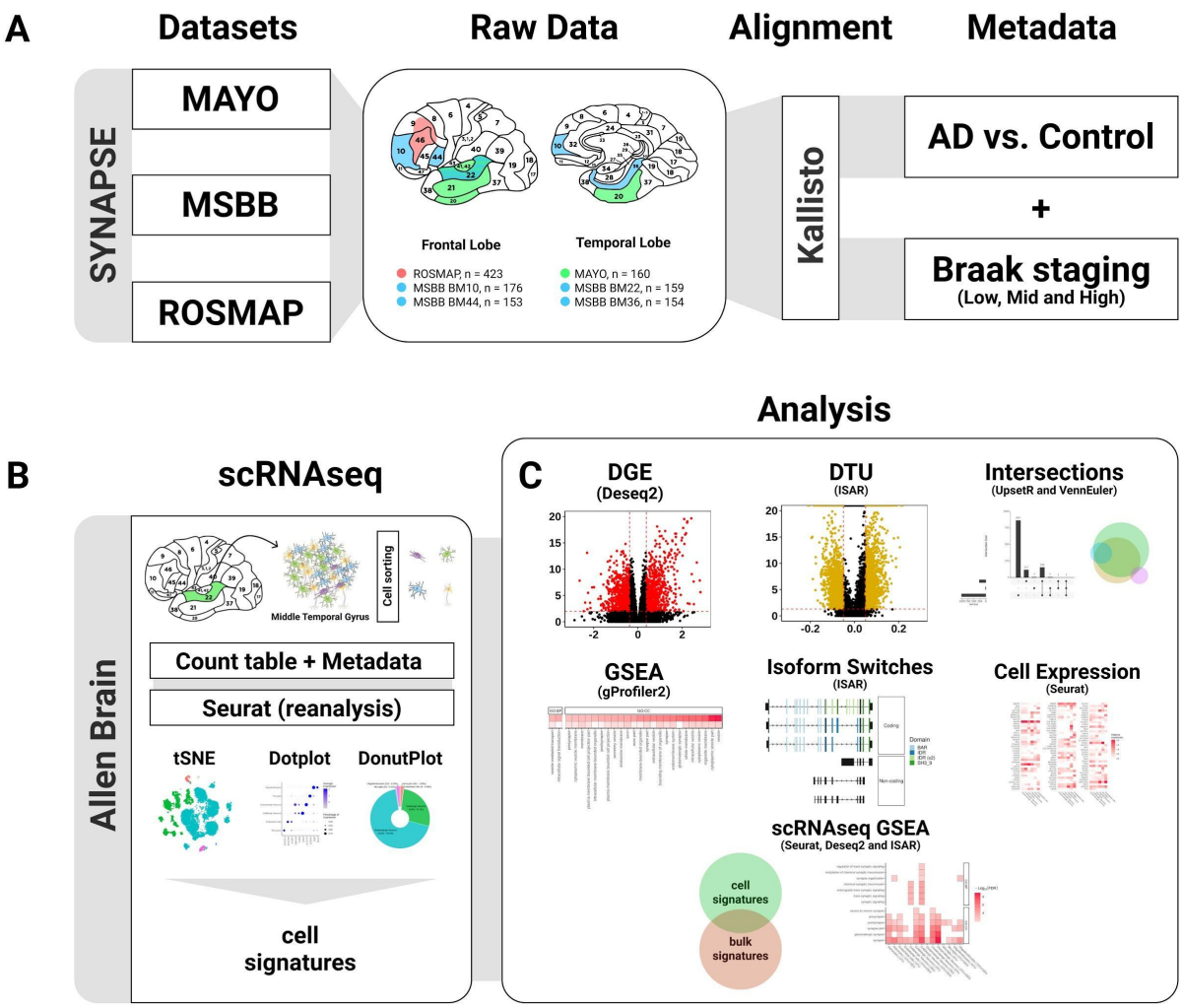


Figure 1. Schematic summary of the methodology. Datasets obtained from 3 consortia (Mayo, MSBB and ROSMAP) were grouped according to the brain region sampled in frontal lobe (FL) or temporal lobe (TL). Next, RNAseq data was pseudo-aligned using Kallisto and analyzed using the packages from R (version 3.6) DESeq2, IsoformSwitchAnalyzeR (ISAR) and gene-set enrichment analysis (GSEA). Assignment of differentially expressed genes or isoform switches to specific cell types was performed indirectly using scRNAseq signatures obtained from the Allen Brain Atlas transcriptomic data.

We found 3,348 and 2,172 DEGs in the TL of AD individuals compared to controls in the MSBB_TL and Mayo datasets, respectively (Figure 2A-B; Supplementary table 1). Of those DEGs, 734 genes were identified in both datasets (15,33% of overlap; p= 8.56 x 10^{-59},

hypergeometric test). In contrast, only 327 (MSBB_FL) and 209 (ROSMAP) DEGs were detected in the frontal lobe of AD subjects compared to controls, 31 genes were found in both datasets (7,34% of overlap; $p = 1.67 \times 10^{-14}$, hypergeometric test) (Figure 2A-B; Supplementary table 1). These differences in the number of DEGs detected in the FL and TL cannot be attributed to lack of statistical power, since the number of samples in the FL is larger than in the TL groups. Notably, 62.5% of the DEGs detected in the FL are also detected in the TL (Figure 2B), suggesting that transcriptional changes in the FL affect similar genes compared to TL.

Next, we performed gene set enrichment analyses for DEGs in the intersection MAYO/MSBB TL (temporal lobe intersection - TLI) or ROSMAP/MSBB FL (frontal lobe intersection - FLI). We found that genes altered in the temporal lobe are enriched for terms (GO:BP, GO:CC and KEGG) associated with important biological processes, including cell-signaling pathways, cell-cell signaling and vesicle transport , whereas the DEGs in the FL were not significantly enriched for the same terms (Figure 2C; see also Supplementary table 2 for detailed information and GSEA using alternative sources). Together, these data indicate that changes in gene expression are much more significant in the TL as compared to the FL. Importantly, this is true even in tissues obtained from the same donors, as it is the case for the MSBB_TL and MSBB_FL samples.

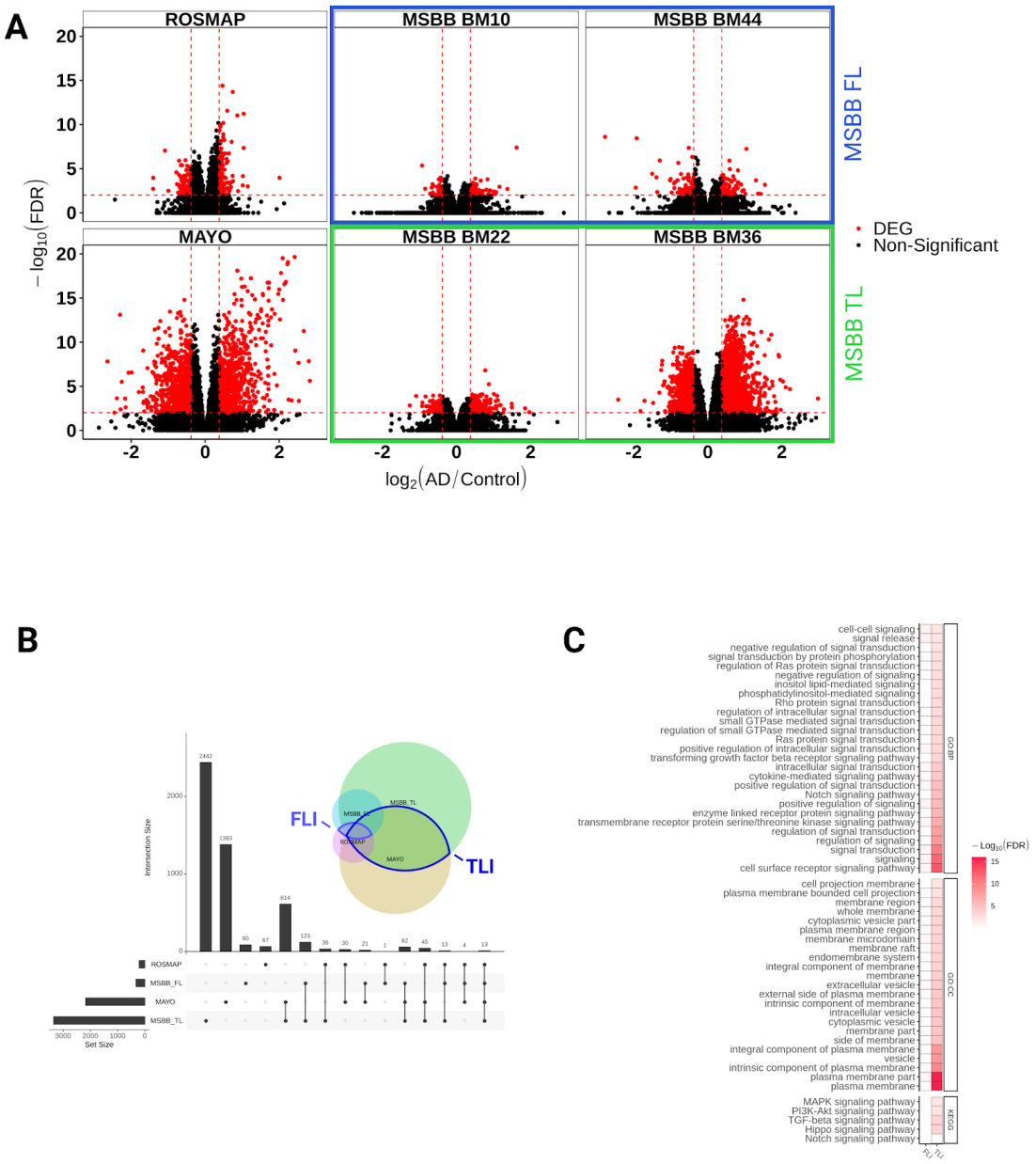


Figure 2. Gene expression alterations are more prominent in the temporal than frontal lobe of AD patients. A) Volcano plots showing differentially expressed genes (DEG, red dots; FC > 1.3 and FDR < 0.01) in the frontal lobe (ROSMAP and MSBB FL - BM10 and BM44) and temporal lobe (MAYO and MSBB TL - BM22 and BM36). Non-significant genes are represented by black dots. Note the large number of DEGs in the MAYO and MSBB 36 datasets. B) Venn diagram showing the number of DEGs identified in the different datasets. The intersection between MAYO and MSB TL is highlighted in red (TLI - temporal lobe intersection). C) Gene ontology terms enriched for DEGs in the TL or FL intersections (TLI and FLI, respectively).

Differential transcript usage analyses reveal novel genes associated with AD pathology

Gene-level expression analysis lacks the sensitivity to detect possible changes at the transcript-level caused, for example, by alterations in alternative splicing (Vitting-Seerup & Sandelin, 2017; Yi et al., 2018). To overcome this limitation, we used differential transcript usage (DTU) analysis to identify additional alterations of gene expression in the AD brains compared to controls. We observed 2,509 and 1,843 genes with differential transcript usage (gDTU) in the temporal lobe of AD brains studied in the Mayo and MSBB datasets, respectively (Figure 3A-B; Supplementary table 1). Similar to what we observed for DEGs, a much smaller number of gDTUs were detected in the frontal lobe, both in ROSMAP and MSBB studies (59 and 855 genes with transcripts altered, respectively). We found 435 gDTUS in TLI (11,1% of overlap; $p= 6.16 \times 10^{-25}$, hypergeometric test) and 13 gDTUs in FLI (1,47% of overlap; $p= 2.56 \times 10^{-3}$, hypergeometric test). In both TLI and FLI, most gDTUs were not DEGs (TL - 34 gDTUs that are DEGs out of 435 gDTUs; FL - gDTUs and DEGs don't have intersection).

Gene set enrichment in the TLI using gDTUs revealed different biological processes than those identified using only DEGs (Figure 3C-D; Supplementary table 2). Noteworthy, despite the widely documented alterations in synapse functions in AD (Canter et al., 2016), we failed to observe any enrichment for synapse terms in GOs based solely on DEGs (Figure 2 and Supplementary table 2). In contrast, we found several GO terms associated with synaptic transmission when we imputed gDTUs (Figure 3C-D; Supplementary table 2). Altogether, these observations indicate that the use of DTU analysis, in combination with DEG, increases the sensibility of gene expression analyses to detect transcriptional alterations relevant for the understanding of AD pathology.

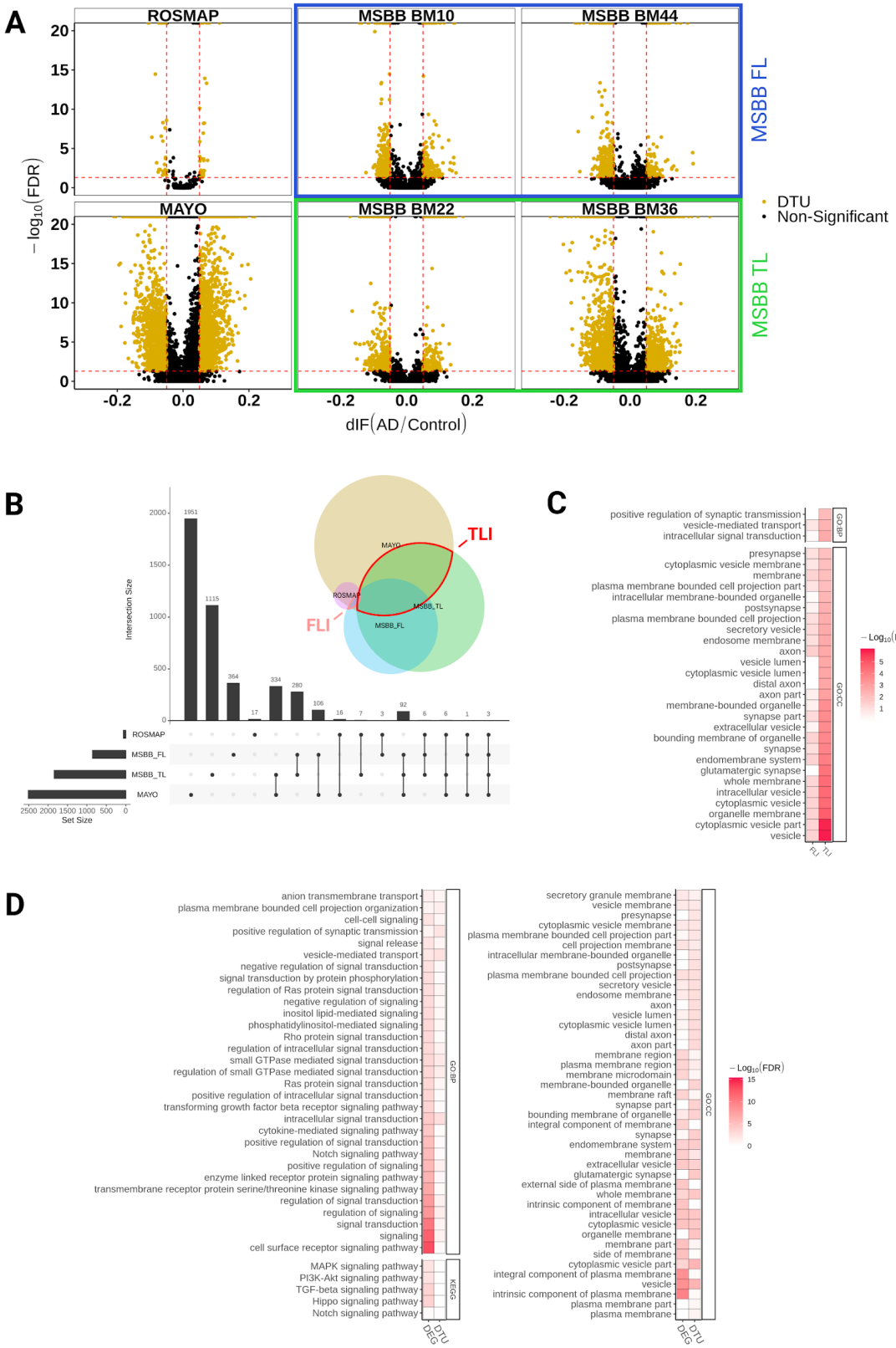
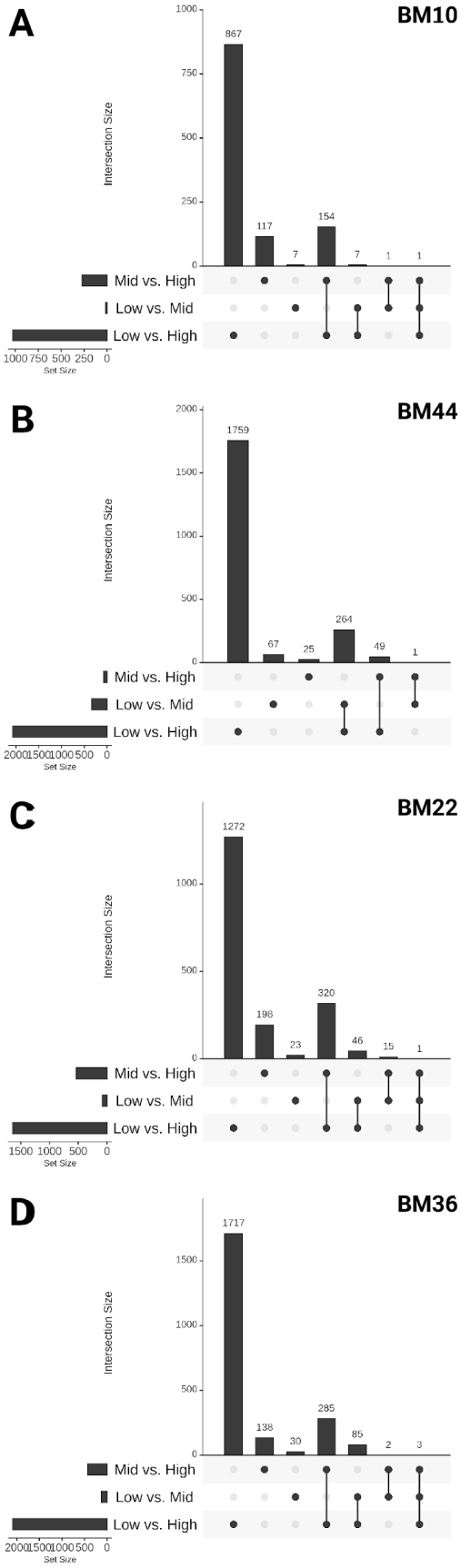

A
ROSMAP
MSBB BM10
MSBB BM44
MAYO
MSBB BM22
MSBB BM36
MSBB FL
MSBB TL
DTU
Non-Significant
−log10(FDR)
dIF(AD/Control)
B
TLI
FLI
Intersection Size
Set Size
ROSMAP
MSBB_FL
MSBB_TL
MAYO
C
GO:BP
GO:CC
−Log10(FDR)
FLI
TLI
D
GO:BP
KEGG
GO:CC
DEG
DTU

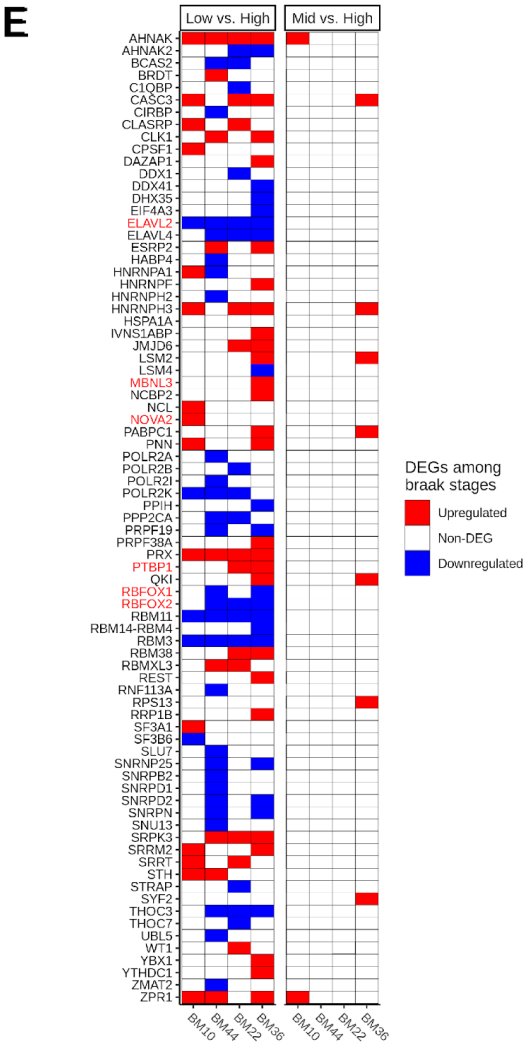
Figure 3. Differential transcript usage analysis in AD brains reveals gene expression alterations overlooked in DEG analysis. A) Volcano plots showing genes with differential transcript usage (gDTU, yellow dots; Differential isoform fraction (dIF) > 0.05 and FDR < 0.05) in the frontal lobe (ROSMAP and MSBB_FL - BM10 and BM44) and temporal lobe (MAYO and MSBB_TL - BM22 and BM36). Non-significant genes are represented by black dots. B) Venn diagram showing the number of gDTUs identified in the different datasets. The intersection between MAYO and MSBB_TL is highlighted in blue (TLI - temporal lobe intersection). C) Synapse-related terms are enriched for gDTUs in both the TLI and FLI. D) Gene ontology and KEGG terms enriched for DEGs and gDTUs in the TLI.

Differential expression of genes involved in alternative splicing correlates with isoform switch

Altered splicing has been recently associated with AD susceptibility (Raj et al., 2018). To evaluate whether gDTUs may be associated with AD hallmarks, we measured gDTUs observed at different disease stages using Braak classification in the MSBB dataset (Figure 4; Supplementary Table 3). For this purpose, we subdivided AD and control samples in three groups: low Braak (0, 1 and 2) - 196 samples (clinical diagnosis: 15 AD and 181 controls); mid Braak (3 and 4) - 133 samples (clinical diagnosis: 58 AD and 75 controls); and high Braak (5 and 6) - 308 samples (clinical diagnosis: 305 AD and 3 controls). Next, we evaluate the number of gDTUs when comparing individuals at these different stages (Figure 4). We observed that most gDTUs were detected only at high Braak stages (Figure 4A-D) both in the FL (BM10 and BM44) and TL (BM22 and BM36), suggesting that gene isoform switch could be associated with AD progression. Next, we set out to evaluate possible alterations in the expression of genes encoding for proteins associated with splicing machineries. We found that among 441 genes related to 'splicing' or 'spliceosome' terms, 79 genes were DEGs at high Braak stage compared to low or mid Braak stages (Figure 4E). In contrast, we could not detect any DEG in the comparison of mid vs low Braak stages (data not shown). Noteworthy, we found that several genes associated with the neuronal splicing regulatory network (B. Raj & Blencowe, 2015) were differentially expressed at high Braak stages (Figure 4E, highlighted in red). Among them, RBFOX1, RBFOX2 and ELAVL2 were downregulated both in the FL and TL, whereas MBNL3 and PTBP1 were upregulated in the TL (BM36) and NOVA2 was upregulated in the FL (BM10). A similar pattern was observed when we evaluated the

differential expression of the same 441 genes in all AD vs control subjects in our different datasets (Figure 4F). Although the number of DEGs was smaller, likely due to the effects of combining low, mid and high Braak stages in the AD group, we could still observe that DEGs were more prominent in the TL than in the FL. Altogether, these data suggest that changes in the expression of splice regulator proteins are more frequent in areas with high stage of pathology and could help to explain the predominant gene isoform switch observed in AD brains.


A
BM10
Intersection Size
867
117
7
154
7
1
1
Mid vs. High
Low vs. Mid
Low vs. High
Set Size
B
BM44
1759
67
25
264
49
1
C
BM22
1272
198
23
320
46
15
1
D
BM36
1717
138
30
285
85
2
3


E
Low vs. High
Mid vs. High
AHNAK
AHNAK2
BCAS2
BRDT
C1QBP
CASC3
CIRBP
CLASRP
CLK1
CPSF1
DAZAP1
DDX1
DDX41
DHX35
EIF4A3
ELAVL2
ELAVL4
ESRP2
HABP4
HNRNPA1
HNRNPF
HNRNPH2
HNRNPH3
HSPA1A
IVNS1ABP
JMJD6
LSM2
LSM4
MBNL3
NCBP2
NCL
NOVA2
PABPC1
PNN
POLR2A
POLR2B
POLR2I
POLR2K
PPIH
PPP2CA
PRPF19
PRPF38A
PRX
PTBP1
QKI
RBFOX1
RBFOX2
RBM11
RBM14-RBM4
RBM3
RBM38
RBMXL3
REST
RNF113A
RPS13
RRP1B
SF3A1
SF3B6
SLU7
SNRNP25
SNRPB2
SNRPD1
SNRPD2
SNRPN
SNU13
SRPK3
SRRM2
SRRT
STH
STRAP
SYF2
THOC3
THOC7
UBL5
WT1
YBX1
YTHDC1
ZMAT2
ZPR1
BM10
BM44
BM22
BM36
DEGs among braak stages
Upregulated
Non-DEG
Downregulated

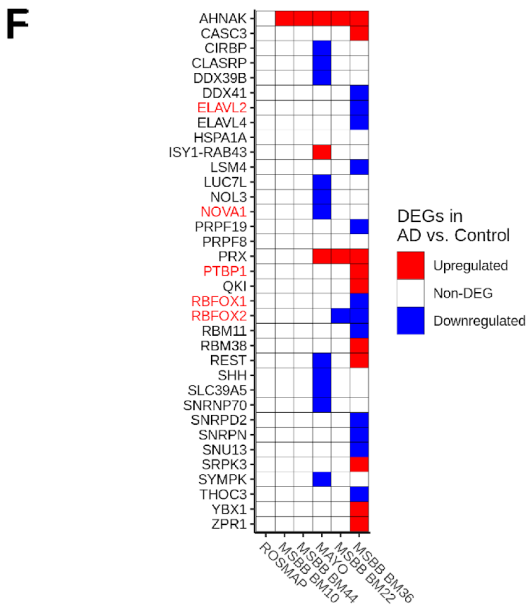

F
AHNAK
CASC3
CIRBP
CLASRP
DDX39B
DDX41
ELAVL2
ELAVL4
HSPA1A
ISY1-RAB43
LSM4
LUC7L
NOL3
NOVA1
PRPF19
PRPF8
PRX
PTBP1
QKI
RBFOX1
RBFOX2
RBM11
RBM38
REST
SHH
SLC39A5
SNRNP70
SNRPD2
SNRPN
SNU13
SRPK3
SYMPK
THOC3
YBX1
ZPR1
ROSMAP
MSBB BM10
MSBB BM44
MAYO
MSBB BM22
MSBB BM36
DEGs in AD vs. Control
Upregulated
Non-DEG
Downregulated

Figure 4. MSBB brain areas with gDTUs and DEGs related to braak stages indicate a possible alternative splicing mechanism. A-D) Total number of gDTUs identified following comparison of different Braak stages (Low vs. High, Low vs. Mid and Mid vs. High) in the 4 brain regions sampled in the MSBB study. E-F) Differential expression of genes associated with splicing/spliceosome after comparison of different Braak stages (E) or AD vs Controls in different datasets (F). Red and blue squares indicate, respectively, up and downregulated genes. Gene symbols highlighted in red indicate genes belonging to the neuronal splicing machinery.

Differential gene expression in separate cell types of the human brain

To further refine our analysis considering the cell diversity in the brain, we took an indirect approach to sort DEGs and gDTUs according to cell types. To that, we used scRNAseq data (Allen Brain) obtained from the adult human brain to identify cell types expressing the genes altered in our DEG/gDTU analysis (Figure 4A; supplementary figure 3). We found that, out of the 1135 genes in the intersections, gDTU + DEG, of temporal lobe (Figures 2 and 3), 839 were found in at least one cell-type (scTLI) using as cut-off the expression in more than 10% of cells assigned for a specific cell-type (Supp figure XX). From these, 239 were identified in unique cell-types, 396 in multiple cell-types (2-4 cell-types) and 211 in ubiquitous (more than 5 cell-types) (Figure 5A; Supplementary Figure 3; Supplementary table 4). Confirming our strategy, GO analyses using cell-type specific genes revealed that DEGs/gDTUs in the TLI of AD patients were significantly enriched for biological processes associated with inflammation in microglial cells, whereas those associated with cell adhesion were enriched in endothelial cells (Figure 5B; Supplementary table 5). Similarly, DEGs/gDTUs identified in neuronal cells were enriched for GO terms such as synaptic signaling, synaptic plasticity and synapse vesicle cycle (Figure 5C). Notably, these enrichments were more significant in GABAergic neurons, which could suggest a more pronounced pathological burden on these cells compared to glutamatergic neurons (Figure 5C).

To validate our strategy, we compared our results with two previous work using scRNAseq in AD (Mathys et al., 2019; Grubman et al., 2019). We observed a similar intersection among DEGs identified in either of the two different scRNAseq studies and our DEG+gDTU analysis, similar to what has been reported for the comparison of the two scRNAseq studies (Grubman et al., 2019). Next, we performed GO analyses for all genes altered in unique cell types

identified in the scRNAseq studies and compared to our results using DEGs and gDTUs (Supplementary Figures 1 and 2; Supplementary table 6). We also observed a high consistency for enriched terms using DEGs identified in either of the scRNAseq studies or cell-type specific DEGs/gDTUs identified in our indirect approach (Supplementary Figures 1 and 2). These data indicate that gene expression alterations identified through different strategies are related to similar biological processes.

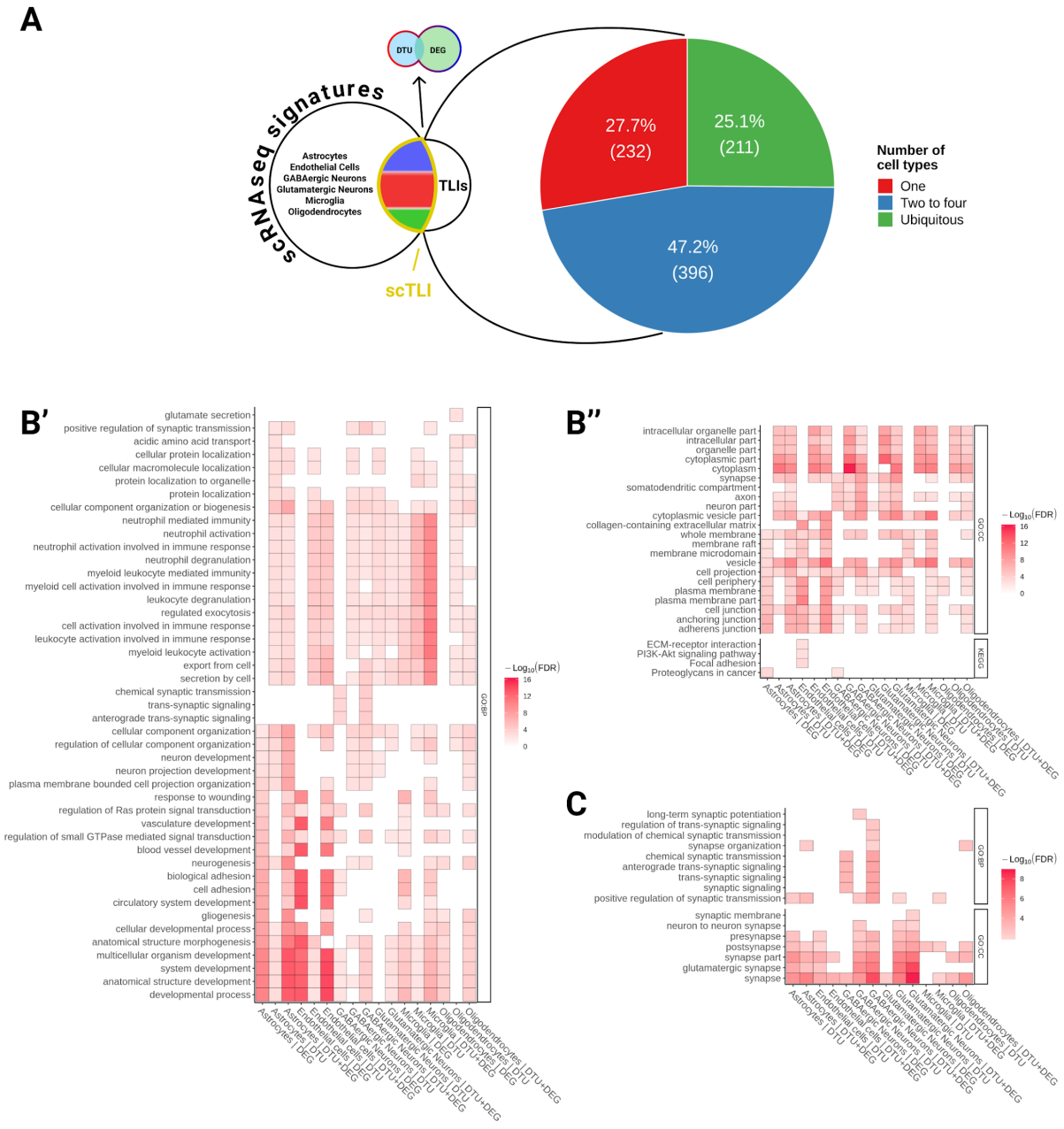


Figure 5. Cell-type expression pattern for genes altered in AD brains. A) Schematic representation showing our strategy to assign DEGs and gDTUs identified in the TLI to different cell-types of the adult human brain (see also Supplementary Figure 3). Out of 839 single-cell TLI genes (scTLI), 281 were expressed in a unique cell-type, 249 in 2 to 4 cell-types and 77 in a ubiquitous fashion. B) Gene ontology terms enriched for scTLI DEGs, gDTUs or both per cell type. C) Selected GO terms associated with synaptic transmission.

DEG/gDTU analyses identify cell-type specific alterations in AD risk factors

Genomic associated studies have revealed about 45 loci containing variants associated with the risk to develop AD (Kunkle et al., 2019; Lambert et al., 2013). However, the functional variants and their target genes remain mostly elusive (Dourlen et al., 2019). To contribute with the identification of target genes, we first evaluate the expression of 179 putative AD risk genes (Supplementary table 7) in individual cell types of the adult human brain. We found that only 116 out of the 179 AD risk genes were expressed by at least one of the major cell types identified in the brain (Figure 6A). Subsets of these genes were exclusively expressed in microglial cells (14 out of 116), neurons (12), astrocytes (2), oligodendrocytes (6) or endothelial cells (6), suggesting cell-type specific roles for these AD risk genes.

Next, we set out to evaluate the differential expression or transcript usage for these genes. We observed 54 DEGs/gDTUs in the list of 116 AD risk genes analyzed among all three bulk RNAseq datasets and also comprised in our list of genes expressed by brain cell types (Figure 6B). Taking into account the regional variations in these datasets (Figure 2-3), we focused on the AD risk genes identified in the temporal lobe for further analysis. In this region, we identified 27 and 17 DEGs/gDTUs, respectively in the MSBB_TL and Mayo datasets (Figure 6), including some well-characterized AD risk genes, such as ADAM10, BIN1, CLU and TREM2. Eight genes (15,38% of overlap) were altered in both datasets (Figure 6B, yellow box) and were selected for further analysis of isoform switch. Using ISAR to identify the isoforms altered in the AD brains compared to controls, we observed some patterns of isoform switch that could have important functional relevance (Figure 6C and D). For instance, while BIN1 transcripts ENST00000316724.9 (NP_647593.1 - isoform 1) and ENST00000409400.1 (NP_647600.1 - isoform 9) were downregulated, transcripts ENST00000393040.7 (NP_647598.1 - isoform 6) and ENST00000462958.5,

ENST0000046611.5 and ENST00000484253.1 (Retained introns) were upregulated (Figure 6C). This pattern could lead to an overall decrease of BIN1 expression, given that retained introns are non-coding sequences. Moreover, expression of BIN1 isoform 1 is neuronal specific (Zhou et al., 2014), which could suggest that the isoform switch observed for BIN1 in AD patients is likely associated with alternative splicing in that cell type.

Similarly, we also observed AD-associated isoform switches in the APP gene that could indicate changes in alternative splicing in neuronal cells. While two APP isoforms were downregulated (ENST00000348990 and ENST00000354192) in Mayo and MSBB datasets, the isoforms ENST00000346798 and ENST00000357903 were upregulated (Figure 6D). With the exception of ENST00000354192, the other transcripts are mostly expressed in neurons (Marques-Coelho and Costa, unpublished data). Noteworthy, significantly downregulated APP isoforms lack exon 6, which contains the Kunitz protease inhibitor (KPI) domain (Figure 6). KPI is one of the main serine protease inhibitors and increased KPI(+)APP mRNA and protein expression levels have been described in AD brains and are associated with increased amyloid beta deposition (Tanzi et al., 1989; Johnson et al., 1989; Kitaguchi al., 1988).

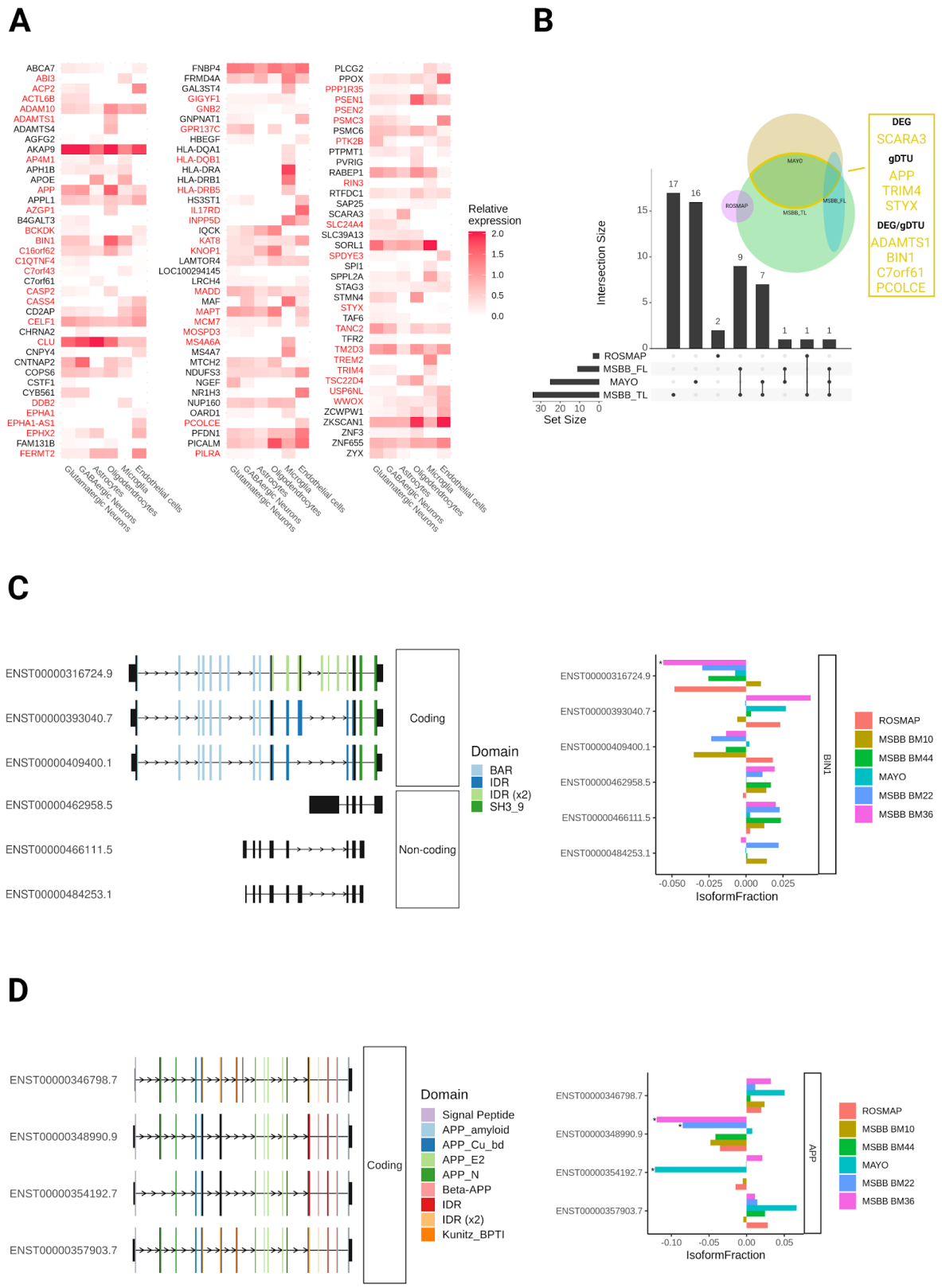

A
ABCA7
ABI3
ACP2
ACTL6B
ADAM10
ADAMTS1
ADAMTS4
AGFG2
AKAP9
AP4M1
APH1B
APOE
APP
APPL1
AZGP1
B4GALT3
BCKDK
BIN1
C16orf62
C1QTNF4
C7orf43
C7orf61
CASP2
CASS4
CD2AP
CELF1
CHRNA2
CLU
CNPY4
CNTNAP2
COPS6
CSTF1
CYB561
DDB2
EPHA1
EPHA1-AS1
EPHX2
FAM131B
FERMT2
FNBP4
FRMD4A
GAL3ST4
GIGYF1
GNB2
GNPNAT1
GPR137C
HBEGF
HLA-DQA1
HLA-DQB1
HLA-DRA
HLA-DRB1
HLA-DRB5
HS3ST1
IL17RD
INPP5D
IQCK
KAT8
KNOP1
LAMTOR4
LOC100294145
LRCH4
MADD
MAF
MAPT
MCM7
MOSPD3
MS4A6A
MS4A7
MTCH2
NDUFS3
NGEF
NR1H3
NUP160
OAR1D1
PCOLCE
PFDN1
PICALM
PILRA
PLCG2
PPOX
PPP1R35
PSEN1
PSEN2
PSMC3
PSMC6
PTK2B
PTPMT1
PVRIG
RABEP1
RIN3
RTFDC1
SAP25
SCARA3
SLC24A4
SLC39A13
SORL1
SPDYE3
SPI1
SPPL2A
STAG3
STMN4
STYX
TAF6
TANC2
TFR2
TM2D3
TREM2
TRIM4
TSC22D4
USP6NL
WWOX
ZCWPW1
ZKSCAN1
ZNF3
ZNF655
ZYX
Relative expression
2.0
1.5
1.0
0.5
0.0
Glutamatergic Neurons
GABAergic Neurons
Astrocytes
Oligodendrocytes
Microglia
Endothelial cells
B
Intersection Size
Set Size
ROSMAP
MSBB_FL
MAYO
MSBB_TL
DEG
SCARA3
gDTU
APP
TRIM4
STYX
DEG/gDTU
ADAMTS1
BIN1
C7orf61
PCOLCE
C
ENST00000316724.9
ENST00000393040.7
ENST00000409400.1
ENST00000462958.5
ENST00000466111.5
ENST00000484253.1
Coding
Non-coding
Domain
BAR
IDR
IDR (x2)
SH3_9
BIN1
IsoformFraction
ROSMAP
MSBB BM10
MSBB BM44
MAYO
MSBB BM22
MSBB BM36
D
ENST00000346798.7
ENST00000348990.9
ENST00000354192.7
ENST00000357903.7
Coding
Domain
Signal Peptide
APP_amyloid
APP_Cu_bd
APP_E2
APP_N
Beta-APP
IDR
IDR (x2)
Kunitz_BPTI
APP
IsoformFraction

Figure 6. Expression of AD risk genes is mostly altered in the TL of patients. A) Heatmap showing the expression of predicted AD risk genes in different cell types of the adult human brain. Out of 180 imputed genes (see Materials and Methods), 116 were expressed by at least one cell type. Genes highlighted in red were DEG or gDTU in at least one of the datasets used. B) Venn diagram showing the DEGs and gDTUs in the different datasets for the 116 AD risk genes expressed in specific cell types (A). The intersection between MAYO and MSBB TL is highlighted in yellow and genes identified are shown in the yellow box. C) Representation of the 4 most significant APP isoforms altered (left) and quantification of the change in isoform fraction in AD brains compared to controls (right). Main protein domains are indicated with different colors. D) Similar representation for BIN1. Differential isoform fraction (dIF) > 0.05 and FDR > 0.01.

Discussion

A comprehensive knowledge of transcriptional alterations associated with the onset and progression of human diseases is a key step towards the understanding of their cellular and molecular mechanisms (T. I. Lee & Young, 2013). In this work, we provide a novel framework to identify cell-type specific gene expression alterations in AD using patient-derived bulk RNAseq. Comparing RNA-sequencing data obtained from distinct brain regions of control and AD patients, we show that changes in gene expression are more significant in the temporal than frontal lobe. We also show that a large number of genes presents isoform switch without changes in the global expression levels. As a consequence, these genes are overlooked in classical differential expression analysis, but can be detected through differential transcript usage analysis. Gene isoform switches are mostly evident at late stages of the pathology and are associated with altered expression of genes encoding for splicing-related proteins. Using an indirect approach to assign genes to unique cell types, we are also able to map DEGs/gDTUs to unique cell populations of the adult brain, and our results are comparable to previously published scRNAseq data (Grubman et al, 2019; Mathys et al, 2019). Finally, we show that a subset of AD risk factors such as APP or BIN1, identified in genomic association studies and expressed by brain cells, are transcriptionally altered in the AD brain. Altogether, our work provides a comprehensive description of transcriptional alterations in the AD brain and suggests that alternative splicing could be an important process for pathological progression.

Despite the availability of RNAseq datasets generated from healthy subjects and AD patients (Allen et al, 2016; Wang et al, 2018; De Jager et al, 2018), a systematic evaluation of the transcriptional changes in the AD brain, as well as comparisons of transcriptional changes in distinct brain region, had not yet been performed. To the best of our knowledge, only one study aimed at comparing transcriptional changes in different AD brain regions (Haroutunian et al., 2009), but this work was based on microarray data which has a limited gene coverage. We here show, using bulk tissue RNAseq data, that alterations in gene expression are highly prominent in biological samples obtained from the temporal lobe, which harbors the first brain regions affected in the AD pathogenesis (Braak and Braak, 1991). Conversely, few transcriptional changes are present in biological samples derived from the frontal lobe, whose cells are affected only at advanced stages of the AD. These

observations are in line with recent data showing that changes in protein expression levels in AD brains are much more prominent in the temporal lobe (hippocampus, entorhinal cortex and cingulate gyrus) than in the frontal lobe (motor cortex) (Xu et al., 2019). They can also help to explain the low number of DEGs identified in scRNAseq data obtained from the frontal lobe compared to a similar study in the entorhinal cortex (Figures S1 and S2; Mathys et al., 2019; Grubman et al., 2019).

Changes in gene expression detected in RNAseq experiments are often identified by using DESeq or edgeR. Both pipelines rank differentially expressed genes considering all gene transcripts, including non-coding sequences (Costa-Silva et al., 2017). Therefore, the classical DEG analysis overlooks potential changes in the dynamic of gene expression that could, for example, lead to isoform switches with important functional consequences (Vitting-Seerup and Sandelin, 2017). Accordingly, we show that a large set of genes in the AD brain, especially in the temporal lobe, have isoform switches but are not detected in the classical DEG analysis. In other words, comparisons of the total expression of these genes in AD vs control subjects does not show statistical significance, whereas the dynamics individual transcripts/isoforms of the same genes may be profoundly altered, leading to statistically significant differences in transcript usage.

By combining both DEG and DTU analyses, we are able to identify gene sets associated with key biological processes associated with AD pathogenesis, such as synaptic communication, immune response, inflammation, endocytosis and cell-signaling (Canter et al., 2016). Notably, a significant proportion of gDTUs, but not DEGs, are associated with synapse GO terms. This could help to explain why previous analyses based solely on DEG analysis of bulk RNAseq data have failed to identify robust transcriptional alteration in synapse-related genes (Mostafavi et al., 2018; T. Raj et al., 2018), despite the central role of synapses in AD pathology.

This limitation in the detection of altered gene expression associated with synapse-related terms is partly overcome by analyzing separate cell populations in scRNAseq data (Mathys et al., 2018; Grubman et al., 2018). However, technical aspects hamper the analysis of these data to detect isoform switches (Arzalluz-Luqueángeles & Conesa, 2018). We here used an indirect approach to assign DEGs and gDTUs to unique cell types and provide evidence that such strategy is comparable to previous scRNAseq studies (Mathys et al., 2018; Grubman et

al., 2018). Interestingly, compared to the intersection between these two scRNAseq studies, our strategy shows a better similarity with Mathys et al (2018) and Grubman et al (2018) to detect alterations in glutamatergic and GABAergic neurons, respectively. Considering also the prominence of transcriptional alterations in the TL, our results could suggest a more pronounced pathological burden on GABAergic to glutamatergic neurons. Interestingly, it has been suggested that parvalbumin-expressing GABAergic neuron dysfunction critically contributes to abnormalities in oscillatory rhythms, network synchrony, and memory in hAPP mice and possibly in AD (Verret et al., 2012). Therefore, our observation at the transcriptional level could be reflecting a general hallmark of AD pathology.

Gene isoform switches could be explained by alternative splicing. In fact, we provide evidence for alterations in the splicing machinery at mid to late stages of the disease that could help to explain the high number of gDTUs observed at later stages. Again, these alterations are more abundant in the TL than in the FL, which could help to explain the low number of gDTUs in the latter brain region identified in our work and in previous study using a different strategy to detect isoform switch (Raj et al., 2018). Also in accordance with our results, it has been shown that APP-695 transcript is reduced in the hippocampus, but not cerebellum, of AD patients (Johnson et al., 1989). Thus, there is a regional susceptibility to isoform switch correlated with changes in the expression of genes encoding for proteins of the splicing machinery in the AD brain.

Particularly interesting, genes involved in the control of alternative splicing in neurons are highly affected in the temporal lobe. Decreased expression of Rbfox1 and 2, for example, could lead to an increased rate of exon inclusion (Alam et al., 2014; B. Raj & Blencowe, 2015) as observed for the APP gene in our study. Indeed, reduced expression of Rbfox1 has been associated with an increased inclusion of exon 7 in the APP gene, leading to an enhanced expression of APP isoforms 770 and 751 containing the KPI domain (Alam et al., 2014). A similar switch in the APP isoforms has also been associated with somatic gene recombination in AD (M. H. Lee et al., 2018), further indicating that increased ratios of APP isoforms containing the KPI domain could be detrimental to neurons. Considering these findings and the well-established associations between APP-KPI expression levels, amyloid plaque deposition and AD pathology progression (Tanzi et al., 1988, Johnson et al. 1990; Kitaguchi et al., 1988), it is tempting to speculate that controlling APP isoform switches by

manipulating Rbfox family proteins could be a potential therapeutic strategy to hamper disease progression.

Another interesting isoform switch detected in our study is for the AD risk gene BIN1, which is a major risk factor for AD (Lambert et al., 2013; Kunkle et al., 2019). BIN1 comprises a N-BAR domain involved in membrane curvature sensing, an SH3 domain that binds to proline-rich motifs, and a clathrin-binding domain (CLAP) specific of the neuronal isoform 1 (Zhou et al., 2014). We show that the transcript encoding for this latter isoform is significantly reduced in the temporal lobe, suggesting that expression of BIN1 isoform 1 in neurons is reduced. This observation is in line with a previous work reporting a protein decrease in BIN1 isoform 1 expression in the brain of AD case when compared with controls (Glennon et al., 2013). This would be also in agreement with the observation that an overexpression of the BIN1 isoform 1 may be protective in a model of Tauopathy (Sartori et al., 2019).

Altogether, our work indicates that isoform switch is an important mechanism for gene expression alteration in AD and genes associated with the splicing machinery in neurons could be interesting therapeutic targets to be exploited.

Acknowledgments

The results published here are in whole or in part based on data obtained from the AMP-AD Knowledge Portal (https://adknowledgeportal.synapse.org/). Study data were provided by the following sources: The Mayo Clinic Alzheimer's Disease Genetic Studies, led by Dr. Nilufer Taner and Dr. Steven G. Younkin, Mayo Clinic, Jacksonville, FL using samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimer's Disease Research Center, and the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data includes samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinson's Disease and Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimer's Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer's Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium) and the Michael J. Fox Foundation for Parkinson's Research. Study data were also provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161 (ROS), R01AG15819 (ROSMAP; genomics and RNAseq), R01AG17917 (MAP), the Illinois Department of Public Health (ROSMAP), and the Translational Genomics Research Institute (genomic). For ROSMAP, additional phenotypic data can be requested at www.radc.rush.edu. And part based on data generated from postmortem brain tissue collected through the Mount Sinai VA Medical Center Brain Bank and were provided by Dr. Eric Schadt from Mount Sinai School of Medicine.

References

Alam, S., Suzuki, H., & Tsukahara, T. (2014). Alternative splicing regulation of APP exon 7 by RBFox proteins. *Neurochemistry International*, *78*, 7–17. https://doi.org/10.1016/j.neuint.2014.08.001

Allen, M., Carrasquillo, M. M., Funk, C., Heavner, B. D., Zou, F., Younkin, C. S., Burgess, J. D., Chai, H. S., Crook, J., Eddy, J. A., Li, H., Logsdon, B., Peters, M. A., Dang, K. K., Wang, X., Serie, D., Wang, C., Nguyen, T., Lincoln, S., … Ertekin-Taner, N. (2016). Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Scientific Data*, *3*. https://doi.org/10.1038/sdata.2016.89

Anders, S., Reyes, A., & Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Research*, *22*(10), 2008–2017. https://doi.org/10.1101/gr.133744.111

Arzalluz-Luqueángeles, & Conesa, A. (2018). Single-cell RNAseq for the study of isoforms-how is that possible? *Genome Biology*, *19*(1), 1–19. https://doi.org/10.1186/s13059-018-1496-z

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. In *Acta Neuropathol* (Vol. 82).

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527. https://doi.org/10.1038/nbt.3519

Canter, R. G., Penney, J., & Tsai, L. H. (2016). The road to restoring neural circuits for the treatment of Alzheimer's disease. In *Nature* (Vol. 539, Issue 7628, pp. 187–196). Nature Publishing Group. https://doi.org/10.1038/nature20412

Costa-Silva, J., Domingues, D., & Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE*, *12*(12), 1–18. https://doi.org/10.1371/journal.pone.0190152

De Jager, P. L., Ma, Y., McCabe, C., Xu, J., Vardarajan, B. N., Felsky, D., Klein, H. U., White, C. C., Peters, M. A., Lodgson, B., Nejad, P., Tang, A., Mangravite, L. M., Yu, L., Gaiteri, C., Mostafavi, S., Schneider, J. A., & Bennett, D. A. (2018). Data descriptor: A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Scientific Data*, *5*. https://doi.org/10.1038/sdata.2018.142

Dourlen, P., Kilinc, D., Malmanche, N., Chapuis, J., & Lambert, J. C. (2019). The new genetic

landscape of Alzheimer's disease: from amyloid cascade to genetically driven synaptic failure hypothesis? In *Acta Neuropathologica* (Vol. 138, Issue 2, pp. 221–236). Springer Verlag. https://doi.org/10.1007/s00401-019-02004-0

Glennon, E. B. C., Whitehouse, I. J., Miners, J. S., Kehoe, P. G., Love, S., Kellett, K. A. B., & Hooper, N. M. (2013). BIN1 Is Decreased in Sporadic but Not Familial Alzheimer's Disease or in Aging. *PLoS ONE*, *8*(10), 1–11. https://doi.org/10.1371/journal.pone.0078806

Grubman, A., Chew, G., Ouyang, J. F., Sun, G., Choo, X. Y., McLean, C., Simmons, R. K., Buckberry, S., Vargas-Landin, D. B., Poppe, D., Pflueger, J., Lister, R., Rackham, O. J. L., Petretto, E., & Polo, J. M. (2019). A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nature Neuroscience*, *22*(12), 2087–2097. https://doi.org/10.1038/s41593-019-0539-4

Haroutunian, V., Katsel, P., & Schmeidler, J. (2009). Transcriptional vulnerability of brain regions in Alzheimer's disease and dementia. *Neurobiology of Aging*, *30*(4), 561–573. https://doi.org/10.1016/j.neurobiolaging.2007.07.021

Johnson, S. A., Rogers, J., & Finch, C. E. (1989). APP-695 transcript prevalence is selectively reduced during Alzheimer's disease in cortex and hippocampus but not in cerebellum. In *Neurobiology of Aging* (Vol. 10, Issue 6). ~ Pergamon Press plc. https://doi.org/10.1016/0197-4580(89)90017-1

Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., Boland, A., Vronskaya, M., van der Lee, S. J., Amlie-Wolf, A., Bellenguez, C., Frizatti, A., Chouraki, V., Martin, E. R., Sleegers, K., Badarinarayan, N., Jakobsdottir, J., Hamilton-Nelson, K. L., Moreno-Grau, S., … Pericak-Vance, M. A. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nature Genetics*, *51*(3), 414–430. https://doi.org/10.1038/s41588-019-0358-2

Lambert, J. C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A. L., Bis, J. C., Beecham, G. W., Grenier-Boley, B., Russo, G., Thornton-Wells, T. A., Jones, N., Smith, A. V., Chouraki, V., Thomas, C., Ikram, M. A., Zelenika, D., … Seshadri, S. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics*, *45*(12), 1452–1458. https://doi.org/10.1038/ng.2802

Lee, M. H., Siddoway, B., Kaeser, G. E., Segota, I., Rivera, R., Romanow, W. J., Liu, C. S., Park, C., Kennedy, G., Long, T., & Chun, J. (2018). Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature*, *563*(7733), 639–645. https://doi.org/10.1038/s41586-018-0718-6

Lee, T. I., & Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. In *Cell*. https://doi.org/10.1016/j.cell.2013.02.014

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/s13059-014-0550-8

Love, M. I., Soneson, C., Patro, R., Vitting-seerup, K., & Oshlack, A. (2019). *Swimming downstream: statistical analysis of differential transcript usage following Salmon quantification [version 1 ; peer review : 3 approved with reservations] Referee Status : This article is included in the RPackage gateway .* (Issue May).

Masters, C. L., Bateman, R., Blennow, K., Rowe, C. C., Sperling, R. A., & Cummings, J. L. (2015). Alzheimer's disease. In *Nature Reviews Disease Primers* (Vol. 1). Nature Publishing Group. https://doi.org/10.1038/nrdp.2015.56

Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J. Z., Menon, M., He, L., Abdurrob, F., Jiang, X., Martorell, A. J., Ransohoff, R. M., Hafler, B. P., Bennett, D. A., Kellis, M., & Tsai, L. H. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*, *570*(7761), 332–337. https://doi.org/10.1038/s41586-019-1195-2

Mostafavi, S., Gaiteri, C., Sullivan, S. E., White, C. C., Tasaki, S., Xu, J., Taga, M., Klein, H. U., Patrick, E., Komashko, V., McCabe, C., Smith, R., Bradshaw, E. M., Root, D. E., Regev, A., Yu, L., Chibnik, L. B., Schneider, J. A., Young-Pearse, T. L., … De Jager, P. L. (2018). A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nature Neuroscience*, *21*(6), 811–819. https://doi.org/10.1038/s41593-018-0154-9

Raj, B., & Blencowe, B. J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. In *Neuron* (Vol. 87, Issue 1, pp. 14–27). Cell Press. https://doi.org/10.1016/j.neuron.2015.05.004

Raj, T., Li, Y. I., Wong, G., Humphrey, J., Wang, M., Ramdhani, S., Wang, Y. C., Ng, B., Gupta, I., Haroutunian, V., Schadt, E. E., Young-Pearse, T., Mostafavi, S., Zhang, B., Sklar, P., Bennett,

D. A., & De Jager, P. L. (2018). Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nature Genetics*, *50*(11), 1584–1592. https://doi.org/10.1038/s41588-018-0238-1

Sartori, M., Mendes, T., Desai, S., Lasorsa, A., Herledan, A., Malmanche, N., Mäkinen, P., Marttinen, M., Malki, I., Chapuis, J., Flaig, A., Vreulx, A. C., Ciancia, M., Amouyel, P., Leroux, F., Déprez, B., Cantrelle, F. X., Maréchal, D., Pradier, L., … Lambert, J. C. (2019). BIN1 recovers tauopathy-induced long-term memory deficits in mice and interacts with Tau through Thr348 phosphorylation. *Acta Neuropathologica*, *138*(4), 631–652. https://doi.org/10.1007/s00401-019-02017-9

Soneson, C., Love, M. I., & Robinson, M. D. (2016). Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences [version 2; referees: 2 approved]. *F1000Research*, *4*, 1–18. https://doi.org/10.12688/F1000RESEARCH.7563.2

Strotzer, M. (2009). One century of brain mapping using Brodmann areas. *Clinical Neuroradiology*, *19*(3), 179–186. https://doi.org/10.1007/s00062-009-9002-3

Tanzi, R. E., McClatchey, A. I., Lamperti, E. D., Villa-Komaroff, L., Gusella, J. F., & Neve, R. L. (1988). Protease inhibitor domain encoded by an amyloid protein precursor mRNA associated with Alzheimer's disease. In *Nature* (Vol. 331, Issue 6156). https://doi.org/10.1038/331528a0

Thromb, A., & Biol, V. (2017). *Downloaded from*. https://doi.org/10.94

Van den Berge, K., Soneson, C., Robinson, M. D., & Clement, L. (2017). stageR: A general stage-wise method for controlling the gene-level false discovery rate in differential expression and differential transcript usage. *Genome Biology*, *18*(1), 1–14. https://doi.org/10.1186/s13059-017-1277-0

Verret, L., Mann, E. O., Hang, G. B., Barth, A. M. I., Cobos, I., Ho, K., Devidze, N., Masliah, E., Kreitzer, A. C., Mody, I., Mucke, L., & Palop, J. J. (2012). Inhibitory interneuron deficit links altered network activity and cognitive dysfunction in alzheimer model. *Cell*, *149*(3), 708–721. https://doi.org/10.1016/j.cell.2012.02.046

Vitting-Seerup, K., & Sandelin, A. (2017). The landscape of isoform switches in human cancers. *Molecular Cancer Research*, *15*(9), 1206–1220. https://doi.org/10.1158/1541-7786.MCR-16-0459

Vitting-Seerup, K., & Sandelin, A. (2019). IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics (Oxford, England)*, *35*(21), 4469–4471. https://doi.org/10.1093/bioinformatics/btz247

Wang, M., Beckmann, N. D., Roussos, P., Wang, E., Zhou, X., Wang, Q., Ming, C., Neff, R., Ma, W., Fullard, J. F., Hauberg, M. E., Bendl, J., Peters, M. A., Logsdon, B., Wang, P., Mahajan, M., Mangravite, L. M., Dammer, E. B., Duong, D. M., … Zhang, B. (2018). The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Scientific Data*, *5*. https://doi.org/10.1038/sdata.2018.185

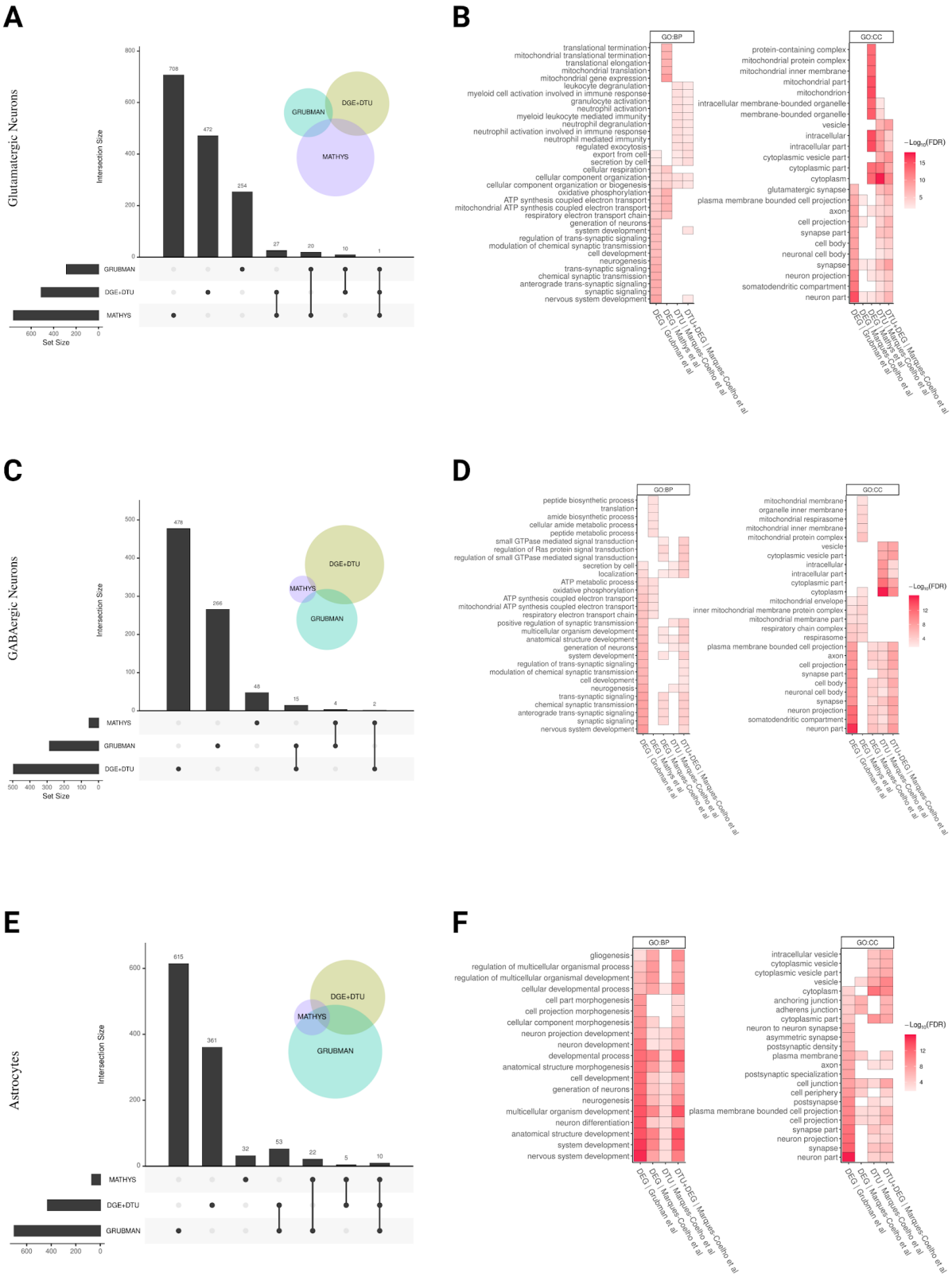
Xu, J., Patassini, S., Rustogi, N., Riba-Garcia, I., Hale, B. D., Phillips, A. M., Waldvogel, H., Haines, R., Bradbury, P., Stevens, A., Faull, R. L. M., Dowsey, A. W., Cooper, G. J. S., & Unwin, R. D. (2019). Regional protein expression in human Alzheimer's brain correlates with disease severity. *Communications Biology*, *2*(1). https://doi.org/10.1038/s42003-018-0254-9

Yi, L., Pimentel, H., Bray, N. L., & Pachter, L. (2018). Gene-level differential analysis at transcript-level resolution. *Genome Biology*, *19*(1). https://doi.org/10.1186/s13059-018-1419-z

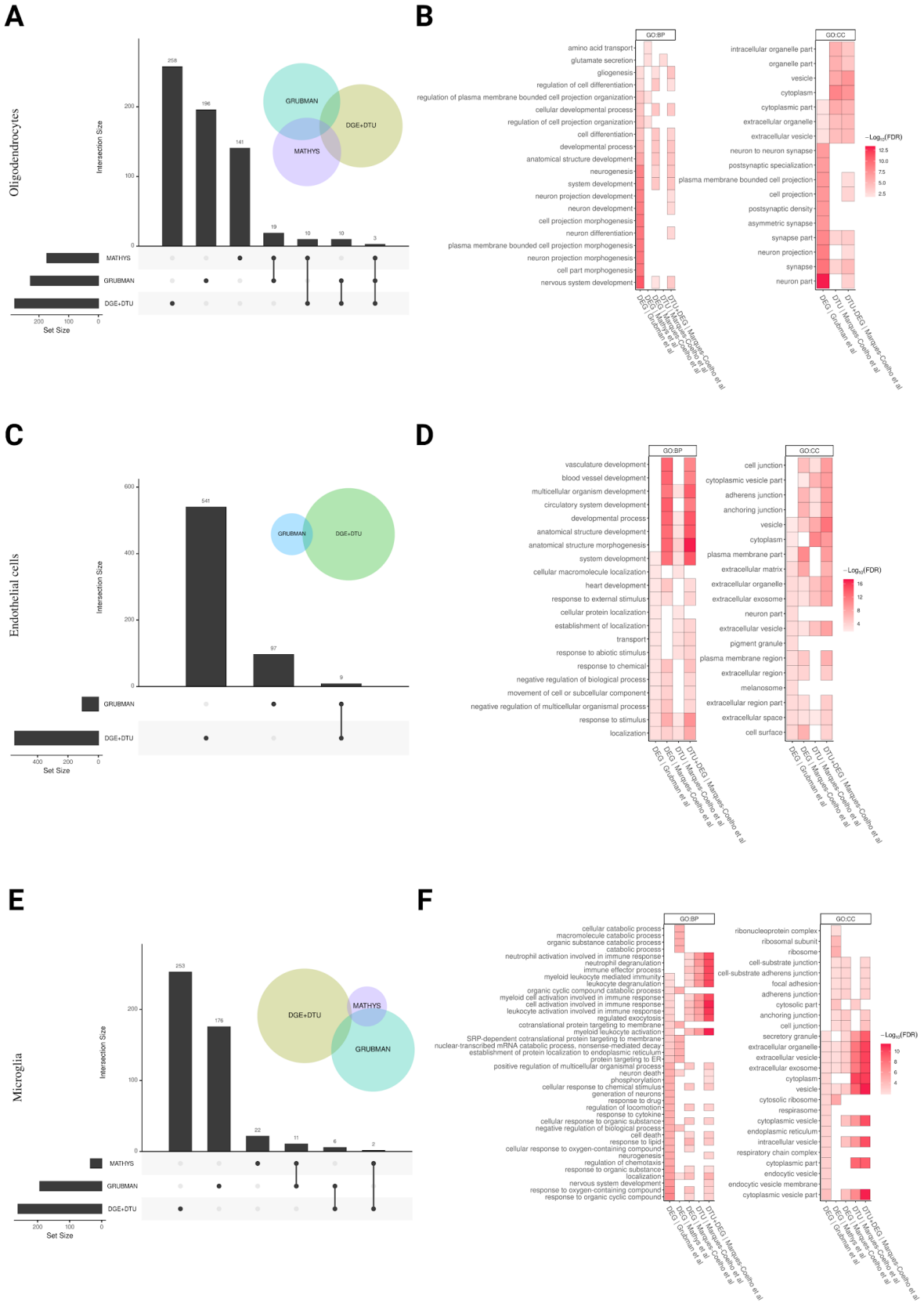
Zhou, Y., Hayashi, I., Wong, J., Tugusheva, K., Renger, J. J., & Zerbinatti, C. (2014). Intracellular clusterin interacts with brain isoforms of the bridging integrator 1 and with the microtubule-associated protein Tau in Alzheimer's Disease. *PLoS ONE*, *9*(7). https://doi.org/10.1371/journal.pone.0103187

Suplementary Figures:

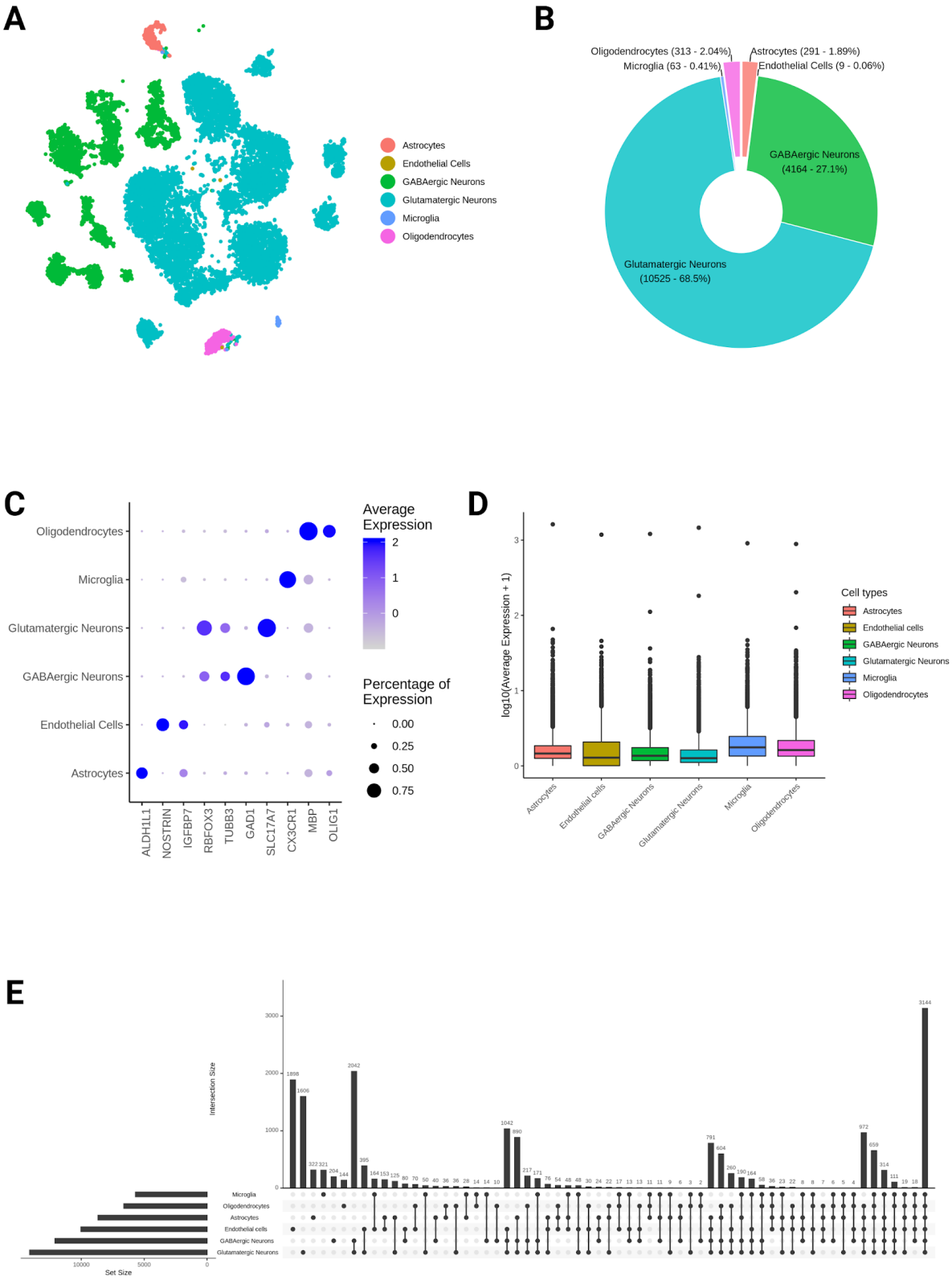
Supplementary Figure 1. scRNAseq signatures comparison with Grubman and Mathys et al works in Glutamatergic Neurons, GABAergic Neurons and Astrocytes. A,C,E) UpsetR and Venn Diagram illustrate intersection among different studies; B,D,F) Gene-set enrichment analysis (GSEA) with cell-type signatures among all studies. As threshold for GSEA, FDR < 0.01, intersection size > 3 and precision > 0.05. Analysis were made for glutamatergic neurons (A-B), gabaergic neurons (C-D) and astrocytes (E-F).


A
Glutamatergic Neurons
Intersection Size
Set Size
GRUBMAN
DGE+DTU
MATHYS
708
472
254
27
20
10
1
B
GO:BP
GO:CC
translational termination
mitochondrial translational termination
translational elongation
mitochondrial translation
mitochondrial gene expression
leukocyte degranulation
myeloid cell activation involved in immune response
granulocyte activation
neutrophil activation
myeloid leukocyte mediated immunity
neutrophil degranulation
neutrophil activation involved in immune response
neutrophil mediated immunity
regulated exocytosis
export from cell
secretion by cell
cellular respiration
cellular component organization
cellular component organization or biogenesis
oxidative phosphorylation
ATP synthesis coupled electron transport
mitochondrial ATP synthesis coupled electron transport
respiratory electron transport chain
generation of neurons
system development
regulation of trans-synaptic signaling
modulation of chemical synaptic transmission
cell development
neurogenesis
trans-synaptic signaling
chemical synaptic transmission
anterograde trans-synaptic signaling
synaptic signaling
nervous system development
protein-containing complex
mitochondrial protein complex
mitochondrial inner membrane
mitochondrial part
mitochondrion
intracellular membrane-bounded organelle
membrane-bounded organelle
vesicle
intracellular
intracellular part
cytoplasmic vesicle part
cytoplasmic part
cytoplasm
glutamatergic synapse
plasma membrane bounded cell projection
axon
cell projection
synapse part
cell body
neuronal cell body
synapse
neuron projection
somatodendritic compartment
neuron part
DEG | Grubman et al
DEG | Mathys et al
DTU | Marques-Coelho et al
DTU-DEG | Marques-Coelho et al
C
GABAergic Neurons
Intersection Size
Set Size
MATHYS
GRUBMAN
DGE+DTU
478
266
48
15
4
2
D
GO:BP
GO:CC
peptide biosynthetic process
translation
amide biosynthetic process
cellular amide metabolic process
peptide metabolic process
small GTPase mediated signal transduction
regulation of Ras protein signal transduction
regulation of small GTPase mediated signal transduction
secretion by cell
localization
ATP metabolic process
oxidative phosphorylation
ATP synthesis coupled electron transport
mitochondrial ATP synthesis coupled electron transport
respiratory electron transport chain
positive regulation of synaptic transmission
multicellular organism development
anatomical structure development
generation of neurons
system development
regulation of trans-synaptic signaling
modulation of chemical synaptic transmission
cell development
neurogenesis
trans-synaptic signaling
chemical synaptic transmission
anterograde trans-synaptic signaling
synaptic signaling
nervous system development
mitochondrial membrane
organelle inner membrane
mitochondrial respirasome
mitochondrial inner membrane
mitochondrial protein complex
vesicle
cytoplasmic vesicle part
intracellular
intracellular part
cytoplasmic part
cytoplasm
mitochondrial envelope
inner mitochondrial membrane protein complex
mitochondrial membrane part
respiratory chain complex
respirasome
plasma membrane bounded cell projection
axon
cell projection
synapse part
cell body
neuronal cell body
synapse
neuron projection
somatodendritic compartment
neuron part
E
Astrocytes
Intersection Size
Set Size
MATHYS
DGE+DTU
GRUBMAN
615
361
32
53
22
5
10
F
GO:BP
GO:CC
gliogenesis
regulation of multicellular organismal process
regulation of multicellular organismal development
cellular developmental process
cell part morphogenesis
cell projection morphogenesis
cellular component morphogenesis
neuron projection development
neuron development
developmental process
anatomical structure morphogenesis
cell development
generation of neurons
neurogenesis
multicellular organism development
neuron differentiation
anatomical structure development
system development
nervous system development
intracellular vesicle
cytoplasmic vesicle
cytoplasmic vesicle part
vesicle
cytoplasm
anchoring junction
adherens junction
cytoplasmic part
neuron to neuron synapse
asymmetric synapse
postsynaptic density
plasma membrane
axon
postsynaptic specialization
cell junction
cell periphery
postsynapse
plasma membrane bounded cell projection
cell projection
synapse part
neuron projection
synapse
neuron part

Supplementary Figure 2. scRNAseq signatures comparison with Grubman and Mathys et al works in Oligodendrocytes, Endothelial cells and Microglia. A,C,E) UpsetR and Venn Diagram illustrate intersection among different studies; B,D,F) Gene-set enrichment analysis (GSEA) with cell-type signatures among all studies. As a threshold for GSEA, FDR < 0.01, intersection size > 3 and precision > 0.05. Analyses were made for oligodendrocytes (A-B), endothelial cells (C-D) and microglia (E-F).


A
Oligodendrocytes
Intersection Size
Set Size
MATHYS
GRUBMAN
DGE+DTU
258
196
141
19
10
10
3
B
GO:BP
GO:CC
amino acid transport
glutamate secretion
gliogenesis
regulation of cell differentiation
regulation of plasma membrane bounded cell projection organization
cellular developmental process
regulation of cell projection organization
cell differentiation
developmental process
anatomical structure development
neurogenesis
system development
neuron projection development
neuron development
cell projection morphogenesis
neuron differentiation
plasma membrane bounded cell projection morphogenesis
neuron projection morphogenesis
cell part morphogenesis
nervous system development
intracellular organelle part
organelle part
vesicle
cytoplasm
cytoplasmic part
extracellular organelle
extracellular vesicle
neuron to neuron synapse
postsynaptic specialization
plasma membrane bounded cell projection
cell projection
postsynaptic density
asymmetric synapse
synapse part
neuron projection
synapse
neuron part
−Log10(FDR)
DEG | Grubman et al
DTU | Marques-Coelho et al
DTU-DEG | Marques-Coelho et al
C
Endothelial cells
Intersection Size
Set Size
GRUBMAN
DGE+DTU
541
97
9
D
GO:BP
GO:CC
vasculature development
blood vessel development
multicellular organism development
circulatory system development
developmental process
anatomical structure development
anatomical structure morphogenesis
system development
cellular macromolecule localization
heart development
response to external stimulus
cellular protein localization
establishment of localization
transport
response to abiotic stimulus
response to chemical
negative regulation of biological process
movement of cell or subcellular component
negative regulation of multicellular organismal process
response to stimulus
localization
cell junction
cytoplasmic vesicle part
adherens junction
anchoring junction
vesicle
cytoplasm
plasma membrane part
extracellular matrix
extracellular organelle
extracellular exosome
neuron part
extracellular vesicle
pigment granule
plasma membrane region
extracellular region
melanosome
extracellular region part
extracellular space
cell surface
−Log10(FDR)
DEG | Grubman et al
DTU | Marques-Coelho et al
DTU-DEG | Marques-Coelho et al
E
Microglia
Intersection Size
Set Size
MATHYS
GRUBMAN
DGE+DTU
253
176
22
11
6
2
F
GO:BP
GO:CC
cellular catabolic process
macromolecule catabolic process
organic substance catabolic process
catabolic process
neutrophil activation involved in immune response
neutrophil degranulation
immune effector process
myeloid leukocyte mediated immunity
leukocyte degranulation
organic cyclic compound catabolic process
myeloid cell activation involved in immune response
cell activation involved in immune response
leukocyte activation involved in immune response
regulated exocytosis
cotranslational protein targeting to membrane
myeloid leukocyte activation
SRP-dependent cotranslational protein targeting to membrane
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
establishment of protein localization to endoplasmic reticulum
protein targeting to ER
positive regulation of multicellular organismal process
neuron death
phosphorylation
cellular response to chemical stimulus
generation of neurons
response to drug
regulation of locomotion
response to cytokine
cellular response to organic substance
negative regulation of biological process
cell death
response to lipid
cellular response to oxygen-containing compound
neurogenesis
regulation of chemotaxis
response to organic substance
localization
nervous system development
response to oxygen-containing compound
response to organic cyclic compound
ribonucleoprotein complex
ribosomal subunit
ribosome
cell-substrate junction
cell-substrate adherens junction
focal adhesion
adherens junction
cytosolic part
anchoring junction
cell junction
secretory granule
extracellular organelle
extracellular vesicle
extracellular exosome
cytoplasm
vesicle
cytosolic ribosome
respirasome
cytoplasmic vesicle
endoplasmic reticulum
intracellular vesicle
respiratory chain complex
cytoplasmic part
endocytic vesicle
endocytic vesicle membrane
cytoplasmic vesicle part
−Log10(FDR)
DEG | Grubman et al
DTU | Marques-Coelho et al
DTU-DEG | Marques-Coelho et al

Supplementary Figure 3. Allen Brain Atlas scRNAseq MTG dataset information. A) tSNE representation showing distribution of single cells in two dimensions; B) Donut plot explaining percentage of cell types among all groups with absolute number of cells; C) Dotplot showing cell type known markers elucidating cell clusters findings; D) Boxplot demonstrating distribution of average expression among all cell types; E) UpsetR exhibiting number of genes expressed in each cell type combination.


A
Astrocytes
Endothelial Cells
GABAergic Neurons
Glutamatergic Neurons
Microglia
Oligodendrocytes
B
Oligodendrocytes (313 - 2.04%)
Astrocytes (291 - 1.89%)
Microglia (63 - 0.41%)
Endothelial Cells (9 - 0.06%)
GABAergic Neurons (4164 - 27.1%)
Glutamatergic Neurons (10525 - 68.5%)
C
Oligodendrocytes
Microglia
Glutamatergic Neurons
GABAergic Neurons
Endothelial Cells
Astrocytes
ALDH1L1
NOSTRIN
IGFBP7
RBFOX3
TUBB3
GAD1
SLC17A7
CX3CR1
MBP
OLIG1
Average Expression
2
1
0
Percentage of Expression
0.00
0.25
0.50
0.75
D
log10(Average Expression + 1)
Cell types
Astrocytes
Endothelial cells
GABAergic Neurons
Glutamatergic Neurons
Microglia
Oligodendrocytes
E
Intersection Size
Set Size
Microglia
Oligodendrocytes
Astrocytes
Endothelial cells
GABAergic Neurons
Glutamatergic Neurons