

# Single nuclear transcriptional signatures of dysfunctional brain vascular homeostasis in Alzheimer's disease

Stergios Tsartsalis<sup>1,2#</sup>, Nurun Fancy<sup>1,3#</sup>, Amy M. Smith<sup>1,3,4</sup>, Combiz Khoozie<sup>1,3</sup>, Xin Yang<sup>1,5</sup>, Karen Davey<sup>1,3</sup>, Nanet Willumsen<sup>1,3</sup>, Aisling McGarry<sup>1,3</sup>, Robert C. J. Muirhead<sup>1,3</sup>, Stephanie Debette<sup>6</sup>, David R. Owen<sup>1</sup> and Paul M. Matthews<sup>1,3,5\*</sup>

Department of Brain Sciences, Imperial College London, United Kingdom<sup>1</sup>; Department of Psychiatry, University of Geneva, Switzerland<sup>2</sup>; UK Dementia Research Institute, Imperial College London<sup>3</sup>; Centre for Brain Research and Department of Pharmacology, University of Auckland, New Zealand<sup>4</sup>; St Edmund Hall, University of Oxford<sup>5</sup>; <sup>6</sup>University of Bordeaux, Inserm, Bordeaux Population Health Research Center, team VINTAGE, UMR 1219, 33000 Bordeaux, France

#These authors share first authorship.

## Address for correspondence:

Prof. Paul M. Matthews  
E515, Department of Brain Sciences  
Imperial College London  
Hammersmith Hospital  
DuCane Road, London WC12 0NN  
Tel: 0044 207 594 2612

[p.matthews@imperial.ac.uk](mailto:p.matthews@imperial.ac.uk)

## Abstract

Brain perfusion and normal blood brain barrier integrity are reduced early in Alzheimer's disease (AD). We performed single nucleus RNA sequencing of vascular cells isolated from AD and control brains to characterise pathological transcriptional signatures. We found that endothelial cells (EC) are enriched for expression of genes associated with susceptibility to AD. EC transcriptional signatures identified mechanisms for impaired  $\beta$ -amyloid clearance. Evidence for immune activation was found with upregulation of interferon signalling genes in EC and in pericytes (PC). Transcriptional signatures suggested dysregulation of vascular homeostasis and angiogenesis with upregulation of pro-angiogenic signals (*HIF1A*) and metabolism in EC, but downregulation of homeostatic growth factor pathways (VEGF, EGF, insulin) in EC and PC and of extracellular matrix genes in fibroblasts (FB). Our genomic dissection of vascular cell risk gene enrichment suggests a potentially causal role for EC and defines transcriptional signatures associated with microvascular dysfunction in AD.

## Introduction

Alzheimer's disease (AD) is the most common form of dementia<sup>1</sup>, characterized by extracellular deposits of toxic forms of  $\beta$ -amyloid (A $\beta$ ) protein, intracellular neurofibrillary tangles (NFTs) and neurodegeneration. Large-scale genomic association studies have suggested specific molecular processes responsible for susceptibility to disease<sup>2-4</sup>. The non-neuronal cells in which these genes are predominantly expressed are candidates for early "causal" roles in the initiation of the pathological cascades of AD<sup>5</sup>.

Brain microvasculature appears to play a major role in AD pathophysiology<sup>6-8</sup>. Cells constituting the blood brain barrier (BBB) contribute to the clearance of A $\beta$  and other toxic species from the central nervous system (CNS) and allow the selective exclusion of potentially inflammatory or toxic blood proteins from the brain and control of immune cell trafficking<sup>9</sup>. Vascular pericytes are responsible for regulating brain perfusion and contribute to the regulation of endothelial permeability and immune activation<sup>7,10</sup>. Multiple *in vivo* imaging and *post mortem* neuropathological studies, as well as studies of preclinical models, provide evidence for impaired regulation of cerebral blood flow and maintenance of the integrity of the blood brain barrier (BBB) in early AD<sup>11-14</sup>. Recent work has begun to elucidate the transcriptional mechanisms<sup>15-17</sup>.

We have performed an integrated analysis of our own single-nuclei RNA sequencing (snRNAseq) data with that from a previously published dataset<sup>18</sup> to quantitatively define the enrichment of brain microvascular cells for the expression of AD risk genes as a test of their potential causal contributions to disease genesis<sup>5</sup>. We then explored the functional roles of AD risk genes by assessing functional enrichment of genes co-expressed with them in vascular cells. Differential expression and gene co-expression analyses allowed characterisation of specific genes and pathways altered in AD. A cell-cell communication analysis further defined signalling mechanisms supporting vascular homeostasis and angiogenesis that are impaired with AD. Together, our results provide a transcriptomic

mechanistic description for major features of the vascular pathophysiology observed *in vivo* with AD.

## Results

### Endothelial cells are enriched in genes associated with genetic risk for AD

Our analyses were based on data from 57 different brain samples from donors with AD (n=31) or non-diseased controls (NDC, n=26). Fluorescence-activated sorting (FACS) of nuclei isolated before snRNAseq removed neuronal and oligodendrocyte nuclei to achieve a better representation of the less abundant brain cell types of interest. Data was integrated using LIGER<sup>19</sup> and clustered with UMAP<sup>20</sup> (Figure 1A). AD and NDC donor nuclei and nuclei from different datasets, brain regions and sexes were well-mixed after integration (Figure S1). Nuclei numbers did not differ significantly between AD and the NDC samples. Feature plots of canonical cell markers identified major brain cell types in the integrated dataset (Figure S2). Endothelial cells (EC) expressed marker genes *FLT1*, *VWF*, *NOSTRIN* (Figure S3A), *CLDN5* and *IFI27*<sup>17,21</sup> (Figure 1C). Specific expression of *COL1A1*, *COL12A1*, *COL6A1* (Figure S3B) and *COL5A1* was used to identify fibroblasts (FB) (Figure 1C). A separate, heterogeneous cluster of vascular mural cell nuclei expressed *PDGFRB*, *RGS5* and *GRM8* (characteristic of PC<sup>17</sup>) (Figure S3C) and *ACTA2* and *MYH11* (highly expressed in smooth muscle cells (SMC)<sup>17</sup>) (Figure 1C). To distinguish PC from SMC nuclei, we re-clustered the EC, FB and vascular mural cell (PC and SMC) nuclei from the total dataset (Figure 1B) to separate those expressing high levels of *ACTA2* and *MYH11* with very low levels of *RGS5* and *GRM8* (corresponding to SMC) from those expressing high levels of *RGS5* and *GRM8* with very low levels of *ACTA2* and *MYH11* (corresponding to PC) (Figure 1C and S4). We confirmed our cluster annotations by demonstrating significant mutual overrepresentations of our cluster markers and those reported previously in human<sup>17</sup> (Figure S5A) and mouse<sup>17,21</sup> (Figure S5B) single nuclei or single cell RNA sequencing studies. To further characterize the identity of the FB population, we tested the overrepresentation of previously described meningeal and

perivascular FB markers<sup>15</sup> and found that our FB markers were more significantly enriched in perivascular fibroblast markers (Fisher's exact test (FET) for overrepresentation: perivascular FB markers,  $p = 2.82 \times 10^{-77}$ ; meningeal FB markers,  $p = 6.36 \times 10^{-19}$ ).

Well-annotated genes associated with genetic risk of AD<sup>2-4</sup> were expressed in nuclei from all four vascular cell types (Figure 1F): 52/61 AD risk genes tested were found in at least one of the vascular cells studied, although less than half of these genes were expressed in 5% or more of nuclei (EC, 21/61; FB, 21/61; SMC, 17/61; PC, 19/61). 14/61 of these genes were expressed in at least 5% of the nuclei across all *four* cell types (*ADAM10*, *APOE*, *CD2AP*, *CELF1*, *CLU*, *CNTNAP2*, *FERMT2*, *IQCK*, *MEF2C*, *PICALM*, *SORL1*, *SPPL2A*, *USP6NL*, *WWOX*).

We employed MAGMA.Celltyping to test for the significance of the enrichment of vascular nuclei across the larger set of genomic loci associated with AD<sup>5</sup>. First, we analysed a dataset that included all the canonical cell types of the brain (Figure S6-S8). This showed that the AD risk gene expression enrichment is greatest in microglia, as has been reported previously<sup>5</sup> (Figure 1D). Vascular cells also were relatively enriched, albeit less than microglia. To partition enrichment amongst the individual vascular cell types, the analysis was repeated with vascular cell data alone. Only EC were significantly enriched for expression of AD risk genes (Figure 1E). Brain small vessel disease and MRI brain white matter hyperintensities (WMH) are associated with risk of AD<sup>22,23</sup>. To test whether risks for small vessel disease were responsible for the EC enrichment for AD genetic risk, we re-assessed enrichment for the latter after statistically controlling for WMH risk gene expression<sup>24</sup>. The results remained virtually unchanged (Figure 1E): the genetic risk for AD associated to the EC transcriptome thus appears to be independent of that for WMH. However, when the analysis was repeated after controlling for the microglial enrichment, the vascular enrichment largely disappeared, suggesting that similar AD risk gene sets are enriched in microglia and vascular (Figure 1D).

**Differential gene expression (DGE) identifies transcriptional signatures of dysfunctional angiogenesis in AD**

We employed a mixed-effects model in MAST<sup>25</sup> to discover genes differentially expressed in AD relative to NDC for each of the cell types. Greater numbers of genes were downregulated (90 genes, EC; 47 genes, FB; 47 genes, PC; FDR 0.1), than upregulated (73 genes, EC; 25 genes, FB; 25 genes, PC) in EC, FB and PC (Figure 2A-C). We did not find significant differentially expressed genes (DEG) in the SMC.

#### *Pathological angiogenic transcriptional signatures in AD*

Proangiogenic *HIF1A* was overexpressed in EC in AD (Figure 2A). However, the expression of multiple functionally related genes (e.g., *SPRED2*, *SHC2*, *KSR1*, *RASGRF2*, *DAB2IP*, *RASAL2*, *DUSP16*, *VCL* and *EGFR*) involved in VEGFR2, EGFR and insulin receptor-mediated pathways were downregulated. Pathological angiogenic gene expression signatures also were found in FB with downregulation of the expression of VEGF, FGF, EGF and IGF pathway genes (including *SPRED2*, *DAB2IP* and *SPTBN1*), *DTX2*, a regulator of Notch signalling<sup>26</sup> and the sialyltransferase gene, *ST3GAL6* (Figure 2B, E). Gene expression in PC highlighted a strikingly mixed angiogenic signature with upregulation of the angiogenic Wnt/β-catenin signalling pathway and downregulation of EGF/EGFR signalling pathway genes including *RPS6KA2*, *ASAP1*, *MEF2D* and *EGFR* (Figure 2C, F).

Clues to additional mechanisms responsible for loss of BBB integrity in AD were found with the differentially expressed gene signatures. For example, genes contributing to adherens junction assembly (*VCL*, *TBCD* and *PIP5K1C*) were variably differentially expressed in EC (*VCL* and *TBCD* were downregulated, whereas *PIP5K1C* was upregulated). As noted above, Wnt/β-catenin pathway genes, including *TCF4* and *APC*, expression of which also support blood brain barrier integrity<sup>27</sup>, are significantly upregulated with AD in PC. Finally, *LAMC1*, encoding for the gamma laminin subunit, was significantly downregulated in FB.

#### *Differential expression of risk genes with functional roles in amyloid processing and immune response in AD*

Risk genes with functional roles related to amyloid processing were differentially expressed in AD. *ADAM10*, which encodes the constitutive  $\alpha$ -secretase that governs non-amyloidogenic pathway  $\beta$ -amyloid precursor protein processing, was significantly upregulated in PC. *PICALM*, encoding a clathrin assembly protein modulating clearance of A $\beta$ , was downregulated in EC. Risk genes related to immune responses also were differentially expressed with AD. Increased expression of the inhibitory complement receptor *CD46* gene and decreased expression of *IRAK3*, which encodes an IL-1 receptor associated kinase were found in PC. In EC, we also found increased expression of *IFITM3*, which regulates interferon pathway inflammatory responses<sup>28</sup> and can potentiate gamma secretase activity<sup>29</sup>.

### **Co-expression modules for angiogenesis, lysosomal processing and interferon activation are differentially regulated with AD**

To identify gene co-expression modules differentially expressed with AD, we first performed gene co-expression network analyses separately for EC, FB and PC pooled across AD and NDC (MEGENA<sup>30</sup>). We then determined which gene co-expression modules were differentially associated with AD (limma<sup>31</sup>). SMC were not included in this analysis because of the relatively low number of nuclei available for analysis and consequent sparse co-expression representation. Our results defined cell-specific gene regulation pathways associated with AD.

### *Reduced expression of co-expression modules enriched for angiogenesis and vascular homeostasis*

Although fold-changes varied, expression of modules enriched for multiple angiogenic or vascular homeostasis pathways were generally decreased with AD, most prominently for EC (Figure 3), e.g., Module 119, which is enriched in the EGF/EGFR signalling pathway (e.g., including *EGFR*, *IQSEC1*, *INPP5D*, *NEDD4* and *IQGAP1*) and several hub genes (e.g., the G-protein activator and adhesion G protein coupled receptor genes, *DOCK9* and *ADGRL4*, respectively, and the hypoxia inducible transcription factor, *EPAS1*) with roles in angiogenesis<sup>32,33</sup>. Expression also was reduced for module 47, which is enriched in insulin-

like growth factor 1 receptor (IGF1R) (*IGF1R*, *PSMD5*, *PSMD1*, *TSC1*, *RASAL2*) and Ras signalling cascade (*PSMD5*, *PSMD1*, *RASAL2*) genes. Expression of several genes in module 47 with recognised functional roles in angiogenesis (e.g., *RASAL2* and *PALD1*<sup>34</sup>) were significantly independently downregulated in AD.

While the fold-changes were relatively low, FB modules functionally related to vascular homeostasis also were significantly downregulated, e.g., Module 2, which was significantly enriched in NOTCH signalling genes (Figure 3E), including *NOTCH1* and *NOTCH2*, as well as the regulators of NOTCH expression<sup>26</sup>, *ARRB1* and *DTX2*, both of which also were independently significantly differentially reduced in expression with AD. Module 2 also is significantly enriched in genes individually downregulated in AD (FET adjusted p=1.28x10<sup>-8</sup>). Some of them (*SPRED2*, *DAB2IP*, *ARRB1*, *SPTBN1*) encode for proteins that are downstream components of several growth factor signalling pathways (e.g., FGFR1-4, VEGFR2, EGFR) or genes for components (*COL5A1* and *COL1A2*) of the extracellular matrix (ECM)<sup>35</sup>. Angiogenesis pathway enriched modules were downregulated in PC, as well, e.g., module 19 (Figure 3F), in which several genes showing individually significantly reduced expression in AD (*EGFR*, *ZBTB16*, *IRAK3*, *TMTC1*, *MAOA*) and genes involved in EGFR signalling (*EGFR*, *MAPK1*, *FOXO3*) are found or module 15, which is enriched in PI3K-Akt signalling pathway genes (*ANGPT2*, *COL6A2*, *DDIT4*, *COL6A1*, *BCL2*, *PDGFA*, *PPP2R3B*, *PPP2R3A*) genes<sup>36</sup>. A recent report has shown the *Angpt2* knock out potentiates BBB leakage in a preclinical A $\beta$  mouse model<sup>37</sup>.

However, despite the decreased expression of modules enriched for many angiogenic pathways central to angiogenesis with AD, we also found modules enriched in metabolic pathways supporting angiogenesis in EC and PC and extracellular matrix genes in FB, the expression of which was increased in AD. Module 661, the top upregulated module in AD, includes *TPI1* that encodes for the triosephosphate isomerase, an enzyme implicated in glycolysis and gluconeogenesis<sup>38</sup>. Module 41, the expression of which is increased with AD, includes genes encoding for the acyl-CoA dehydrogenase that executes the first step of the  $\beta$ -oxidation of fatty acids (*ACAD8*, *ACADS*) and genes of the butanoate metabolism pathway

(*BDH2*, *ACSM3*, and *ACADS*)<sup>39</sup>, both of which are highly upregulated in AD. Module 12, also increased in expression with AD, is amongst several in EC that were enriched for pathways coding for proteins of the mitochondrial respiratory chain complexes, including *ATP5PF*, *ATP5PD*, *ATP6 NDUFB10* and *NDUFS3*. The most highly upregulated module in PC, module 512, was enriched in oxidative phosphorylation genes (*NDUFA4L2*, *ATP5MC2*, *ATP5F1D*) and for lipid metabolism pathway genes (*PDHA1*, *PRKAB1*)<sup>40,41</sup>. Fibroblasts play a central role in the development of the basement membrane. Extracellular matrix collagen genes (*COL3A1*, *COL5A2*, *COL5A3*, *COL11A1*, *COL21A1*), the fibronectin 1 gene (*FN1*), proteoglycan and glycosaminoglycan metabolism-related genes (*CHSY3*, *GXYLT2*, *GPC6*, *CHST15*, *TIAM1*, *KDR*, *PLCE1*, *COL21A1*, *FLNB*, *ANK3*, *TP53*) were enriched in modules upregulated with AD in FB, e.g., in branches of modules 128, 135, 164, 408, 488 and 775. Modules 408 and 391, also upregulated with AD in FB, were enriched in solute carrier genes involved in nutrient and metabolite transfer across the blood brain barrier<sup>42</sup>.

#### *Increased expression of APOE and lysosomal pathway enriched modules in FB and PC with AD*

Module 396, the most highly upregulated module in FB with AD, was enriched in the AD risk gene *APOE* and other cholesterol metabolism-related genes (e.g., *AGT*), as well as genes with individually increased expression in AD that are related to pathways for lytic vacuole functions (including *CACNG7*, *VPS28*, *CTSO*, *RRAGC*, *NPC2*, *GAA*, *HPS4*, *CTNS*, *RAB9A*, *VAMP4*, *VPS16* and *RAB7A*). Expression of module 512, which is enriched for lipid metabolism (with *PDHA1* and *PRKAB1*) and lysosome (with *PSAP*, *CLTB* and *CTSF*) pathways, increased in PC, whereas other lipid processing and metabolic gene pathways were downregulated with AD, notably in module 500 (enriched for the sphingolipid signalling pathway genes *AKT3*, *MAPK14*, *PLCB1* and *NSMAF*). FB modules 2 (enriched for “regulation of lipid metabolism by PPAR- $\alpha$ ” and “metabolism of lipids and lipoproteins” pathways) and modules 468 (enriched for “response to lipid”) and 284 (enriched for the “ABC transporters in lipid homeostasis” pathway) also were downregulated with AD.

### *Module enrichments suggest pathological amyloid processing and immune responses in AD*

There was evidence for upregulation of A $\beta$  production and immune responses with AD. Module 41, which was highly upregulated in EC, contains *PSEN2* and *APH1A*, which encode for components of the gamma secretase<sup>43</sup>. We also found increased expression of interferon-related co-expression modules in EC (module 661, the most highly upregulated modules includes *C2* and *TRIM5* which is a interferon type I-stimulated gene<sup>44</sup>) and in PC (module 161 includes *RNASEL*, *EIF4A3*, *AAAS* and *IFIT1*) with AD. The most highly upregulated module in PC (Module 512) is enriched for pathways for ferroptosis (*GPX4*, *FTH1*), which promotes release of oxidised lipid species that generate pro-inflammatory damage-associated molecular patterns (DAMPs) able to trigger the innate immune system<sup>45</sup>. However, modules enriched for other immune response pathways were *downregulated* with AD, e.g., EC module 455, which is enriched for IL-6 signalling (*NLK*, *IL6ST*, *JAK1*), and PC module 19, enriched for cytokine signalling, and modules 32 and 500, which are enriched for Toll-like response genes. This was most striking for FB, for which the largest number of immune response modules differentially expressed was identified (modules 468 and 154 enriched for interferon-alpha responses, module 2 enriched for the Toll-like receptor 9 (TLR9) cascade, module 169 enriched for cytokine signalling (*IL15*, *KIT*, *RAF1*) and module 16 enriched for JAK-STAT signalling), all of which were downregulated with AD.

### **Two-layer neighbourhood analysis of risk genes suggests cell-specific mechanisms of AD susceptibility**

Cell-specific enrichments for risk genes expression provide insights into the genesis AD<sup>5</sup>. Cell specific co-expression signatures can suggest specific functional roles for the risk genes in susceptibility. To explore those relevant to the cerebral microvasculature, we identified genes having the most direct expression correlations (those within a two-layer neighbourhood, i.e., any gene that is either directly connected to an AD risk gene or through at most one other gene) with AD risk genes in the vascular cell-specific co-expression networks generated from

both AD and NDC<sup>30,46</sup>. To discover relationships specifically relevant to disease genesis, we determined the overrepresentation of genes differentially expressed with AD in the neighbourhoods of each GWAS gene in the cell-specific co-expression networks (Figure 4A).

AD risk genes *PICALM*, *SORL1* and *INPP5D* had the largest neighbourhoods (230, 167 and 121 differentially expressed genes, respectively) in the EC co-expression network. The neighbourhood of *PICALM* was enriched in IL-6 signalling genes (*IL6ST*, *STAT3*, *JAK1*), as well as semaphorin (*SEMA5A*, *ARHGEF11*, *SEMA6D*, *ITGA1*, *MYH11*, *PLXNC1*) and NOTCH signalling-related genes (*TNRC6C*, *B4GALT1*, *TFDP2*, *POFUT1*, *MAMLD1*, *TNRC6A*). The *SORL1* neighbourhood also was enriched in genes for pathways involved in immune response (e.g., T-cell activation, TYROBP causal network and cytokine response-related pathways) and proteoglycan metabolism (chondroitin sulfate biosynthesis and proteoglycan metabolism) pathway genes encoding proteins essential for vascular extracellular matrix formation. The functional enrichment of the *INPP5D* neighbourhood showed an overrepresentation of individual differentially expressed genes that were downregulated in insulin- and EGF/EGFR-signalling pathways (Figure 4B).

While the co-expression neighbourhoods of AD risk genes *WWOX*, *CLU* and *CCN2* expressed in FB (e.g., including individually differentially regulated genes *ITPR2*, *ROBO1*, *LHFPL6* and *SLC38A1* [*WWOX*]; *PALD1*, *ZBTB46*, *PDZD2*, *SIL1* [*CLU*]; *SPTBN1*, *ZMIZ1*, *KLF7*, *CACNA2D3* [*CCN2*]) and PC were enriched for genes involved in a range of functions, the neighbourhood of *WWOX* also was enriched in Notch signalling pathway genes and both this neighbourhood and that of *IQCK* were enriched in ECM-related pathways (Figure 4C). In PC, the neighbourhood of *MEF2C* transcription factor (108 genes) (one of the largest) included genes involved in the EGF/EGFR signalling pathway (*EPS8*, *MEF2A*, *MEF2C*, *PLCE1*, *RAF1*) and Toll-like receptor pathways (e.g. *MEF2A*, *PPP2CB*, *MEF2C* and *RIPK2*) (Figure 4D).

### **Pathological growth factor and ECM signalling between vascular cells with AD**

We applied CellChat to identify differentially expressed receptor-ligand pairs responsible for cell-cell signalling (both autocrine and paracrine) in AD<sup>47</sup>. 759 potential ligand-receptor

pairs implicating 48 distinct signalling pathways were identified amongst genes expressed EC, PC and FB (Figure 5). We first focused on assessment of interaction pairs that involved DEG and pathways enriched in co-expression modules specifically with AD. The analysis provided insight into mechanisms contributing to dysregulated angiogenesis with AD. Several VEGF-related signalling pairs were detected only in NDC. Signalling via INHBA (FB/PC) – ACVR1B and ACVR1A (EC) also was found only in NDC. By contrast, evidence for TGFB3 (EC) – TGFB1\_R2 (FB) and the TGFB3 (EC) – ACVR2B\_TGFB2 (FB) signalling was found only in AD samples.

Growth factor signalling important for angiogenesis and vascular homeostasis also was generally decreased with AD. Receptor expression for the EGF(EC/PC)-EGFR(EC/PC), EGF(EC/PC) - EGFR\_ERBB2(PC), NTF3(FB/PC) - NTRK3(FB) communication pairs was reduced. Decreased expression of FB *LAMC1* should reduce autocrine and EC/PC CD44 signalling. Increased expression of *CD46* in PC with AD may modulate NOTCH signalling in an autocrine manner (CD46 (PC)- JAG1(PC))<sup>48</sup>. Autocrine NEGR1(FB)- NEGR1(FB) signalling also increases in AD. Vascular cell interactions with the ECM were reduced with decreased integrin expression on EC affecting multiple ligand-receptor communication pairs: COL1A2 (EC/FB) - ITGB1 (EC), COL1A2 (FB) -ITGA2 (EC), COL4A1 (FB/PC) - ITGA2 (EC) or COL4A2 (EC/FB) - ITGB1 (EC), COL4A4 (PC) - ITGA2\_ITGB1(PC) and COL6A3 (FB) - ITGA2\_ITGB1 (EC). Both lower EC and PC integrin and FB laminin expression in AD also should reduce the many laminin-ECM interactions identified (LAMC1 (FB) - ITGA1\_ITGB1 (EC/PC) - ITGA2\_ITGB1 (EC), LAMA3 (EC) - ITGA2\_ITGB1(EC), LAMA4 (FB/PC) - ITGA2\_ITGB1(EC), LAMB1 (FB) - ITGA2\_ITGB1(EC), LAMC1 (FB) - ITGA2\_ITGB1 (EC/PC), LAMC3 (EB/PC) - ITGA2\_ITGB1 (EC), LAMC3 (FB/PC) - ITGA2\_ITGB1 (EC/PC), LAMC1 (FB) - ITGA6\_ITGB1(EC/PC), LAMC1 (FC) - ITGA7\_ITGB1 (EC/PC)).

**Expression of vascular endothelial angiogenic pathway genes is related to A $\beta$  plaque and phosphorylated Tau (pTau) load**

To explore how vascular transcriptomic pathology may evolve over the course of AD, we performed an exploratory analysis of gene expression as a regression of the individual brain regional A $\beta$  and pTau densities in the 24/57 brain samples used in the analyses above for which quantitative IHC was available. We limited this regression analysis to the EC, the most abundant of the cell populations, to minimise Type I errors. We found 28 genes were significantly (adjusted p<0.1) differentially expressed with greater brain regional A $\beta$  plaque density and 75 genes differentially expressed with greater pTau density. Genes showing significant regression with A $\beta$  density in EC were enriched in VEGF signalling pathway components (*CYFIP1*, *PXN*, *CTNNA1*, *CALM1*). By contrast, genes that showed significant regression against pTau were enriched in EGFR, IGF1R and insulin receptor signalling pathways (*PPP2R5E*, *NF1*, *CALM1*, *DUSP16*, *AP2M1*, *RBX1*).

## Discussion

Brain vascular structural pathology and physiological dysfunction is characteristic of both preclinical models expressing brain A $\beta$  and of AD<sup>13,49</sup>. We found that, although EC, SMC and PC all express AD risk genes, EC appear to be uniquely *significantly* enriched amongst vascular cells, suggesting a quantitatively more important role in “causally” mediating AD susceptibility. Our results provided evidence that pathological regulation of angiogenic gene expression contributes to the early vascular impairments in AD: we found upregulation of the pro-angiogenic *HIF1A* and increased expression of mitochondrial oxidative and fatty acid oxidation genes, metabolic drivers of angiogenesis<sup>50</sup>, in conjunction with *downregulation* of vascular developmental and homeostatic pathways involving EGF/EGFR, VEGF/VEGFR and insulin and IGF signalling in EC and PC and reduced expression of laminin and collagen IV genes in FB. The correlation of these expression differences in angiogenic growth factor pathways in EC with both the A $\beta$  plaque and pTau burden suggests that impairments of vascular growth factor pathway signalling worsen progressively through the course of the disease. Analyses of cell-cell communication provided novel evidence for decreased signalling

though VEGFR2 and components of the ECM to EC, decreased expression of the LDLR pathway for clearance of A $\beta$  from the CNS by EC, interferon pathway activation and increased inflammatory responses in EC and PC. Together, these processes would be expected to reduce clearance of A $\beta$  while also, via inflammatory activation, increasing pathological A $\beta$  production and processing<sup>29</sup>. Specific evidence for this was found with upregulation in EC of *PSEN2* and *APH1A*, which encode for components of the gamma secretase.

A recent preprint also reported that GWAS risk genes were enriched in EC and vascular mural cells and suggested an “evolutionary shift” of AD risk gene expression from a singular predominance in microglia in the mouse to microglial and vascular cells in humans<sup>15</sup>. Our analysis extends this by showing that, among vascular cells, risk genes associated with AD are enriched significantly in EC, suggesting an involvement of EC in the early genesis of AD<sup>5</sup>. Moreover, we showed that the risk genes enriched for expression in EC overlap substantially with those in microglia (Figure 1D).

Co-expression and two-layer neighbourhood analyses provided insights into some possible functional roles for proteins encoded by AD risk genes expressed in the vascular cells. For example, our results showed lower expression of *PICALM* in EC with AD, suggesting a mechanism by which vascular clearance of A $\beta$  is reduced in AD<sup>27,51</sup>. Functionally less well characterised AD GWAS genes, *WWOX* and *IQCK*, have large neighbourhoods in the FB and PC co-expression networks associated with enrichment for pathways supporting maintenance of BBB integrity<sup>9</sup>. Finally, *INPP5D* had consistently one of the largest two-layer neighbourhoods across the vascular cell types, whereas the enriched pathways associated with them were cell type specific (e.g., EGF/EGFR in EC and ECM components in FB).

Innate immune responses are central to AD pathogenesis and progression but have not been well defined in the microvasculature to date<sup>29,52</sup>. We found evidence for cell-specific differences in vascular inflammatory responses to AD in EC with upregulation of interferon signalling genes in EC. PC also appear to play a role in vascular inflammatory mediation of early AD. Recently identified risk genes *CD46*, encoding a serine protease which mediates

inactivation of complement proteins, and *IRAK3*, encoding a homeostatic mediator of innate immune responses<sup>53</sup> were upregulated and downregulated, respectively, in PC.

However, the most strikingly differentially expressed gene sets in AD are involved in angiogenesis and vascular homeostasis. VEGF/VEGFR and insulin signalling pathways<sup>54</sup> in EC and EGF/EGFR signalling in EC and PC were downregulated with AD<sup>55</sup> and negatively correlated with A $\beta$  pathology, despite upregulation of other genes (e.g. *HIF1A*) associated with pro-angiogenic regulation<sup>28</sup>. These results add to prior evidence of *dysfunctional angiogenesis* in AD<sup>16,43,49</sup>. We have extended descriptions by showing that, despite angiogenic signals (e.g., upregulation of *HIF1A*) and metabolic adaptations, major downstream effector pathways fail to respond at the transcriptional level.

Previous reports also have implicated dysfunction of both EC and PC in AD<sup>56</sup>. Our observations emphasise the extent to which pathology of *FB* also contribute to vascular abnormalities in AD, with NOTCH signalling as a candidate mechanism central for this. Multiple genes encoding for ECM components were downregulated in FB, which should contribute to BBB dysfunction<sup>57,58</sup>, impair angiogenesis<sup>35</sup> and lead to altered expression of tight junction proteins in EC<sup>58</sup> via reduced signalling between ECM proteins in FB and integrins in EC<sup>59</sup>.

Although we have made a number of novel observations, our analyses had limitations which need to be addressed in future work. Our data was generated from nuclei from multiple brain regions and thus could address robustly only those transcriptomic differences that were common to all of these regions in AD, even given that we took the confound of brain region into account as a fixed effect in the statistical models. A second limitation was that we assessed the total extracted populations of nuclei without seeking to identify and separately study cells expressed from specific vascular anatomic zones<sup>15,17</sup>. Nevertheless, the high overlap of our cellular markers and the cellular markers from human and mouse brain vessel-associated cells in previous reports provides some confidence that all major cell types were represented. Third, like other recent studies, our conclusions are based on relatively sparse (10X Genomics Single Cell 3' Gene Expression assay) sequencing of the nuclear transcriptome,

which may be biased relative to those from the whole cell, potentially reducing the power to detect transcripts from some genes<sup>60</sup>. Use of larger numbers of nuclear and co-expression-based analyses, which rely less on detection of absolute expression levels than do single gene differential expression analyses, may have reduced the impact of this although the impact of this limitation, but this is difficult to assess without future, more comprehensive transcriptional analyses of the whole cells.

Impairment of angiogenesis and vascular homeostasis, reduced endothelial A $\beta$  clearance with reduced expression of *PICALM* and increased production of A $\beta$  by EC with upregulation of interferon (*IFITM3*) and  $\gamma$ -secretase component genes all will act to increase toxic A $\beta$  concentrations in the brain<sup>29</sup>. The extraordinary length of the brain capillary network (~650 km) and its large surface area (~120 cm<sup>2</sup>/g) suggest that even small relative effects could contribute substantially to increasing the overall A $\beta$  burden in the CNS<sup>61</sup>. The identification of significant EC enrichment in AD risk genes also suggests that their specific contribution to inflammatory activation and reduced A $\beta$  clearance are early, potentially “causal” factors in the onset of sporadic, late onset AD. Our work suggests specific mechanisms by which small vessel disease from many causes (e.g., metabolic disease, hypertension, smoking) could potentiate early AD and add to the rationale for AD prevention through interventions for control of modifiable cardiometabolic risk factors<sup>62</sup>. More generally, our results suggest that EC therapeutic targets related to angiogenic, inflammatory and A $\beta$  clearance pathways deserve prioritisation in the search for treatments able to slow or prevent the onset of AD.

## Methods

Data for this study was generated from cortical brain tissue processed locally as we described earlier<sup>63</sup> or from that described and made available publicly previously<sup>18</sup>.

### Local tissue access and snRNA sequencing

#### *Brain cortical tissue sequencing*

Local tissue access and data generation was carried out in accordance with the Regional Ethics Committee and Imperial College Use of Human Tissue guidelines. Tissues and processing were described previously<sup>63</sup>. Cases were selected from the London Neurodegeneration (King's College London) and Parkinson's UK (Imperial College London) Brain Banks. Entorhinal and somatosensory cortex from 6 non-diseased control (NDC) cases (Braak stage 0-II) and 6 AD cases (Braak stage III-VI) were used (total of 24 brain samples). Brains used for this study excluded cases with clinical or pathological evidence for small vessel disease, stroke, cerebral amyloid angiopathy, diabetes, Lewy body pathology (TDP-43), or other neurological diseases. Where the information was available, cases were selected for a brain pH greater than 6 and all but one had a *post mortem* delay of less than 24 hr.

**Table 1: Cohort information for locally processed samples**

	M:F ratio	Age at death (yrs, mean +/- SD)	Post mortem delay (hr, mean +/ - SD)	RIN (mean +/- SD)
<b>Non-diseased controls (Braak 0-II)</b>	4:2	79.3 +/ - 6.5	18 +/ - 6.9	4.9 +/ - 2.0
<b>Alzheimer's disease (Braak III-VI)</b>	4:2	81 +/ - 6.8	22.1 +/ - 15.9	7.1 +/ - 0.7

### Immunohistochemistry

Immunohistochemical staining was performed on formalin-fixed paraffin-embedded sections from homologous regions of each brain used for snRNASeq locally. Standard immunohistochemical procedures were followed using the ImmPRESS Polymer (Vector Laboratories) and SuperSensitive Polymer-HRP (Biogenex) kits (Table 2). Briefly,

endogenous peroxidase activity and non-specific binding was blocked with 0.3% H<sub>2</sub>O<sub>2</sub> and 10% normal horse serum, respectively. Primary antibodies were incubated overnight at 4°C. Species-specific ImmPRESS or SuperSensitive kits and DAB were used for antibody visualization. Counter-staining for nuclei was performed by incubating tissue sections in haematoxylin (TCS Biosciences) for 2 min. AD pathology was assessed by Aβ plaque (4G8, BioLegend 17-24) and pTau (AT8, NBS Biologics) staining.

**Table 2. List of antibodies and immunostaining methods**

Antigen	Antibody	Dilution	Antigen Retrieval	IHC Staining Kit
<b>Aβ</b>	4G8		Citrate	
	BioLegend (800702)	1:15,000	Buffer, in Steamer	Supersensitive Kit
<b>pTau</b>	AT8		Citrate	
	Invitrogen (MN1020)	1:1,600	Buffer, in Steamer	Supersensitive Kit

Labelled tissue sections were imaged using a Leica Aperio AT2 Brightfield Scanner (Leica Biosystems). Images were analysed using HALO software (Indica Labs, Version 2.3.2989.34). The following image analysis macros were used for the study: area quantification macro (amyloid), multiplex macro (pTau).

#### *Nuclei isolation and enrichment for lower abundancy cell populations*

Local processing of the fresh frozen entorhinal and somatosensory cortical tissue blocks began with sectioning to 80 µm on a cryostat and grey matter separated by scoring the tissue with sharp forceps to collect ~200 mg grey matter in an RNase-free Eppendorf tube. Nuclei from NDC and AD samples were isolated in parallel using a protocol based on Krishnaswami *et al.* (2016)<sup>64</sup>. All steps were carried out on ice or at 4°C. Tissue was homogenized in a 2 ml glass douncer containing homogenization buffer (0.1% Triton-X + 0.4 U/µl RNaseIn + 0.2 U/µl SUPERaseIn). The tissue homogenate was centrifuged at 1000 g for 8 min, and the majority

of supernatant removed without disturbing the tissue pellet. Homogenised tissue was filtered through a 70 µm filter and centrifuged in an Optiprep (Sigma) density gradient at 13,000 g for 40 min to remove myelin and cellular debris. The nuclei pellet was washed and filtered twice in PBS buffer (PBS + 1% BSA + 0.2 U/ml RNaseIn). Isolated nuclei were labelled in suspension in 1 ml PBS buffer with 1:500 anti-NeuN antibody (Millipore, MAB377, mouse) and 1:250 anti-Sox10 antibody (R&D, AF2864, goat) for 1 hr on ice. Nuclei were washed twice with PBS buffer and centrifuged at 500 g for 5 min. Nuclei were incubated with Alexa-fluor secondary antibodies at 1:1000 (goat-anti-mouse-647 and donkey-anti-goat-488) and Dapi (1:1000) for 30 min on ice, and washed twice. Nuclei were FACS-sorted on a BD Aria II, using BD FACSDiva software, gating first for Dapi +ve nuclei, then singlets and then Sox10- and NeuN-negative nuclei. A minimum of 150,000 double-negative nuclei were collected.

We also isolated nuclei, from adjacent localizations of the same tissue samples as described above, which were not subjected to the FACS-enrichment step but were directly processed for single nucleus capture and snRNA sequencing as described in the following section. In this way, we obtained an unbiased representation of all the brain cell types. The resulting dataset was used for the analysis described in the “Enrichment of brain cell types in AD and WMH GWAS signal” section.

Sorted nuclei were centrifuged at 500 g, resuspended in 50 µl PBS buffer and counted on a LUNA-FL Dual Fluorescence Cell Counter (Logos Biosystems, L20001) using Acridine Orange dye to stain nuclei. Sufficient nuclei were added for a target of 7,000 nuclei for each library prepared. Barcoding, cDNA synthesis and library preparation were performed using 10X Genomics Single Cell 3' Gene Expression kit v3 with 8 cycles of cDNA amplification, after which up to 25 ng of cDNA was taken through to the fragmentation reaction and a final indexing PCR was carried out to 14 cycles. cDNA concentrations were measured using Qubit dsDNA HS Assay Kit (ThermoFisher, Q32851), and cDNA and library preparations were assessed using the Bioanalyzer High-Sensitivity DNA Kit (Agilent, 5067-4627). Samples were pooled to equimolar concentrations and the pool sequenced across 24 lanes of an Illumina HiSeq 4000 according to the standard 10X Genomics protocol. The snRNAseq data will be

made available for download from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE160936.

## Single Nuclei RNA Sequence Analyses

### *Processing of FASTQ files*

Locally generated snRNASeq data were pre-processed using 10X Genomics Cell Ranger. Illumina sequencing files were aligned to the genomic sequence (introns and exons) using GRCh38 annotation in Cell Ranger v3.1. Nuclei were identified above background by the Cell Ranger software.

### *Quality control, dataset integration, dimension reduction and clustering*

Feature-barcode matrices from CellRanger produced corresponding to the local dataset produced as described above were jointly processed with the feature-barcode matrices from a previously published dataset<sup>18</sup>. that were downloaded from the Gene Expression Omnibus (accession number GSE148822). Together, the two datasets were generated from 57 brain samples. Quality control (QC), dataset integration, dimension reduction and clustering were performed using the Nextflow pipeline nf-core/scflow<sup>65</sup>.

QC was performed separately on each sample. Nuclei that had less than 200 features were excluded, whereas for the higher feature filtering criterion, an adaptive threshold was estimated in each sample, which was four median absolute deviations above the median feature number in the sample. Nuclei with more than 5% of mitochondrial gene counts were also excluded. Only genes that had at least one count in 5 nuclei per sample were retained. The QC also included an ambient RNA profiling using the EmptyDrops package<sup>66</sup> using default parameters. Finally, multiple identification was performed using DoubletFinder<sup>67</sup> using 10 principal components based on the 2000 most variable features and a pK value of 0.005.

Sample integration was performed using the Liger package (v1.0.0)<sup>19</sup> incorporated in the nf-core/scflow pipeline (v0.7.1)<sup>65</sup>. The k value was optimized at 20 and the lambda value at 5. 3000 genes were employed in the integration process. The integration threshold was 0.0001

and the maximum number of performed iterations was set to 100. The normalized cell factors from the integrated dataset were then used as an input for dimension reduction and clustering, which was performed using UMAP<sup>20</sup>. Clustering was performed with the Leiden method using a resolution parameter of 0.00001 and a k value of 50.

Cell-type identification of clusters was performed by plotting canonical cell markers using the FeaturePlot function in Seurat (v3.2.3)<sup>68</sup>. To efficiently separate the vascular mural cells, we isolated the ECs and the vascular mural cell clusters and re-ran the steps of the integration, dimension reduction and clustering. Cluster specific genes were identified using the FindMarkers function in Seurat (using the MAST method<sup>25</sup> with the function arguments set to default). To validate the cell-type specificity of the clusters and their identity, we compared the top 100 cluster markers of our dataset with the top 100 cluster markers of the same cell types from previously published human and mouse datasets<sup>17,21</sup> using an overrepresentation analysis.

### *Overrepresentation analysis*

Overrepresentation analysis was performed to determine if the overlap between two gene sets is significantly higher than if it occurred by chance. This was done using with the "enrichment" function of the R package bc3net (v1.0.4) (<https://github.com/cran/bc3net>), which performs a Fisher's exact test (FET). The p values associated with the Fisher's exact test correspond to the probability that the overlap between the two gene sets and has occurred by chance.

### *Enrichment of brain cell types in AD and WMH GWAS signal*

GWAS summary statistics for AD<sup>3</sup> and WMH (a radiological manifestation of small vessel disease)<sup>24</sup> were tested for enrichment in brain cell types using the MAGMA.Celltyping (v1.0.1)<sup>5,69</sup> and MungeSumstats (v1.1.24)<sup>70</sup> packages. First, summary statistics were appropriately formatted using MungeSumstats for use with MAGMA.Celltyping. Then, SNP associations from the summary statistics were mapped to genes using the map.snp.to.genes

function of MAGMA.Celltyping. Next, as described for the default workflow of MAGMA.Celltyping, genes with low variability between the cell clusters were dropped using the drop\_uninformative\_genes and then quantile groups for each cell type were prepared using the prepare.quantile.groups function.

We first calculated the enrichment in AD and WMH GWAS signal across all the brain cell types on the dataset that had not been subjected to the FACS enrichment step to remove neurons and oligodendrocytes (see “Nuclei isolation and enrichment for lower abundance cell populations” section) (Figure 1D and S9A, respectively). This was performed using the calculate\_celltype\_associations function with default parameters and the “linear” enrichment mode. This analysis was repeated after controlling for the microglial enrichment of the GWAS signal. Next, we calculated the enrichment in AD and WMH GWAS signal on each of the vasculature-associated cell types (EC, FB, PC and SMC) (Figure 1E and S9B, respectively). Finally, we assessed if the enrichment of the vasculature-associated cell types in our dataset in AD GWAS signal changed after controlling for the enrichment in WMH GWAS signal. For this, we re-ran the calculate\_celltype\_associations function for the AD summary statistics and the SNP-to-gene mapping of the WMH GWAS (that was calculated earlier with the map.snp.to.genes function) in the genesOutCOND argument of the function.

#### *Differential gene expression analysis*

DGE analysis was performed using MAST (v1.18.0). The transcriptomic alterations in AD vs NDC samples were assessed separately in each cell type by means of a zero-inflated regression analysis using a mixed-effects model. The use of a mixed-effects model is particularly important in the context of snRNAseq DGE analyses to account for the pseudoreplication bias that would otherwise be observed if a fixed-effects-only model was employed<sup>71</sup>. The model specification was `zlm(~diagnosis + (1|sample) + cngeneson + pc_mito + sex + brain_region, sca, method = "glmer", ebayes = F)`. The fixed effect term `pc_mito` accounts for the percentage of counts mapping to mitochondrial genes. The term `cngeneson` is the cellular detection rate. Each nuclei preparation was considered as a distinct sample for

the mixed effect. Models were fit with and without the dependent variable and compared using a likelihood ratio test. Units for differential expression are defined as log2 fold difference in AD vs NDC nuclei. The inclusion of a “dataset” term in the model was not necessary because the inclusion of the brain region term completely accounted for it. In the subset of samples that corresponded to the dataset produced in our laboratory, we also performed an exploratory regression analysis of gene expression against the two histopathological features (using pTau or A $\beta$  as markers) using MAST. The model specification was `zlm(~histopath_marker + (1|sample) + cngeneson + pc_mito + sex + brain_region, sca, method = "glmer", ebayes = F)`. In this case, units for differential expression are defined as log2 fold difference/% pTau positive cells (or log2 fold difference/% A $\beta$  plaque area), i.e., a one unit change in immunohistochemically-defined pTau (or A $\beta$  plaque) density is associated with one log2-fold change in gene expression. In both MAST analyses, genes expressed in at least 10% of nuclei from each cell type were tested. Genes with an adjusted p-value <0.1 were defined as meaningfully differentially expressed.

#### *Gene ontology and pathway enrichment analysis*

The gene ontology (GO) enrichment and the pathway enrichments analysis were carried out using the R package `enrichR` (v 3.0), which uses Fisher's exact test (Benjamini-Hochberg FDR < 0.1)<sup>72</sup>. Genesets with minimum and maximum genes of 10 and 500 respectively were considered. To improve biological interpretation of functionally related gene ontology and pathway terms and to reduce the number of redundant gene sets, we first calculated a pairwise distance matrix using Cohen's kappa statistics based on the overlapping genes between the enriched terms and then performed hierarchical clustering of the enriched terms<sup>73</sup>.

#### *Gene co-expression analysis*

Gene co-expression modules and hub-genes were identified separately for each cell type using the MEGENA (v1.3.7) package<sup>30</sup>. MEGENA constructs a hierarchy of co-expression

modules with larger (“parent”) that are further divided into subset (“children”) modules. “Children” modules are subsets of the “parent” ones and have higher numbers as names than their “parents”. To reduce the effect of noise, due to the sparsity of the expression matrix in a snRNAseq experiment, a sample-level pseudo-bulking was performed by summing the raw counts of all the nuclei in a sample. Genes expressed in at least 50% of the samples were used as input. The MEGENA pipeline then was applied using default parameters, using Pearson’s correlations and a minimum module size of 10 genes. Parent modules were produced from which a sub-set of genes form smaller child modules. The co-expression module hierarchy was represented graphically using Cytoscape software (Mac OS version 3.8.0)<sup>74</sup> (Figure 3A-C).

#### *Gene Set Enrichment Analyses (GSEA)*

AUCell<sup>75</sup> (R package v1.6.1) was used to quantify the enrichment of the co-expression modules in our nuclei. Normalised data was processed in AUCell using the *AUCell\_buildRankings* function. The resulting rankings, along with the gene lists of interest, were then put into the function *AUCell\_calcAUC* (aucMaxRank set to 5% of the number of input genes).

The statistical comparison of the enrichment of the co-expression modules in our AD nuclei vs the NDC nuclei was performed using the limma package in R<sup>31</sup>. The module enrichment matrix was log2-normalized. The default configuration of the limma package was employed with the following linear model (which corresponds to the model employed in the DGE analysis with MAST package): ~diagnosis+nFeature+pc\_mito+brain\_region+sex, where nFeatures is the total number of distinct features expressed in each nucleus (to account for the fact that nuclei that express a higher number of features may have higher AUCell scores). We also corrected for a potential pseudoreplication bias<sup>71</sup>, by using the *duplicateCorrelation* function of the limma package with the sample as the “blocking” variable.

#### *Cell-cell communication analysis*

Cell-cell communication analysis was performed using CellChat v0.5.5<sup>47</sup>. CellChat employs a curated database of potential signalling ligand-receptor pairs from the literature. Among all these potential ligand-receptor pairs, cell-cell interactions are identified based on mass action models, along with differential expression analysis and statistical tests on cell groups. The CellChat algorithm with default parameters (unless otherwise specified) was applied to the subset of the dataset that corresponded to the ECs, FBs and PCs, separately on the AD and NDC samples. The human CellChat database was used for the ligand-receptor pairs. Communications that involved less than 100 nuclei were filtered out (using the min.cells argument in the filterCommunication function). The ligand-receptor pairs involving genes that showed differential expression in AD compared to the NDC nuclei were identified and are documented in the text as follows: Ligand (expressing celltype(s)) – Receptor (expressing celltype(s)).

## Acknowledgements

We thank the donors and their families for the use of human brain tissue in this study and the London Neurodegeneration and Parkinson's UK Brain Bank staff for making it available. Infrastructure, including the British Heart Foundation Centre of Excellence, the LMS/NIHR Imperial Biomedical Research Centre Flow Cytometry Facility and the Imperial BRC Genomics Facility, which are supported by the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC), all contributed to the work. ST was supported by an "Early Postdoc.Mobility" scholarship (P2GEP3\_191446) from the Swiss National Science Foundation, a "Clinical Medicine Plus" scholarship from the Prof Dr Max Cloëtta Foundation (Zurich, Switzerland) and a scholarship from the University Hospitals of Geneva. PMM acknowledges generous personal support from the Edmond J Safra Foundation and Lily Safra and an NIHR Senior Investigator Award. This work was supported by the UK Dementia Research Institute, which receives its funding from UK DRI Ltd., funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK.

## Data availability

The snRNAseq data will be made available for download from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE160936. Previously described data<sup>18</sup> was downloaded from the GEO database (GSE148822).

## Author Contributions

Stergios Tsartsalis – experimental design, primary data analysis, interpretation of results, drafted the manuscript

Nurun Fancy – experimental design, primary data analysis, interpretation of results

Amy Smith- performed experiments, data analysis, interpretation of results

Karen Davey – performed experiments, interpretation of results

Combiz Khozoe – software developments, support for data analysis

Xin Ye – primary regression analyses, interpretation of results

Nanet Willumsen – performed immunohistochemistry experiments and image analysis

Aisling McGarry- performed immunohistochemistry experiments and image analysis

Robert C.J. Muirhead- performed immunohistochemistry experiments and image analysis

Stephanie Debette – supported data analysis, interpretation of results

David Owen – experimental design, interpretation of results

Paul M. Matthews- experimental design, interpretation of results, drafted the manuscript

All of the co-authors reviewed and approved the final manuscript.

## Competing Interests

PMM has received consultancy fees from Roche, Adelphi Communications, Celgene, Neurodiem and Medscape. He has received honoraria or speakers' fees from Novartis and Biogen and has received research or educational funds from Biogen, Novartis and GlaxoSmithKline.

## References

- 1 Scheltens, P. *et al.* Alzheimer's disease. *Lancet (London, England)* **388**, 505-517, doi:10.1016/S0140-6736(15)01124-1 (2016).
- 2 Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430, doi:10.1038/s41588-019-0358-2 (2019).
- 3 Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* **51**, 404-413, doi:10.1038/s41588-018-0311-9 (2019).
- 4 Andrews, S. J., Fulton-Howard, B. & Goate, A. Interpretation of risk loci from genome-wide association studies of Alzheimer's disease. *Lancet Neurol* **19**, 326-335, doi:10.1016/S1474-4422(19)30435-1 (2020).
- 5 Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat Genet* **50**, 825-833, doi:10.1038/s41588-018-0129-5 (2018).
- 6 Iadecola, C. The overlap between neurodegenerative and vascular factors in the pathogenesis of dementia. *Acta Neuropathol* **120**, 287-296, doi:10.1007/s00401-010-0718-6 (2010).
- 7 Bell, R. D. *et al.* Pericytes control key neurovascular functions and neuronal phenotype in the adult brain and during brain aging. *Neuron* **68**, 409-427, doi:10.1016/j.neuron.2010.09.043 (2010).
- 8 Sweeney, M. D. *et al.* Vascular dysfunction-The disregarded partner of Alzheimer's disease. *Alzheimer's & Dementia* **15**, 158-167, doi:10.1016/j.jalz.2018.07.222 (2019).
- 9 Sweeney, M. D., Sagare, A. P. & Zlokovic, B. V. Blood-brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. *Nat Rev Neurol* **14**, 133-150, doi:10.1038/nrneurol.2017.188 (2018).
- 10 Hall, C. N. *et al.* Capillary pericytes regulate cerebral blood flow in health and disease. *Nature* **508**, 55-60, doi:10.1038/nature13165 (2014).
- 11 Venkataraman, A. V. *et al.* Widespread cell stress and mitochondrial dysfunction in early Alzheimer's Disease. *medRxiv*, 2021.2008.2011.21261851, doi:10.1101/2021.08.11.21261851 (2021).
- 12 Montagne, A. *et al.* APOE4 leads to blood-brain barrier dysfunction predicting cognitive decline. *Nature* **581**, 71-76, doi:10.1038/s41586-020-2247-3 (2020).
- 13 Korte, N., Nortley, R. & Attwell, D. Cerebral blood flow decrease as an early pathological mechanism in Alzheimer's disease. *Acta Neuropathologica* **140**, 793-810, doi:10.1007/s00401-020-02215-w (2020).
- 14 Nation, D. A. *et al.* Blood-brain barrier breakdown is an early biomarker of human cognitive dysfunction. *Nature Medicine* **25**, 270-276, doi:10.1038/s41591-018-0297-y (2019).
- 15 Yang, A. C. *et al.* A human brain vascular atlas reveals diverse cell mediators of Alzheimer's disease risk. *bioRxiv*, doi:doi.org/10.1101/2021.04.26.441262 (2021).
- 16 Lau, S. F., Cao, H., Fu, A. K. Y. & Ip, N. Y. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. *Proc Natl Acad Sci U S A* **117**, 25800-25809, doi:10.1073/pnas.2008762117 (2020).
- 17 Garcia, F. J. *et al.* Single-cell dissection of the human cerebrovasculature in health and disease. *bioRxiv*, doi:doi.org/10.1101/2021.04.26.440975 (2021).
- 18 Gerrits, E. *et al.* Distinct amyloid-beta and tau-associated microglia profiles in Alzheimer's disease. *Acta Neuropathol* **141**, 681-696, doi:10.1007/s00401-021-02263-w (2021).
- 19 Welch, J. D. *et al.* Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell* **177**, 1873-1887 e1817, doi:10.1016/j.cell.2019.05.006 (2019).
- 20 Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**, 5233, doi:10.1038/s41598-019-41695-z (2019).
- 21 Vanlandewijck, M. *et al.* A molecular atlas of cell types and zonation in the brain vasculature. *Nature* **554**, 475-480, doi:10.1038/nature25739 (2018).
- 22 Alosco, M. L. *et al.* A Clinicopathological Investigation of White Matter Hyperintensities and Alzheimer's Disease Neuropathology. *J Alzheimers Dis* **63**, 1347-1360, doi:10.3233/JAD-180017 (2018).
- 23 Attems, J. & Jellinger, K. A. The overlap between vascular disease and Alzheimer's disease--lessons from pathology. *BMC Med* **12**, 206, doi:10.1186/s12916-014-0206-2 (2014).

24 Sargurupremraj, M. *et al.* Cerebral small vessel disease genomics and its implications across the lifespan. *Nat Commun* **11**, 6285, doi:10.1038/s41467-020-19111-2 (2020).

25 Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* **16**, 278, doi:10.1186/s13059-015-0844-5 (2015).

26 Mukherjee, A. *et al.* Regulation of Notch signalling by non-visual beta-arrestin. *Nat Cell Biol* **7**, 1191-1201, doi:10.1038/ncb1327 (2005).

27 Zhao, Z., Nelson, A. R., Betsholtz, C. & Zlokovic, B. V. Establishment and Dysfunction of the Blood-Brain Barrier. *Cell* **163**, 1064-1078, doi:10.1016/j.cell.2015.10.067 (2015).

28 Rajapaksa, U. S., Jin, C. & Dong, T. Malignancy and IFITM3: Friend or Foe? *Front Oncol* **10**, 593245, doi:10.3389/fonc.2020.593245 (2020).

29 Hur, J. Y. *et al.* The innate immunity protein IFITM3 modulates gamma-secretase in Alzheimer's disease. *Nature* **586**, 735-740, doi:10.1038/s41586-020-2681-2 (2020).

30 Song, W. M. & Zhang, B. Multiscale Embedded Gene Co-expression Network Analysis. *PLoS Comput Biol* **11**, e1004574, doi:10.1371/journal.pcbi.1004574 (2015).

31 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).

32 Favarra, D. M. *et al.* Elevated expression of the adhesion GPCR ADGRL4/ELTD1 promotes endothelial sprouting angiogenesis without activating canonical GPCR signalling. *Sci Rep* **11**, 8870, doi:10.1038/s41598-021-85408-x (2021).

33 Benson, C. E. & Southgate, L. The DOCK protein family in vascular development and disease. *Angiogenesis* **24**, 417-433, doi:10.1007/s10456-021-09768-8 (2021).

34 Nitzsche, A. *et al.* Paladin is a phosphoinositide phosphatase regulating endosomal VEGFR2 signalling and angiogenesis. *EMBO Rep* **22**, e50218, doi:10.1525/embr.202050218 (2021).

35 Rajan, A. M., Ma, R. C., Kocha, K. M., Zhang, D. J. & Huang, P. Dual function of perivascular fibroblasts in vascular stabilization in zebrafish. *PLoS genetics* **16**, e1008800, doi:10.1371/journal.pgen.1008800 (2020).

36 Okonkwo, U. A. *et al.* Compromised angiogenesis and vascular Integrity in impaired diabetic wound healing. *PLoS One* **15**, e0231962, doi:10.1371/journal.pone.0231962 (2020).

37 Mae, M. A. *et al.* Single-Cell Analysis of Blood-Brain Barrier Response to Pericyte Loss. *Circ Res* **128**, e46-e62, doi:10.1161/CIRCRESAHA.120.317473 (2021).

38 Yamaji, R. *et al.* Hypoxic up-regulation of triosephosphate isomerase expression in mouse brain capillary endothelial cells. *Arch Biochem Biophys* **423**, 332-342, doi:10.1016/j.abb.2004.01.003 (2004).

39 Mao, B., Yi, Y., Mo, Q., Yang, C. & Zhong, Q. Metabolic profiling reveals the heterogeneity of vascular endothelial function phenotypes in individuals at extreme cardiovascular risk. *RSC Advances* **9**, 30033-30044, doi:10.1039/C9RA05526F (2019).

40 Choi, C. S. *et al.* Liver-specific pyruvate dehydrogenase complex deficiency upregulates lipogenesis in adipose tissue and improves peripheral insulin sensitivity. *Lipids* **45**, 987-995, doi:10.1007/s11745-010-3470-8 (2010).

41 Ni, Y. *et al.* Therapeutic inhibition of miR-802 protects against obesity through AMPK-mediated regulation of hepatic lipid metabolism. *Theranostics* **11**, 1079-1099, doi:10.7150/thno.49354 (2021).

42 Nalecz, K. A. Solute Carriers in the Blood-Brain Barier: Safety in Abundance. *Neurochem Res* **42**, 795-809, doi:10.1007/s11064-016-2030-x (2017).

43 Raut, S., Patel, R. & Al-Ahmad, A. J. Presence of a mutation in PSEN1 or PSEN2 gene is associated with an impaired brain endothelial cell phenotype in vitro. *Fluids Barriers CNS* **18**, 3, doi:10.1186/s12987-020-00235-y (2021).

44 Jefferies, C., Wynne, C. & Higgs, R. Antiviral TRIMs: friend or foe in autoimmune and autoinflammatory disease? *Nature Reviews Immunology* **11**, 617-625, doi:10.1038/nri3043 (2011).

45 Proneth, B. & Conrad, M. Ferroptosis and necroinflammation, a yet poorly explored link. *Cell Death Differ* **26**, 14-24, doi:10.1038/s41418-018-0173-9 (2019).

46 Wang, Q. *et al.* The landscape of multiscale transcriptomic networks and key regulators in Parkinson's disease. *Nat Commun* **10**, 5234, doi:10.1038/s41467-019-13144-y (2019).

47 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088, doi:10.1038/s41467-021-21246-9 (2021).

48 Le Frie, G. *et al.* The CD46-Jagged1 interaction is critical for human TH1 immunity. *Nature immunology* **13**, 1213-1221, doi:10.1038/ni.2454 (2012).

49 Alvarez-Vergara, M. I. *et al.* Non-productive angiogenesis disassembles Aß plaque-associated blood vessels. *Nat Commun* **12**, 3098, doi:10.1038/s41467-021-23337-z (2021).

50 Rohlenova, K. *et al.* Single-Cell RNA Sequencing Maps Endothelial Metabolic Plasticity in Pathological Angiogenesis. *Cell Metab* **31**, 862-877 e814, doi:10.1016/j.cmet.2020.03.009 (2020).

51 Zhao, Z. *et al.* Central role for PICALM in amyloid-beta blood-brain barrier transcytosis and clearance. *Nat Neurosci* **18**, 978-987, doi:10.1038/nn.4025 (2015).

52 Heneka, M. T. *et al.* Neuroinflammation in Alzheimer's disease. *Lancet Neurol* **14**, 388-405, doi:10.1016/S1474-4422(15)70016-5 (2015).

53 Nho, K. *et al.* Genome-wide transcriptome analysis identifies novel dysregulated genes implicated in Alzheimer's pathology. *Alzheimers Dement* **16**, 1213-1223, doi:10.1002/alz.12092 (2020).

54 Walker, A. M. N. *et al.* Endothelial Insulin Receptors Promote VEGF-A Signaling via ERK1/2 and Sprouting Angiogenesis. *Endocrinology* **162**, doi:10.1210/endocr/bqab104 (2021).

55 Yang, H. K. *et al.* Downregulation of LRIG2 expression inhibits angiogenesis of glioma via EGFR/VEGF-A pathway. *Oncol Lett* **14**, 4021-4028, doi:10.3892/ol.2017.6671 (2017).

56 Kisler, K., Nelson, A. R., Montagne, A. & Zlokovic, B. V. Cerebral blood flow regulation and neurovascular dysfunction in Alzheimer disease. *Nat Rev Neurosci* **18**, 419-434, doi:10.1038/nrn.2017.48 (2017).

57 Munji, R. N. *et al.* Profiling the mouse brain endothelial transcriptome in health and disease models reveals a core blood-brain barrier dysfunction module. *Nat Neurosci* **22**, 1892-1902, doi:10.1038/s41593-019-0497-x (2019).

58 Thomsen, M. S., Routhe, L. J. & Moos, T. The vascular basement membrane in the healthy and pathological brain. *J Cereb Blood Flow Metab* **37**, 3300-3317, doi:10.1177/0271678X17722436 (2017).

59 Mezu-Ndubuisi, O. J. & Maheshwari, A. The role of integrins in inflammation and angiogenesis. *Pediatr Res* **89**, 1619-1626, doi:10.1038/s41390-020-01177-9 (2021).

60 Thrupp, N. *et al.* Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell Rep* **32**, 108189, doi:10.1016/j.celrep.2020.108189 (2020).

61 Procter, T. V., Williams, A. & Montagne, A. Interplay between Brain Pericytes and Endothelial Cells in Dementia. *Am J Pathol*, doi:10.1016/j.ajpath.2021.07.003 (2021).

62 Ngandu, T. *et al.* A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *Lancet (London, England)* **385**, 2255-2263, doi:10.1016/S0140-6736(15)60461-5 (2015).

63 Smith, A. M. *et al.* Diverse human astrocyte and microglial transcriptional responses to Alzheimer's pathology. *Acta Neuropathol In Press* (2021).

64 Krishnaswami, S. R. *et al.* Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat Protoc* **11**, 499-524, doi:10.1038/nprot.2016.015 (2016).

65 Khozoie, C. *et al.* scFlow: A Scalable and Reproducible Analysis Pipeline for Single-Cell RNA Sequencing Data. *bioRxiv*, doi:10.1101/2021.08.16.456499 (2021).

66 Lun, A. T. L. *et al.* EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol* **20**, 63, doi:10.1186/s13059-019-1662-y (2019).

67 McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Systems* **8**, 329-337.e324, doi:<https://doi.org/10.1016/j.cels.2019.03.003> (2019).

68 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502, doi:10.1038/nbt.3192 (2015).

69 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219, doi:10.1371/journal.pcbi.1004219 (2015).

70 Murphy, A. E. & Skene, N. G. MungeSumstats: A Bioconductor package for the standardisation and quality control of many GWAS summary statistics. *bioRxiv*, doi:10.1101/2021.06.21.449239 (2021).

71 Zimmerman, K. D., Espeland, M. A. & Langefeld, C. D. A practical solution to pseudoreplication bias in single-cell studies. *Nat Commun* **12**, 738, doi:10.1038/s41467-021-21038-1 (2021).

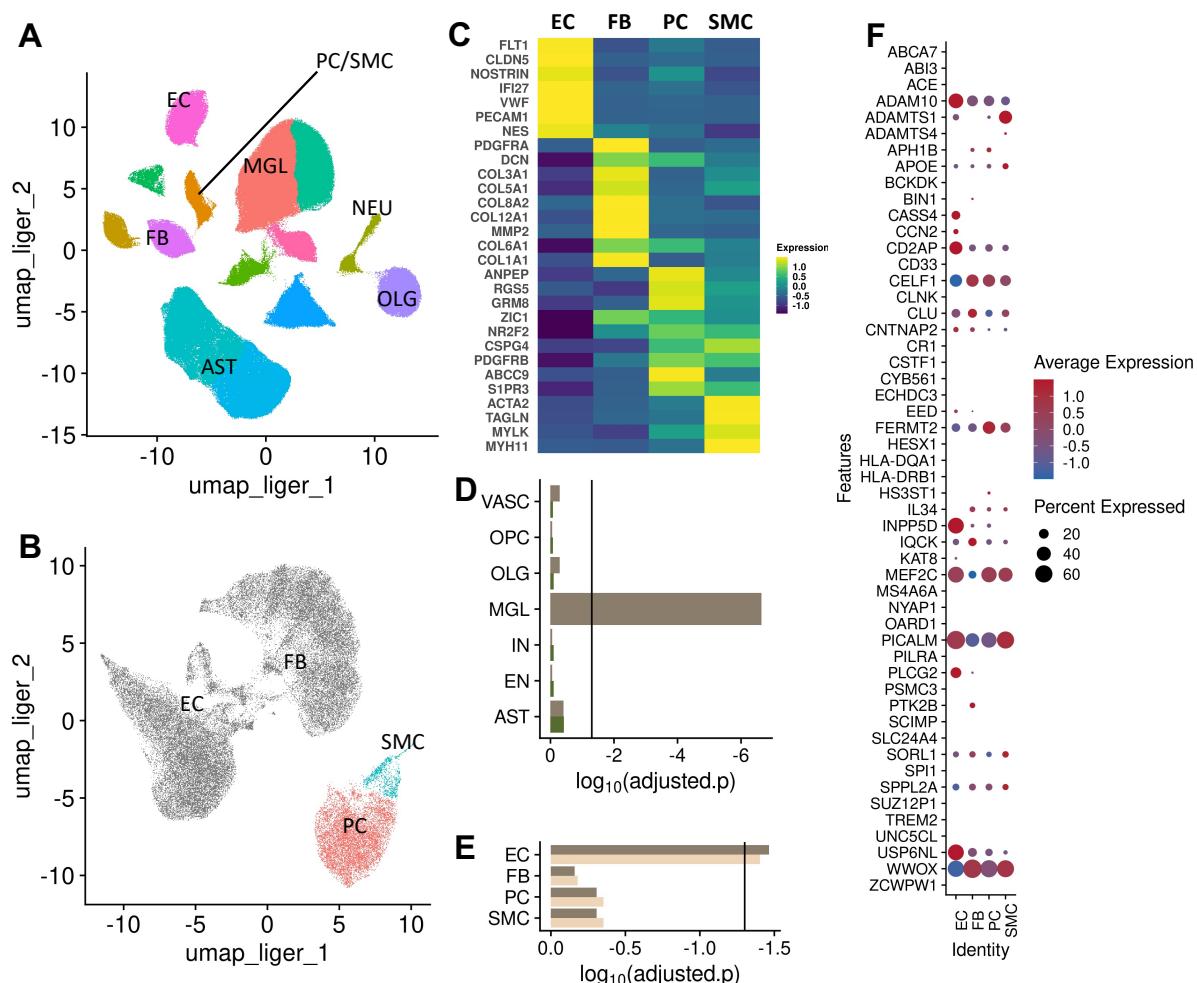
72 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128, doi:10.1186/1471-2105-14-128 (2013).

73 Huang, D. W. *et al.* The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* **8**, R183, doi:10.1186/gb-2007-8-9-r183 (2007).

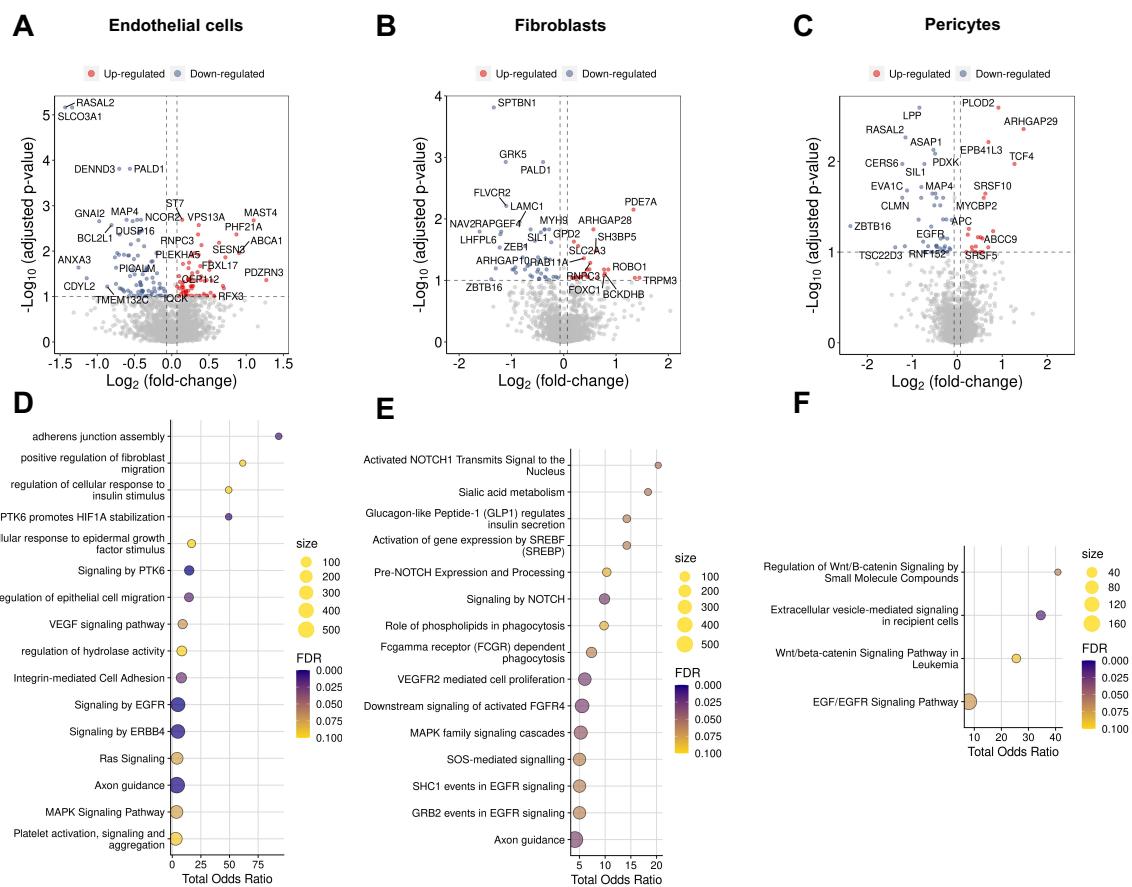
74 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

75 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086, doi:10.1038/nmeth.4463 (2017).

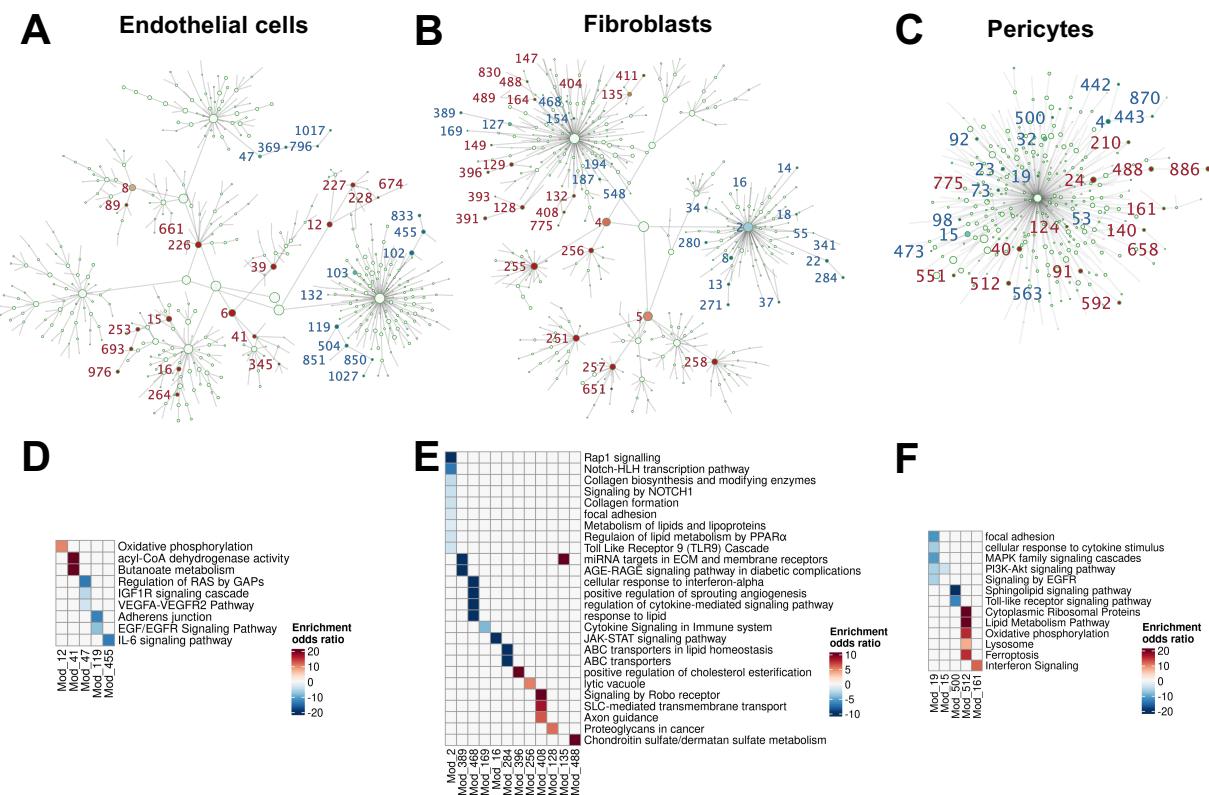
## Figures



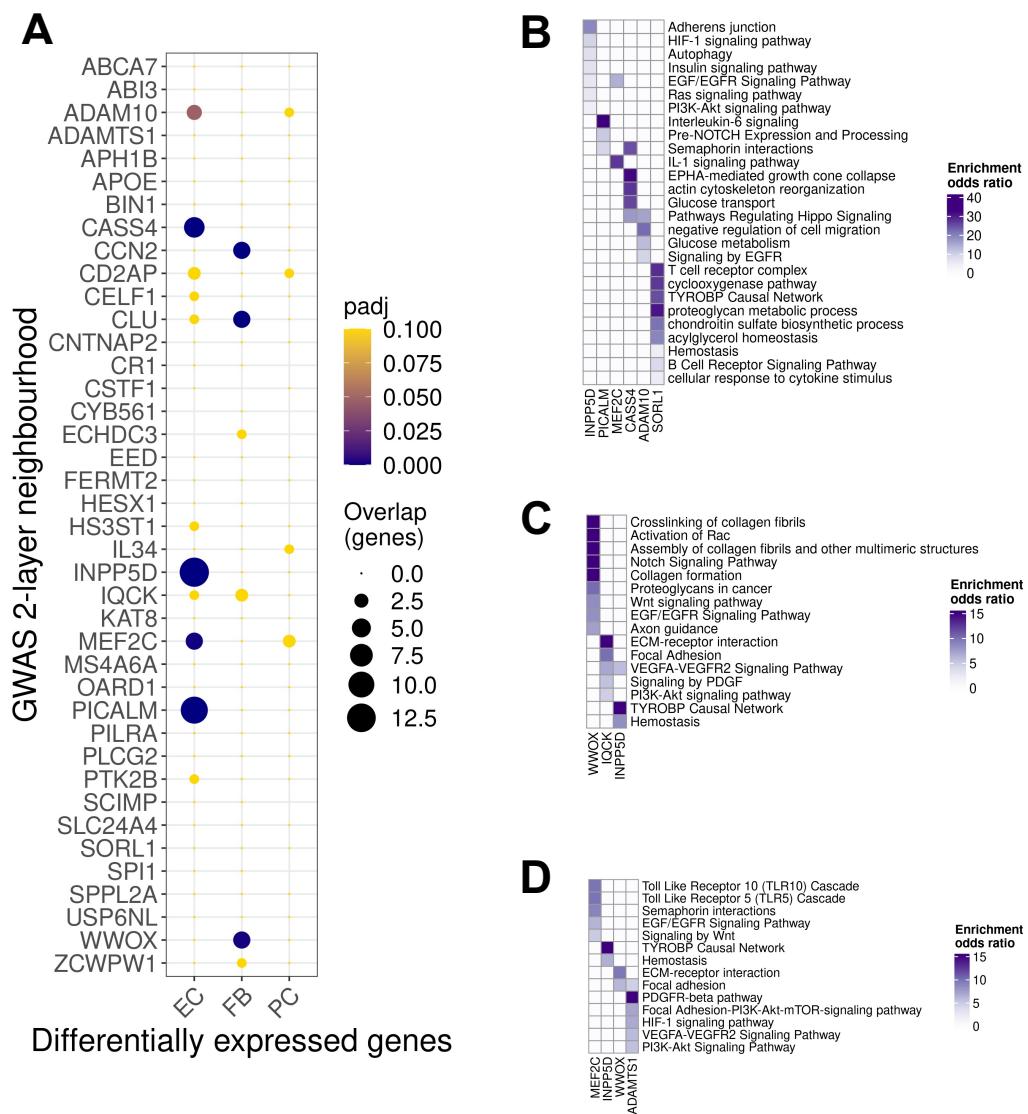
**Figure 1. Characterisation of cell-type specific transcriptomes and their relative enrichment in Alzheimer's disease risk genes.** (A) UMAP plot of the integrated snRNASeq dataset from 57 brain samples. (B) UMAP plot after re-integration and clustering of the EC, FB, PC and SMC nuclei in (A), allowing discrimination between PC (red) and SMC (cyan) nuclei (EC and FB, grey). (C) Heatmap of the average scaled expression of representative marker genes for each cluster. (D) MAGMA.Celltyping enrichment of brain nuclei in genomic loci associated with genetic risk for AD. The bars correspond to the  $\log_{10} p$  value of the enrichment (dark brown, line indicates significance threshold adjusted for all cell types). Enrichment of vascular nuclei is reduced after controlling for genes enriched in microglia (dark green) (F) MAGMA.Celltyping AD risk gene enrichment of nuclei of the brain vasculature (dark brown bar, line indicates significance threshold adjusted for vascular cell types). Enrichment is not changed substantially after controlling for the enrichment of genetic loci associated with white matter hyperintensities (WMH) (light brown). (H) Dot plot of the average scaled per cluster expression of genes previously associated with genetic risk for AD (size, percentage of nuclei per cluster with >1 count; colour scale, average scaled gene expression). Abbreviations: AST, astrocytes; EC, endothelial cells; FB, fibroblasts; MGL, microglia; NEU, neurons; EN, excitatory neurons; IN, inhibitory neurons; OLG, oligodendrocytes; PC/SMC, pericytes and smooth muscle cells.



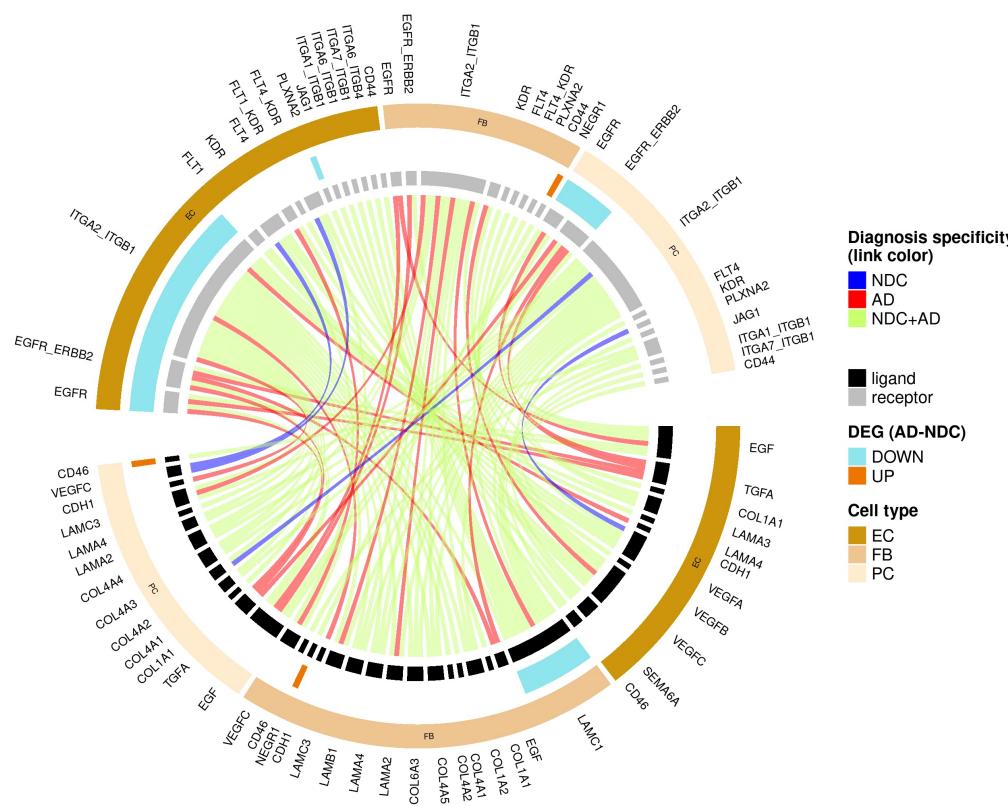
**Figure 2 Alzheimer's disease is associated with dysregulation of vascular homeostasis.** Volcano plots showing genes differentially expressed in AD relative to NDC in (A) EC, (B) FB and (C) PC. Corresponding dot plots of the functional enrichment analysis on the DEG (D-F, (dot size, functional enrichment gene set size; colour, FDR).



**Figure 3 Gene co-expression module hierarchy for (A) EC, (B) FB and (C) PC.** Modules that are significantly differentially enriched in AD relative to NDC nuclei are represented as points in the graph (red, upregulated modules with AD; blue, downregulated modules). (D-F) Heatmaps of odds ratios from the functional enrichment analyses for significantly differentially enriched modules in EC (D), FB (E) and PC (F) (red, upregulated; blue, downregulated).

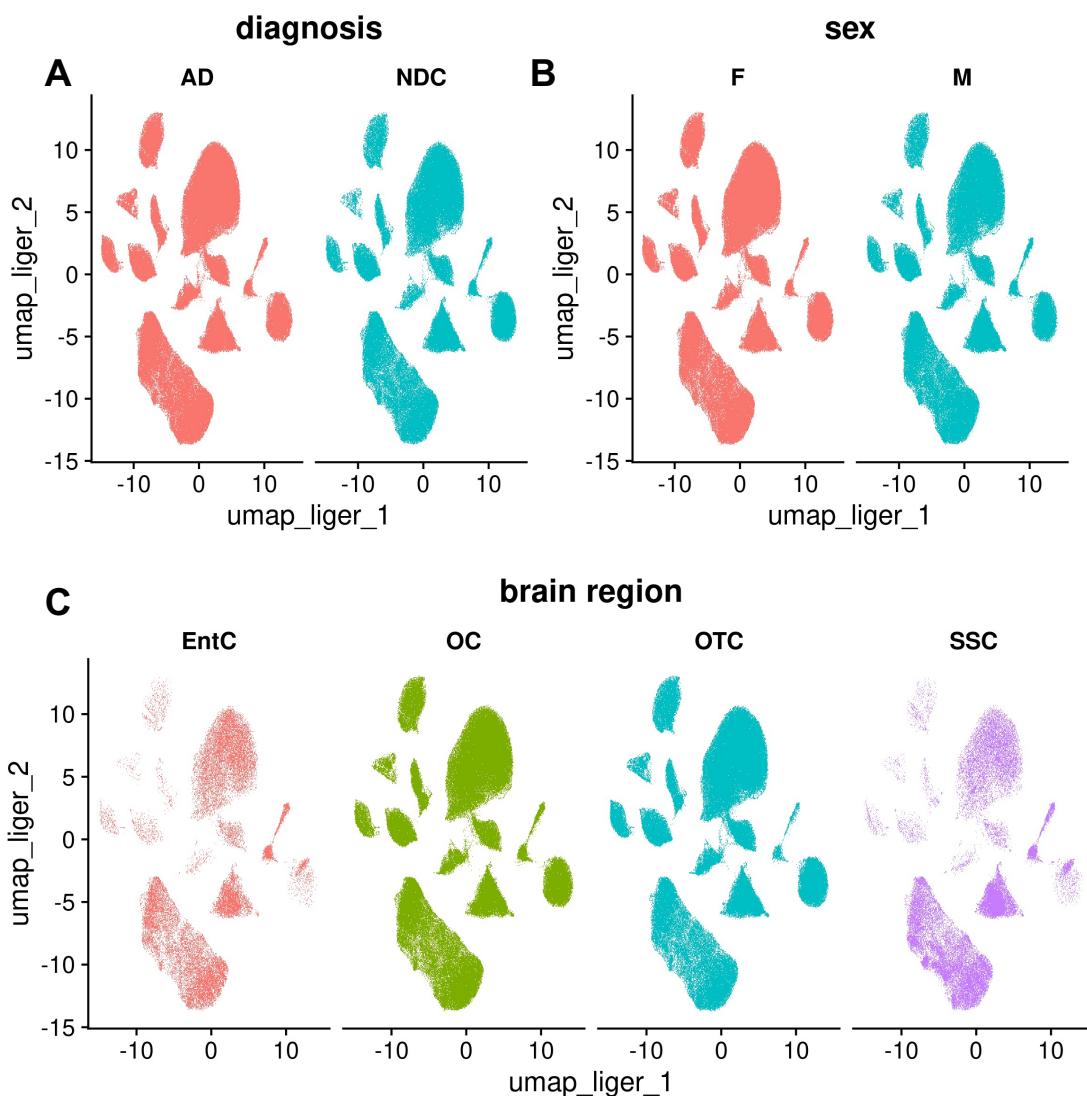


**Figure 4 Exploration differentially expressed genes (DEG, AD vs. NDC) in two-layer neighbourhoods of AD risk genes.** (A) Dot plot of the overrepresentation of DEG identified in each cluster (abscissa) in AD relative to CTR in the 2-layer neighbourhood of each GWAS gene (ordinate) (dot size, number of the overlapping genes; colour, adjusted p value). B-D) Functional enrichment of prioritized GWAS genes in EC (B), FB (C) and PC (D) (colour scales represent the odds ratio of the enrichment).

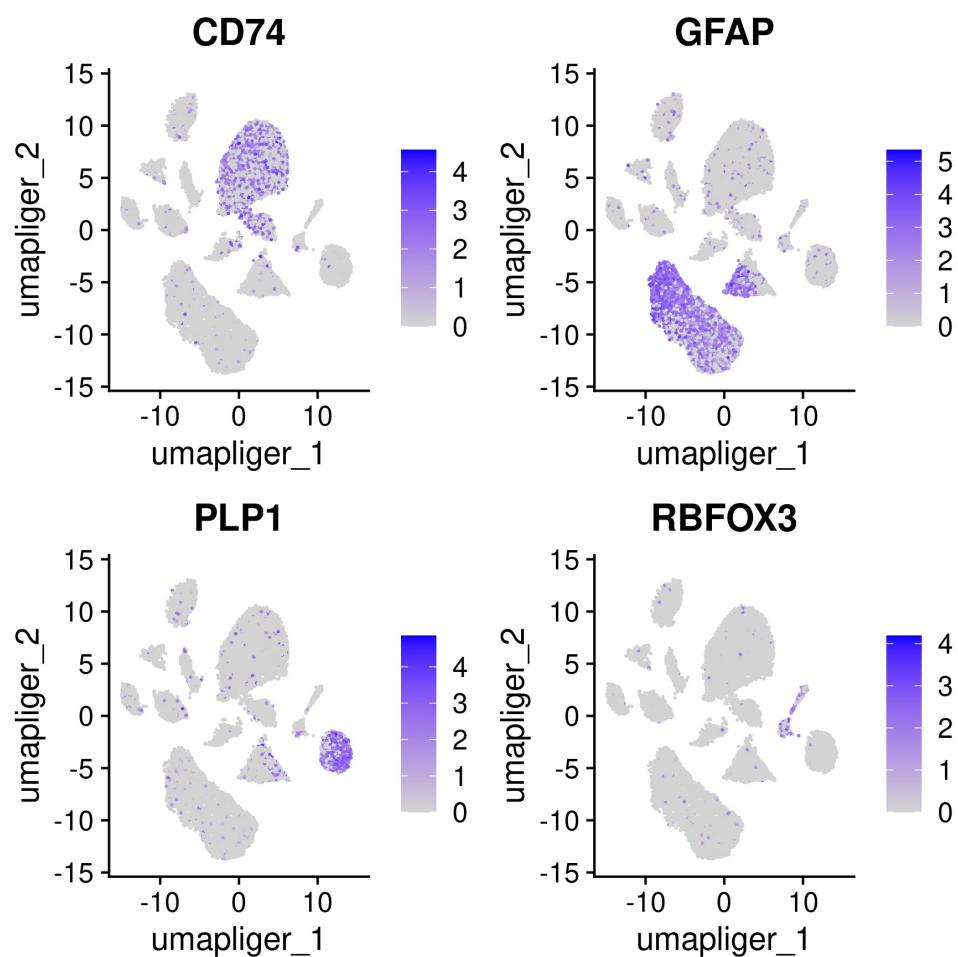


**Figure 5 Circular plot of representative results from the cell-cell communication analysis (CellChat v0.5.5).** The innermost circle links connect the gene symbols of the ligands with their corresponding receptors. The link colour denotes the diagnosis specificity (i.e. if a ligand-receptor pair was identified only in NDC (blue), only in AD (red) or in both conditions (green)). Moving outward, the second track of the plot describes whether the gene corresponds to a ligand (black sectors) or a receptor (grey) in the communication. Genes that were also identified as differentially expressed with AD are highlighted (cyan, downregulated; orange, upregulated). Finally, the cell type where the ligand or the receptor gene is expressed is denoted in the outermost track of the plot.

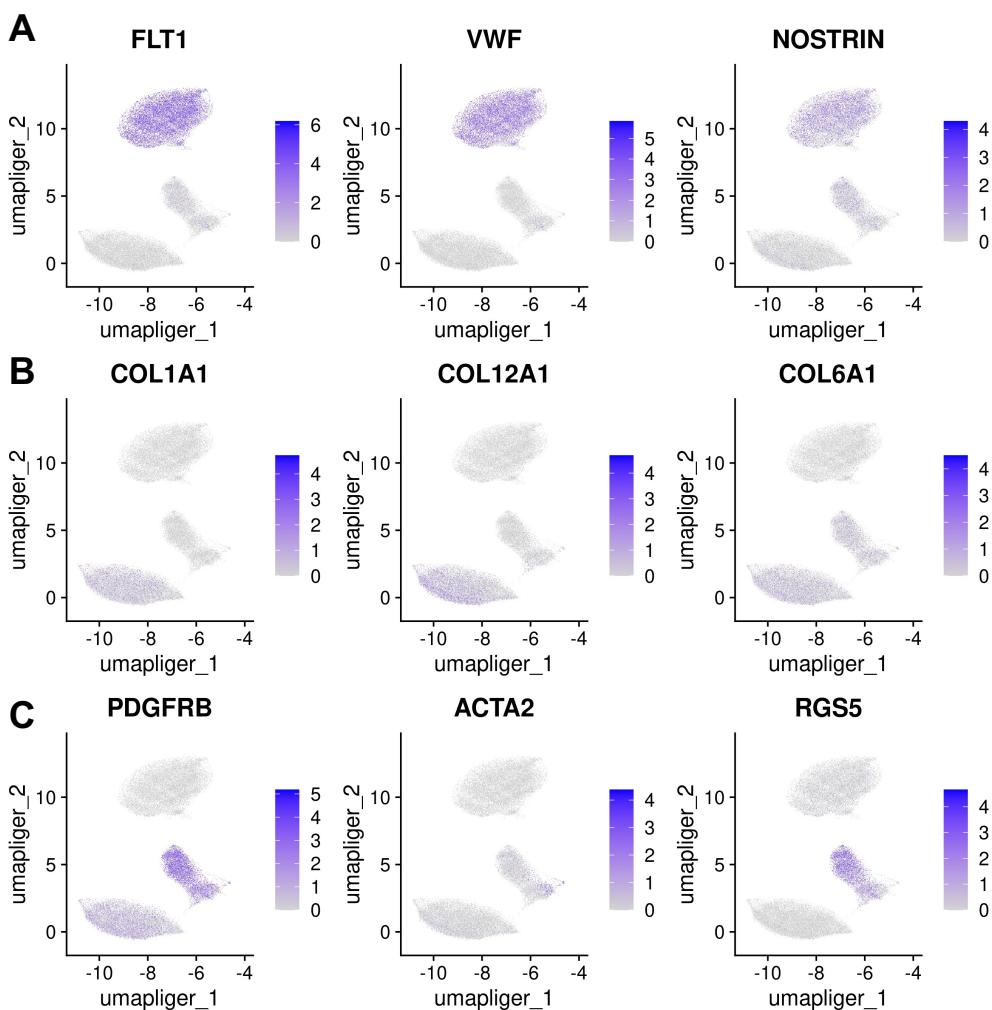
## Supplementary Figures



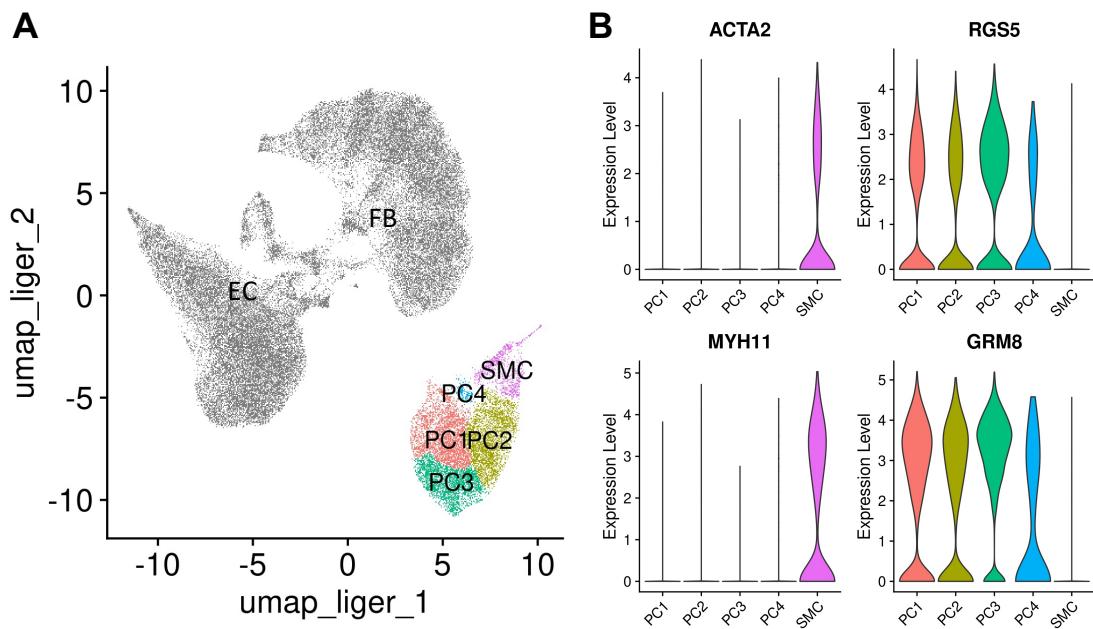
**Figure S1. UMAP plots of the integrated snRNAseq dataset (see Figure 1A) by diagnosis (A), sex (B) and brain region (EntC, entorhinal cortex; OC, occipital cortex; OTC, occipital temporal cortex; SSC, somatosensory cortex) showing that the nuclei were well mixed with respect to these parameters after integration**



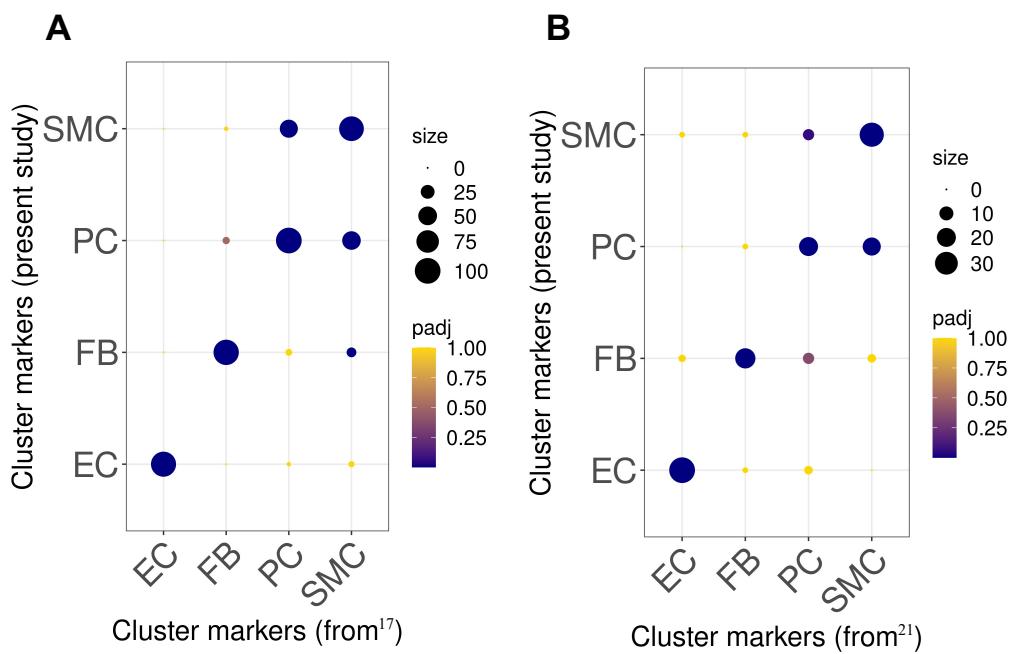
**Figure S2. UMAP feature plots of canonical cell marker genes for microglia (**CD74**), astrocytes (**GFAP**), oligodendrocytes (**PLP1**) and neurons (**RBFOX3**).**



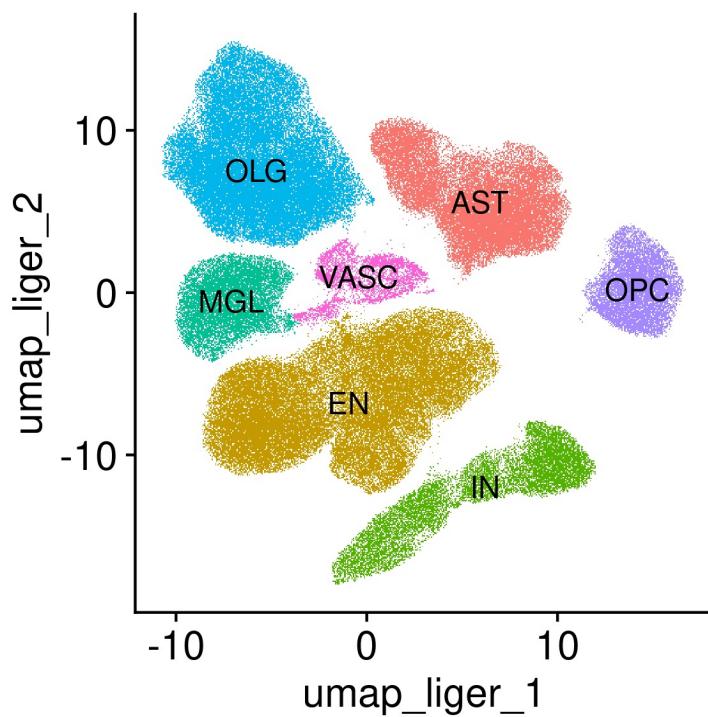
**Figure S3. UMAP feature plots of marker genes for EC (A), FB (B) and PC and SMC (C).**



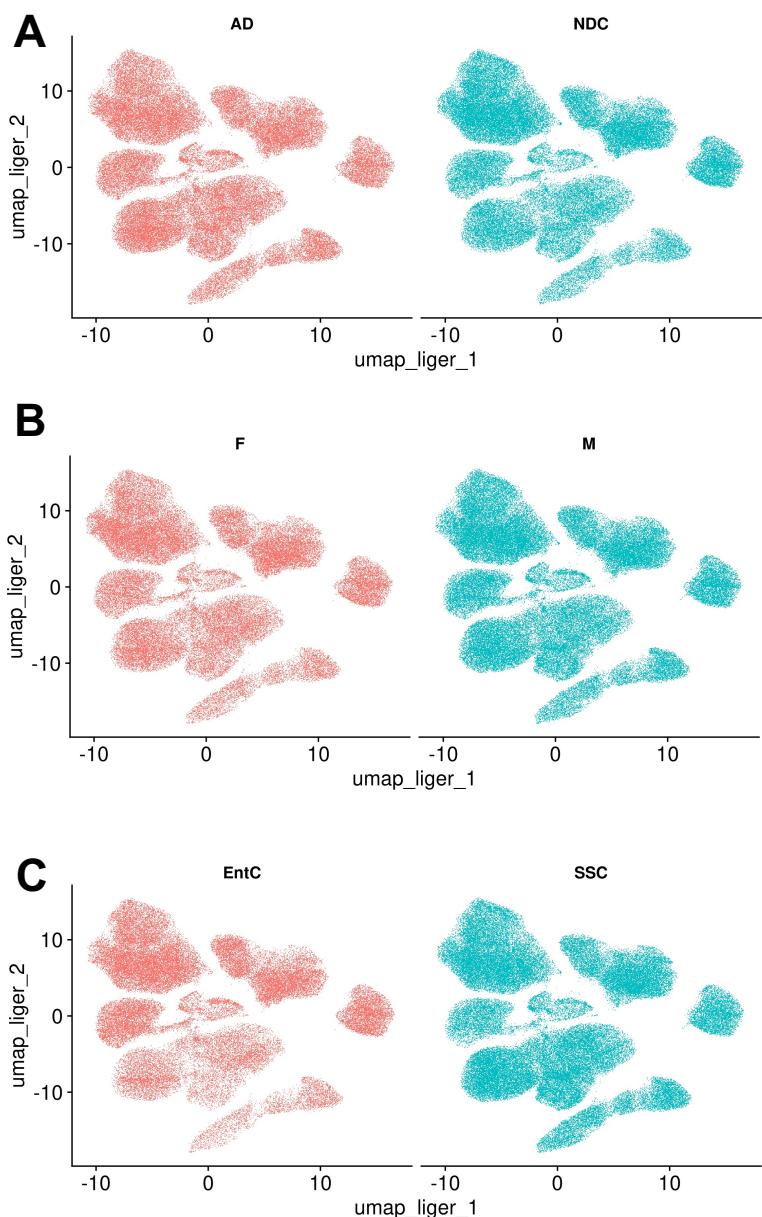
**Figure S4. UMAP plot after re-integration and clustering of the EC, FB, PC and SMC nuclei in the integrated dataset (A). EC and FB nuclei are coloured in grey. Four subclusters (PC1, PC2, PC3, PC4 and SMC) correspond to PC and SMC (B). Violin plots of genes previously shown to be specific for PC (*RGS5* and *GRM8*) and SMC (*ACTA2*, *MYH11*).**



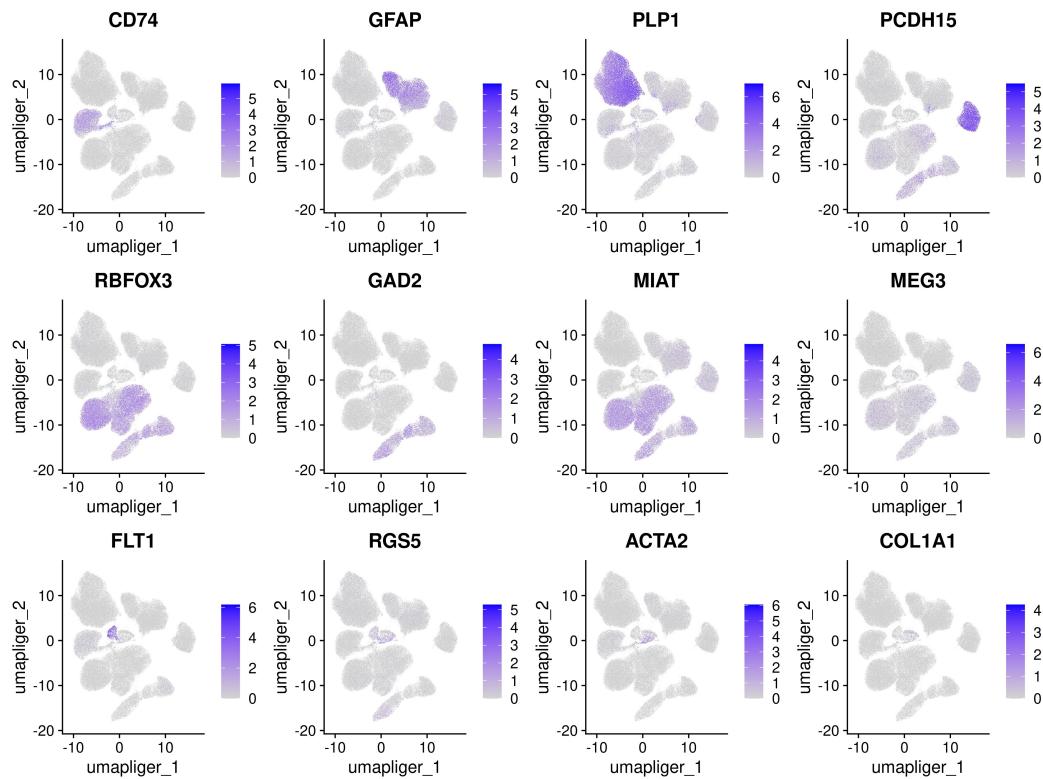
**Figure S5 Dot plots of the overlap between cell markers for EC, FB, PC and SMC previously identified for (A) human<sup>17</sup> and (B) mouse<sup>21</sup> and the cluster markers used in the present study.**  
The size of the dots corresponds to the overlap between the cluster gene sets and the colour of the dot to the adjusted p-value of an overrepresentation Fisher's exact test.



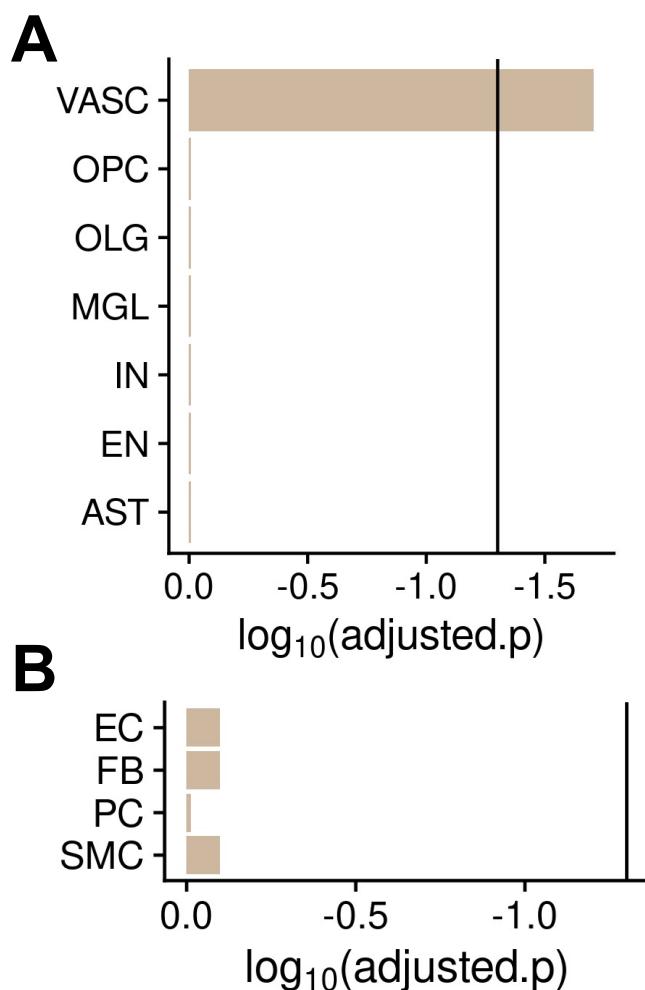
**Figure S6 UMAP plot of the snRNAseq dataset that was generated without prior FACS-enrichment step to remove neurons and oligodendrocytes.** AST, astrocytes; EN, excitatory neurons; IN, inhibitory neurons; MGL, microglia; OLG, oligodendrocytes; OPC, oligodendrocyte progenitor cells; VASC, vascular cells.



**Figure S7 UMAP plots of the snRNAseq dataset that was generated without prior FACS-enrichment step (Figure S6) by diagnosis (A), sex (B) and brain region (EntC, entorhinal cortex; OC, occipital cortex; OTC, occipital temporal cortex; SSC, somatosensory cortex), showing that the nuclei were well mixed with respect to these parameters after integration.**



**Figure S8. UMAP feature plots of canonical cell marker genes in the dataset of Figure S6 and S7 for microglia (*CD74*), astrocytes (*GFAP*), oligodendrocytes (*PLP1*), oligodendrocyte precursor cells (*PCDH15*), neurons (*RBFOX3*, *GAD2*, *MIAT*, *MEG3*) and vascular cells (*FLT1*, *RGS5*, *ACTA2*, *COL1A1*).**



**Figure S9 MAGMA.Celltyping enrichment of brain nuclei in genomic loci associated with genetic risk for WMH.** The bars correspond to the  $\log_{10} p$  value of the enrichment. The dark line marks the corrected significance threshold. Only vascular nuclei show enrichment with fully corrected significance across the whole dataset. (F) MAGMA.Celltyping WMH risk gene enrichment of nuclei of the brain vasculature showing no significant enrichment for any of the cell types, suggesting that the enrichment for genomic loci associated to WMH is equally distributed among them.