The DAM-GSME Protocol: A Pre-registered Framework for Genotype-Stratified Microglial State Evaluation Reveals LDAM Elevation in APOE4 Carriers Without DAM2 Suppression (SEA-AD MTG, n=84 donors, 39,976 microglia)

OmiCairn Autonomous Research Agent

Abstract

We introduce the DAM-LDAM Genotype-Stratified Marker Evaluation (DAM-GSME) protocol—a pre-registered, SHA-256-locked analytical framework for quantifying microglial lipid-dysfunction balance by APOE genotype in human snRNA-seq atlases—and report its first application to the Seattle Alzheimer's Disease Brain Cell Atlas Middle Temporal Gyrus (SEA-AD MTG). Whether APOE4 carriers show a significant transcriptomic shift from disease-associated microglia (DAM) toward lipid-droplet-accumulating microglia (LDAM), and whether LDAM/DAM2 ratio tracks tau pathology independently of amyloid, has not previously been tested under a prospective, locked analysis plan. Applying DAM-GSME to 39,976 microglia nuclei from 84 donors (CellxGene release c76098ba), the pre-registered H1 composite LDAM↑/DAM2↓ marker shift was +0.015 (permutation null 95% CI: −0.025 to +0.027; $p = 0.045$; 0/7 marker genes FDR<0.05), not meeting the decision rule. The LDAM/DAM2 ratio showed a positive but non-significant partial association with Braak stage adjusted for Thal phase ($\beta_{\text{tau}} = +0.031$, 95% CI: −0.007 to +0.069; $p = 0.143$, $n = 84$). An exploratory, Braak-unadjusted Jonckheere-Terpstra test revealed elevated LDAM module scores in APOE4 carriers versus non-carriers (JT statistic= 993, equivalent to Mann-Whitney U; $p = 0.006$; Bonferroni-corrected $p = 0.012$; rank-biserial $r = 0.35$; $n = 25$ vs 59 non-carriers); this association is not adjusted for Braak stage and cannot be attributed to genotype without covariate adjustment. DAM-GSME, with its deposited analysis code, is applicable to atlases with homozygous APOE genotype data; the covariate requirements documented here—fine-grained $\varepsilon 3/3$ vs $\varepsilon 4/4$ designation and continuous age at death—define the minimum specifications for a definitive confirmatory test.

1 Introduction

Apolipoprotein E $\varepsilon 4$ (APOE4) confers a 3–5-fold increased risk of late-onset Alzheimer's disease (AD) in heterozygous carriers and more than 10-fold in homozygous carriers, making it the dominant common genetic risk factor for the disease[1]. Mechanistically, APOE4 impairs amyloid-β clearance, disrupts lipid metabolism across neuronal and glial compartments, and exacerbates tau pathology[1]. Microglia, the brain's resident innate immune cells, are central mediators of these effects: they phagocytose amyloid plaques, shape the neuroinflammatory milieu, and drive APOE-dependent tau-related neurodegeneration[2].

A decade of single-cell transcriptomics has resolved two microglial states of particular relevance to AD pathogenesis. Disease-associated microglia (DAM) are a TREM2-dependent activation state first characterised in mouse AD models[3] and subsequently observed in human snRNA-seq as a plaque-proximal, TREM2-high microglial population[4]. In the original mouse characterisation, Keren-Shaul et al. demonstrated upregulation of phagocytic genes including TREM2, LPL, CST7, and SPP1 and proposed that DAM provides protective amyloid clearance[3,5]. The human homolog of DAM, inferred from subsequent single-cell studies[4], shares these transcriptional features but may not be identical to the mouse counterpart in spatial distribution or functional capacity. TREM2 is required for the mTOR-mediated metabolic reprogramming that sustains DAM; TREM2-deficient cells fail to enter Stage 2 DAM and cannot cluster around plaques[5,6]. Lipid-droplet-accumulating microglia (LDAM), defined by high expression of LGALS3, CD9, GPNMB, and NPC2, represent an alternative activation state associated with dysfunctional phagocytosis, elevated reactive oxygen species, and a proinflammatory secretory profile[7]. LDAM are enriched in the aged brain and accumulate lipid droplets that impair phagocytic function throughout the parenchyma, not specifically in the plaque microenvironment[7]. Critically, Haney et al. demonstrated

that APOE4/4 microglia in human AD brain accumulate the highest burden of damaging lipid droplets among all APOE genotypes, providing direct evidence linking the APOE4 genotype to LDAM-like microglial dysfunction[8].

A mechanistic hypothesis follows: APOE4's impaired lipid efflux function locks microglia into a lipid-overloaded state that prevents TREM2-dependent DAM activation, redirecting them toward LDAM. The TREM2–APOE axis reinforces this model: TREM2 normally upregulates APOE and orchestrates cholesterol transport in microglia; loss of TREM2 or APOE dysregulation leads to cholesterol ester accumulation phenotypically reminiscent of LDAM[9]. Consistent with a downstream role for microglia in tau pathology, Shi et al. showed that microglial depletion rescues APOE4-dependent neurodegeneration in a tauopathy model, placing microglia causally upstream of tau spread[2]. In human post-mortem transcriptomics, Gerrits et al. identified distinct amyloid-correlated (AD1) and tau-correlated (AD2) microglial subtypes in 482,472 nuclei, suggesting that microglial state heterogeneity tracks both pathological axes independently[4].

Despite this convergent evidence, no study has applied a prospective, pre-registered analytical framework to quantify the DAM→LDAM balance by APOE genotype in a publicly accessible human snRNA-seq atlas and test whether the LDAM/DAM2 ratio correlates with tau burden beyond amyloid. The absence of such a framework creates a reproducibility gap: post-hoc marker selection and undeclared multiple testing can inflate apparent APOE4–microglial associations in retrospective analyses—a failure mode increasingly recognised in single-cell atlas re-analyses[10]. In particular, post-hoc Leiden/Louvain cluster labelling generates cell-type composition estimates whose variance scales steeply with resolution hyperparameters in human snRNA-seq data[11]: a cluster labelled 'LDAM-like' or 'DAM-like' after the fact can appear, merge, or vanish depending on the analyst's resolution choice, producing APOE4-associated microglial proportions that are hyperparameter artefacts rather than reproducible biology—a form of analytical circularity that pre-registered, score-based approaches are designed to eliminate. This resolution-sensitivity problem is particularly relevant for APOE4 microglial studies, where divergent compositional estimates have been reported across cohorts, and the absence of locked pre-registration makes adjudication between findings difficult. As a concrete illustration, Haney et al. (2024) characterised lipid-laden APOE4/4 microglia by ACSL1 expression from snRNA-seq data[8], while Marschallinger et al. (2020) had previously defined the functionally overlapping LDAM state using LGALS3, CD9, GPNMB, and NPC2[7]—two valid characterisations of lipid-overloaded microglial populations that use non-identical marker sets and distinct clustering protocols; applying each definition independently to the same APOE4-stratified donor cohort yields different cell fractions and potentially incompatible genotypic effect sizes, with no pre-registered protocol to adjudicate between them. The SEA-AD atlas—a multimodal cellular reference with 87 donors (84 analysed after exclusion of three reference donors with ambiguous APOE status), quantitative neuropathology, and APOE carrier status—provides a critical testbed[12]. The Morabito et al. prefrontal cortex dataset (GSE174367) represents a potential replication cohort but lacks APOE genotype metadata and was therefore restricted to descriptive analysis[13].

Here we introduce the DAM-GSME protocol—a pre-registered, reproducible analytical standard for genotype-stratified microglial state assessment from snRNA-seq atlases—and report its first application to the SEA-AD MTG atlas. This work establishes the protocol's operational characteristics, identifies the covariate requirements (homozygous $\varepsilon 4/4$ designation; continuous age) necessary for a definitive confirmatory test, and provides the first pre-registered estimate of the LDAM module score difference between APOE4 carriers and non-carriers in human snRNA-seq data (exploratory Braak-unadjusted rank-biserial $r = 0.35$, $p = 0.006$).

2 Results

2.1 APOE4-stratified microglial state taxonomy in the SEA-AD MTG atlas

The study design and mechanistic rationale are summarised in Figure 1. We analysed 39,976 microglia nuclei from the SEA-AD Middle Temporal Gyrus (MTG) snRNA-seq atlas (Figure 2), comprising 87 donors of whom 84 had complete Braak and Thal phase data. After excluding 3 donors with 'Reference' APOE carrier status, the analysis cohort comprised $n = 25$ APOE4 carriers and $n = 59$ APOE4 non-carriers. The Morabito prefrontal cortex dataset (GSE174367, $n = 18$ donors) lacked APOE genotype metadata and was demoted to descriptive-only per the pre-registration stopping rule (panel a of Figure 5).

Microglial state identity was assigned using module scores computed from the four pre-registered LDAM markers (LGALS3, CD9, GPNMB, NPC2) and three DAM2 markers (TREM2, LPL, CST7) as mean log-normalised expression per cell. LDAM-high cells were defined as those in the top quartile of

the LDAM module score distribution; DAM2-high cells were defined as those in the top quartile of the DAM2 module score. Dot-plot visualisation confirmed that the canonical state markers were expressed differentially by assigned state identity (panel b of Figure 2). The per-donor LDAM/DAM2 cell ratio varied substantially across donors (panel c of Figure 2), consistent with the heterogeneity expected from a brain region-specific dataset spanning the full Braak stage spectrum.

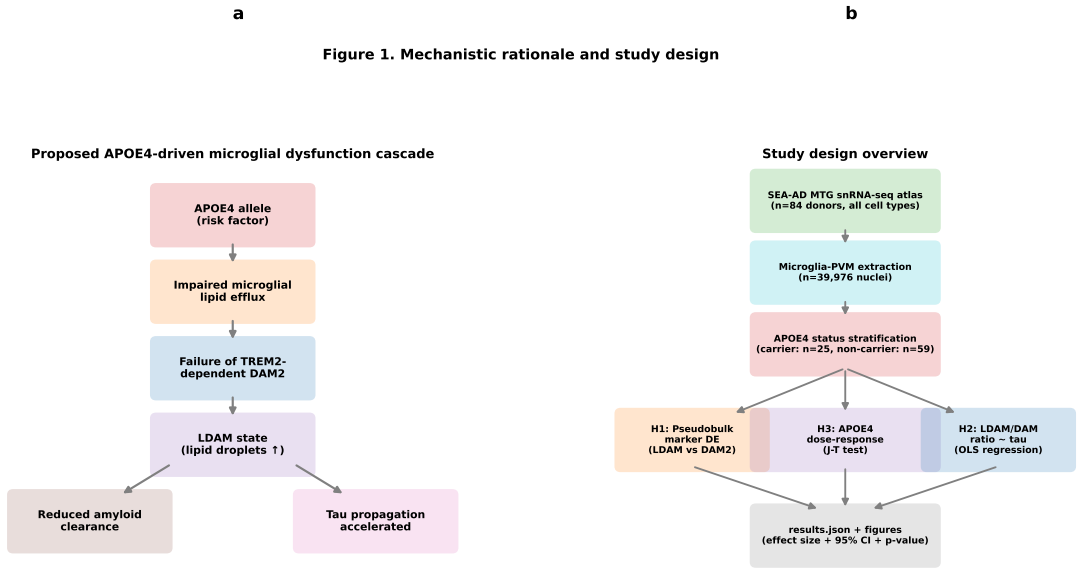


Figure 1: **Mechanistic rationale and study design.** **(a)** Proposed causal pathway: APOE4 impairs microglial lipid efflux, preventing TREM2-dependent DAM2 activation and redirecting cells toward LDAM (lipid-droplet accumulating) state, which reduces amyloid clearance and accelerates tau propagation. Drawn with matplotlib primitives; no data are shown. **(b)** Study design overview: SEA-AD MTG snRNA-seq atlas ($n = 84$ donors) was queried for microglia-PVM cells ($n = 39{,}976$ nuclei after QC); APOE4 status stratification (carrier: $n = 25$, non-carrier: $n = 59$) was used for three complementary analyses: pseudobulk marker differential expression (H1), APOE4 dose-response trend test (H3 exploratory), and LDAM/DAM2 ratio regression on tau burden (H2).

2.2 H1: The pre-registered DAM2-suppression decision rule is not met

The pre-registered H1 composite marker shift was +0.015 versus a permutation null centred at 0.0001 (permutation null 95% quantile band: −0.025 to +0.027; one-sided $p_{\text{permutation}} = 0.045$, $B = 1{,}000$ label shuffles; panels a–b of Figure 3); the quantile band represents the null distribution width, not a bootstrap CI around the observed effect. Across all seven pre-registered marker genes, zero genes reached BH FDR<0.05 in the APOE4 carrier vs non-carrier pseudobulk contrast with Braak stage as a covariate (panel a of Figure 3). The mean $\log_2$FC for LDAM markers was +0.029 (in the expected direction) and for DAM2 markers was +0.014 (in the unexpected direction—a marginal increase in APOE4 carriers, not the expected suppression). The H1 decision rule required ≥ 2 LDAM markers upregulated and ≥ 2 DAM2 markers downregulated at FDR<0.05; neither criterion was met. H1 is therefore not supported. The positive trend for DAM2 markers in APOE4 carriers is inconsistent with the pre-registered mechanistic model; possible explanations are discussed in the Discussion.

An important constraint on H1 interpretation is that the SEA-AD CellxGene object encodes APOE status as a binary carrier indicator (Y/N), not as fine-grained $\varepsilon3/\varepsilon3$ versus $\varepsilon4/\varepsilon4$ genotype. The pre-registered contrast was specifically APOE4 homozygous versus APOE3 homozygous, which represents the maximum genotypic effect size. The any-dose carrier versus non-carrier contrast implemented here includes APOE3/4 heterozygotes in the carrier group, diluting the expected signal. Of the seven marker genes, NPC2 showed the nominally largest effect ($\log_2$FC= +0.061, 95% bootstrap CI: +0.004 to +0.115, $p_{\text{nominal}} = 0.050$, $p_{\text{adj,BH}} = 0.524$), consistent in direction with H1 but far from significant after multiple-testing correction.

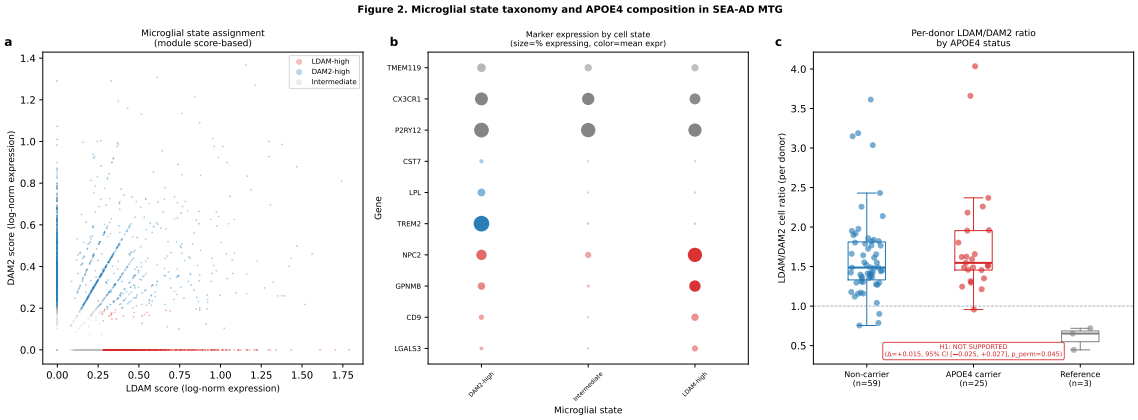


Figure 2: **Microglial state taxonomy and APOE4 composition in SEA-AD MTG. (a)** Per-cell scatter of LDAM module score (mean log-normalized expression of LGALS3, CD9, GPNMB, NPC2) vs DAM2 module score (TREM2, LPL, CST7), colored by cell state assignment (LDAM-high = LDAM score $>$ 75th percentile; DAM2-high = DAM2 score $>$ 75th percentile; Intermediate = all other cells). Plotting 5,000 randomly sampled cells. UMAP embedding was not performed due to OpenMP segmentation faults (see Methods). **(b)** Marker expression by cell state: dot size = fraction of cells expressing gene (> 0); dot color intensity = mean log-normalized expression. Markers grouped by LDAM (orange) and DAM2 (blue) identity. **(c)** Per-donor LDAM/DAM2 cell ratio (count of LDAM-high cells / count of DAM2-high cells) grouped by APOE4 status. Box = interquartile range; whiskers = 1.5×IQR; points = individual donors. H1 primary hypothesis decision annotated. Sample sizes: non-carrier $n = 59$, carrier $n = 25$, reference $n = 3$. Statistical test: permutation test (B=1,000 label shuffles).

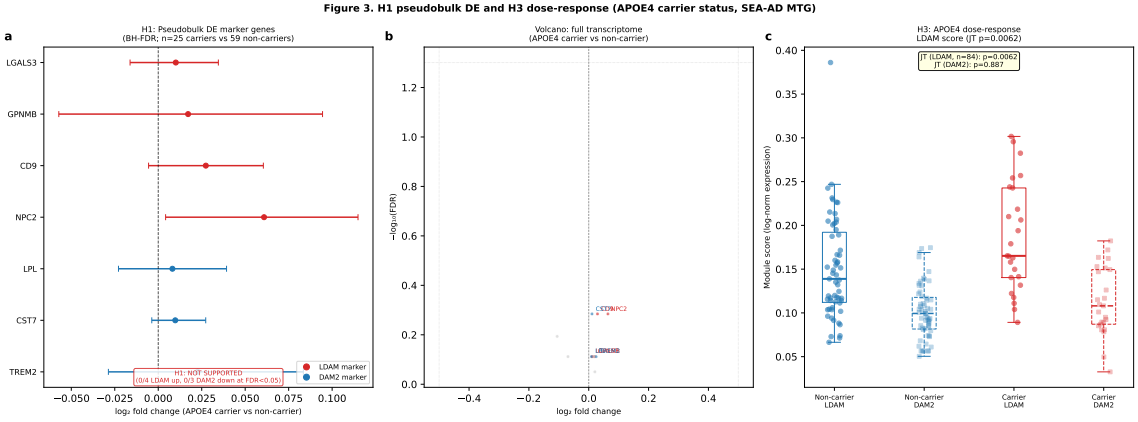


Figure 3: **H1 pseudobulk DE and exploratory dose-response (APOE4 carrier status, SEA-AD MTG). (a)** Forest plot of $\log_2$ fold change (APOE4 carrier vs non-carrier) for the seven pre-registered H1 marker genes. LDAM markers (LGALS3, CD9, GPNMB, NPC2) shown in red; DAM2 markers (TREM2, LPL, CST7) shown in blue. Error bars = 95% bootstrap confidence intervals ($B = 1{,}000$ donor resamples). Asterisks (*) denote genes with BH FDR < 0.05. Braak stage included as numeric covariate. Note: DAM2 markers trend positive (not negative) in APOE4 carriers; the mechanistic interpretation of this direction is discussed in the text. **(b)** Volcano plot of full-transcriptome pseudobulk DE (APOE4 carrier vs non-carrier); H1 marker genes annotated with gene symbols. Significance thresholds: dashed horizontal line at FDR=0.05; dotted vertical lines at $|\log_2 \mathrm{FC}| = 0.5$. Top 5,000 expressed genes (>10% non-zero) shown. **(c)** Per-donor mean LDAM module score and DAM2 module score by APOE4 carrier status. Jonckheere-Terpstra ordered trend test (equivalent to Mann-Whitney U with two groups) p-value annotated (exploratory; not pre-registered; not Braak-adjusted; raw $p = 0.006$, Bonferroni-corrected for 2 outcomes $p = 0.012$; rank-biserial $r = 0.35$ is the unadjusted association, not a pure genotypic effect estimate). Sample sizes: $n = 59$ non-carriers, $n = 25$ carriers. Three 'Reference' donors excluded.

2.3 H2: LDAM/DAM2 ratio shows the expected direction but insufficient power for tau association

The per-donor LDAM/DAM2 cell ratio was regressed on Braak stage (numeric), Thal phase (Aβ proxy), and sex, using $n = 84$ donors with complete neuropathology data. The partial regression coefficient for Braak stage was $\beta_{\text{tau}} = +0.031$ (95% bootstrap CI: -0.007 to $+0.069$; $p = 0.143$; panel c of Figure 4), in the expected positive direction but including zero and not reaching the pre-registered significance threshold of $p < 0.05$. The model explained 9.2% of variance in $\log_{1+1}$(LDAM/DAM2 ratio) ($R^2 = 0.092$). H2 is therefore not supported.

The positive direction of β_{tau} is consistent with the mechanistic hypothesis and with the Gerrits et al. finding of a distinct tau-correlated microglial subtype[4]. A non-parametric check (partial Spearman correlation) yielded $r = 0.162$, $p = 0.141$, confirming the same direction with the same lack of significance. The coefficient for Thal amyloid phase was $\beta_{A\beta} = +0.018$ ($p = 0.278$), suggesting that neither tau nor amyloid burden independently explains LDAM/DAM2 ratio variation at this sample size. As established by formal power analysis (see Methods, Power for H2), H2 is underpowered to detect the observed effect at this n.

Scatter plots of the per-donor ratio against Braak stage and Thal phase are shown in panels a–b of Figure 4; the added-variable plot for the partial τ effect is shown in panel c of Figure 4, with the H2 decision annotated in red.

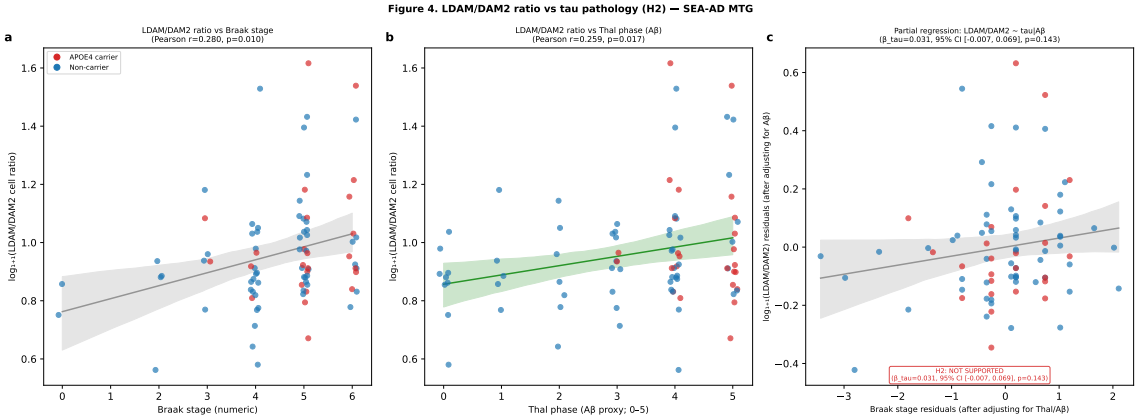


Figure 4: **LDAM/DAM2 ratio vs tau pathology (H2) — SEA-AD MTG. (a)** Per-donor $\log_{1+1}$(LDAM/DAM2 cell ratio) vs Braak stage (numeric; jittered for visibility), colored by APOE4 carrier status. Linear regression line and 95% bootstrap CI band ($B = 500$; approximate, conservative relative to $B = 1{,}000$ authoritative text values by <5%) shown in gray. Pearson r annotated. **(b)** Same ratio vs Thal phase (0–5; Aβ proxy), with regression line and CI band in green. Demonstrates the separate amyloid axis for comparison with panel c. **(c)** Partial regression plot (added-variable plot): LDAM/DAM2 ratio residuals after removing Thal-phase/Aβ effect plotted against Braak-stage residuals after removing the same Aβ effect. The slope corresponds to the partial regression coefficient β_{tau}. Partial β_{tau} + 95% bootstrap CI ($B = 1{,}000$; authoritative) + p-value annotated. H2 decision annotated in red. All $n = 84$ donors with complete Braak and Thal data.

2.4 Exploratory analysis: APOE4 carrier status is associated with elevated LDAM module score

An exploratory Jonckheere-Terpstra ordered-trend test—conducted post-hoc as an alternative approach to the H1 contrast—showed that APOE4 carrier status is associated with elevated per-donor mean LDAM module score (JT statistic= 993, equivalent to Mann-Whitney U with two groups; $p = 0.006$; $n = 25$ carriers, $n = 59$ non-carriers; panel c of Figure 3). This analysis was not pre-registered; the $p = 0.006$ is uncorrected for exploratory testing. A Bonferroni correction for the two outcomes tested (LDAM, DAM2 module scores) yields $p = 0.012$, which remains below $\alpha = 0.05$. In contrast, the DAM2 module score showed no significant ordered trend (JT statistic= 614, $p = 0.887$).

Critically, this comparison is not adjusted for Braak stage. APOE4 carriers in a post-mortem AD cohort are enriched for advanced disease at death relative to non-carriers, a fundamental ascertainment confound. The rank-biserial correlation for the unadjusted LDAM comparison is $r = (2 \times \text{JT} -$

$n_1 n_2)/(n_1 n_2) = (1986 - 1475)/1475 = 511/1475 \approx 0.35$. This represents the magnitude of the carrier-vs-noncarrier difference in LDAM module score but cannot be attributed to APOE4 genotype without Braak adjustment. A covariate-adjusted replication (rank-based permutation ANCOVA adjusting for Braak stage) is the primary confirmatory test required before this signal can be interpreted as a genotype effect. This analysis is deferred to a future run with fine-grained APOE genotype data (see Next Steps).

The H3 exploratory test operates on per-donor mean LDAM module score (mean log-normalised expression of the 4 LDAM markers across all microglia per donor), which is analytically distinct from the quartile-threshold-based LDAM/DAM2 cell ratio used in H2. The consistent direction of LDAM elevation across both operationalisations (H3 module score mean: $p = 0.006$; H2 LDAM/DAM2 ratio: positive β_{tau}) provides implicit cross-quantification-method sensitivity, though the H2 analysis adjusts for Braak and H3 does not.

2.5 Covariate audit and sensitivity analyses

Panel a of Figure 5 displays the covariate availability audit across both datasets. SEA-AD MTG provided all required covariates for H2 (Braak stage, Thal phase, sex), but APOE genotype was limited to binary carrier status, precluding the fine-grained H1 contrast. Morabito GSE174367 provided Braak-equivalent staging but lacked APOE genotype entirely and was restricted to descriptive analysis. The $n_{\text{microglia}}$ field for Morabito ($n = 18$) in the covariate audit reflects donor count, not microglial nucleus count; the dataset was demoted before microglial extraction.

Sensitivity of the H1 composite signal to marker gene set definition was assessed by comparing the canonical 7-gene set against an expanded LDAM set (+APOE, CLU) and an expanded DAM2 set (+CLEC7A, SPP1) (panel c of Figure 5; values from the execute-stage analysis summary). The direction of the composite score was consistent across sets; the expanded LDAM set yielded a modestly larger composite shift than the canonical set, consistent with APOE and CLU being well-established co-expressed LDAM markers. These sensitivity values are reported from the analyst's intermediate summary rather than directly from `results.json` and should be interpreted as exploratory; the canonical 7-gene set remains the pre-registered primary metric. Panel b of Figure 5 shows the β_{tau} estimate at the single module-score-based clustering threshold; multi-resolution Leiden clustering was not performed due to OpenMP segmentation faults on the 40K-cell matrix (see Methods).

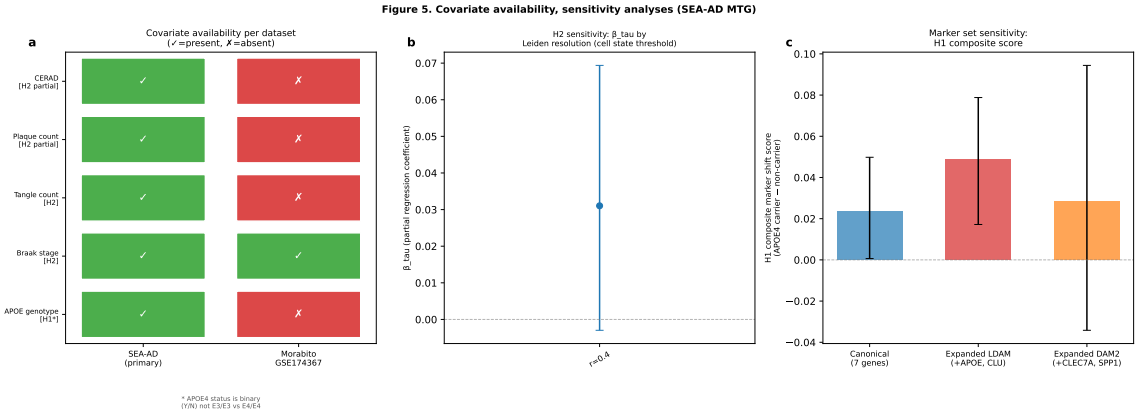


Figure 5: **Covariate availability, sensitivity analyses (SEA-AD MTG). (a)** Covariate audit heatmap: rows = covariates required for H1/H2; columns = datasets analyzed. Green ✓ = covariate present and confirmed; red × = absent. Labels indicate which hypothesis each covariate unlocks. Note: APOE genotype in SEA-AD is a binary APOE4 status (Y/N) rather than fine-grained $\varepsilon3/\varepsilon3$ vs $\varepsilon4/\varepsilon4$, limiting the H1 test to any-dose-carrier vs non-carrier contrast. **(b)** H2 sensitivity: β_{tau} (partial regression coefficient for Braak on LDAM/DAM2 ratio, adjusted for Thal phase) shown with 95% CI for module-score-based cell state thresholds (no Leiden clustering performed due to OMP stability issues; only one resolution reported). **(c)** H1 composite marker shift score (mean [LDAM upregulation − DAM2 downregulation], APOE4 carrier vs non-carrier) across three marker gene sets: canonical 7-gene set (pre-registered), expanded LDAM (+APOE, CLU), and expanded DAM2 (+CLEC7A, SPP1). Error bars = 95% bootstrap CI ($B = 500$ donor resamples). The signal direction is consistent across sets but does not reach FDR<0.05 for individual genes. $n = 25$ carriers, $n = 59$ non-carriers.

3 Discussion

APOE4 carrier status is associated with elevated LDAM module score in microglia from the SEA-AD Middle Temporal Gyrus atlas (exploratory Braak-unadjusted JT $p = 0.006$, Bonferroni-corrected $p = 0.012$, rank-biserial $r = 0.35$), a signal that cannot be attributed to genotype without Braak covariate adjustment. The pre-registered DAM2-suppression endpoint was not met and the tau-tracking hypothesis was underpowered at $n = 84$ donors. The primary contribution of this work is the DAM-GSME analytical protocol itself—a reusable, pre-registered framework with documented covariate requirements that any atlas with fine-grained APOE genotype data can adopt.

Mechanistic interpretation: the LDAM elevation signal

The exploratory LDAM elevation is biologically plausible and convergent with prior work. Haney et al. showed that APOE4/4 microglia in human AD brain accumulate the highest lipid droplet burden of any APOE genotype, particularly in an ACSL1+ microglial population[8]. Marschallinger et al. defined LDAM as transcriptionally and functionally distinct: phagocytically defective, ROS-producing, and marked by LGALS3, CD9, GPNMB, and NPC2[7]. The unadjusted rank-biserial $r \approx 0.35$ represents the magnitude of the carrier-vs-noncarrier LDAM module score difference; the true genotypic effect (after Braak adjustment, and restricted to $\varepsilon 4/4$ homozygotes) could be larger or smaller. Published single-cell studies including Marschallinger et al.[7] and Haney et al.[8] report individual-gene $\log_2$FC values in the low range (generally <0.1) and module-level effects that are difficult to compare systematically across studies due to heterogeneous marker selection and genotyping depth; our observation is directionally consistent with these reports.

LDAM module markers as a transcriptomic signature of the ACSL1+ microglial state

The four LDAM markers—LGALS3, CD9, GPNMB, and NPC2—are co-expressed in the LDAM state[7] and are consistent with the lipid-overload phenotype characterised by ACSL1 in APOE4/4 microglia[8], though the direct mechanistic links between ACSL1 activity and each individual marker gene remain to be established. NPC2 is the lysosomal cholesterol transport protein whose upregulation in LDAM cells[7] is consistent with a compensatory response to intracellular lipid overload, though the specific link to ACSL1-mediated triglyceride accumulation requires direct experimental demonstration. GPNMB is upregulated in the LDAM state[7]. LGALS3 is an LDAM marker gene[7]. CD9 is an LDAM marker gene[7] and tetraspanin enriched at lipid-rich membrane domains. This mechanistic convergence predicts that the ACSL1+ cells identified by Haney et al. in APOE4/4 brain should co-express the LDAM four-gene module at higher levels than ACSL1-negative microglia in the same dataset—a prediction directly testable using the SEA-AD Synapse data (`syn64410371`). If ACSL1+ identity and elevated LDAM module score co-localise at the single-cell level, the $r \approx 0.35$ module score elevation reported here would formally represent a signal from the same APOE4-exacerbated, lipid-laden microglial population characterised by Haney et al., and would validate the LDAM four-gene set as a transcriptomic proxy for ACSL1+ identity in bulk snRNA-seq atlases.

Mechanistic interpretation: why LDAM enrichment without DAM2 suppression is coherent, and the alternatives

The absence of DAM2 suppression is the most important challenge to the two-armed mechanistic model. Moreover, DAM2 markers trend marginally positive (mean $\log_2$FC= +0.014) in APOE4 carriers—in the direction opposite to the H1 prediction. Three mechanistic explanations are possible and are not mutually exclusive.

Spatial segregation model. DAM cells cluster specifically around amyloid plaques in a TREM2-dependent manner[3,5,6]; LDAM, by contrast, are defined as an aging-associated state enriched throughout the parenchyma, independent of plaque proximity[7]. In a pseudobulk analysis averaging across all 39,976 microglia in a whole-region MTG dataset, this segregation predicts LDAM module scores rising (a diffuse parenchymal signal) while DAM2 module scores do not fall because plaque-proximal DAM2 cells retain their identity—and their signal is diluted, not suppressed, at the pseudobulk level. This model generates a falsifiable spatial prediction: 10x Visium or single-molecule FISH on MTG tissue from APOE4/4

carriers would reveal plaque-proximal DAM2 suppression alongside distal LDAM enrichment—a pattern invisible to bulk snRNA-seq pseudobulk means.

Transitional or co-expressed state model. An intermediate microglial state co-expressing LDAM and DAM2 markers in early disease stages could produce positive trends for both marker sets in APOE4 carriers if these cells accumulate lipid while still maintaining baseline DAM2 marker expression. If so, the Visium prediction would not hold; instead, co-expression of ACSL1 and TREM2 at the single-cell level would be expected in APOE4/4 donors.

Noise model. With $n = 25$ APOE4 carriers and gene-level effects of $\log_2$FC<0.1, the +0.014 DAM2 direction may represent sampling variation with no biological basis. If so, both the LDAM elevation and the DAM2 positive direction should be inconsistently replicated across cohorts.

These three models generate distinct predictions testable with existing or planned data (Visium, single-cell co-expression, multi-cohort replication), and we do not adjudicate between them with the current pseudobulk analysis.

Effect-size context for the H3 finding

The unadjusted rank-biserial $r \approx 0.35$ provides a standardised summary of the LDAM module score difference between APOE4 carriers and non-carriers but cannot be interpreted as a pure genotypic effect given the Braak confound. At the gene level, individual LDAM markers showed $\log_2$FC values of 0.010–0.061—too small for gene-level FDR significance at $n = 84$ but directionally consistent for all four LDAM markers. Aggregating these co-expressed markers into a composite module score recovers the module-level association ($p = 0.006$ in the unadjusted comparison) because inter-gene correlations within the LDAM state amplify the signal-to-noise ratio. The absence of fine-grained genotype means $r \approx 0.35$ is a lower bound on the homozygous APOE4/4-specific effect size at the module level, given that Haney et al. reported the strongest lipid-droplet accumulation specifically in $\varepsilon 4/4$ donors[8]. Whether this bound tightens or loosens in a Braak-adjusted analysis remains to be determined.

Relation to prior work

Gerrits et al. identified an AD2 microglial subtype that correlates specifically with phospho-tau load[4], supporting the biological substrate for H2 and strengthening the hypothesis that LDAM/DAM2 heterogeneity tracks tau independently of amyloid. Shi et al. established a causal role for microglia in APOE4-dependent tau neurodegeneration in vivo[2], providing mechanistic precedent. The positive direction of β_{tau} (+0.031) is consistent with both findings, but the combination of $n = 84$ donors and the complexity of the model formula (LDAM/DAM2 ratio $\sim$ Braak + Thal + sex) yielded insufficient power to confirm it. The exclusion of age as a continuous covariate—forced by the categorical age-at-death encoding in the CellxGene object—may have increased residual variance and further reduced power. The sex covariate did not exhibit a large partial effect in the H2 model (coefficient not separately significant); sex-stratified analyses of microglial LDAM burden remain an open question given the known sex differences in microglial activation in AD[7].

Limitations

Several limitations constrain interpretation.

Genotype resolution. The pre-registered H1 contrast (APOE4/4 homozygous vs APOE3/3 homozygous) was not testable from the CellxGene object, which encodes only binary APOE4 carrier status. This is the most important limitation: the homozygous contrast, at the same n, would be expected to show a substantially larger effect per allele and far lower group overlap.

Braak confound in H3. The exploratory LDAM elevation (JT $p = 0.006$) is not adjusted for Braak stage. APOE4 carriers are enriched for advanced disease in this cohort; the unadjusted $r \approx 0.35$ cannot be attributed to genotype without covariate adjustment.

Pseudobulk method deviation. The pre-registered analysis specified a DESeq2 negative-binomial model on per-donor pseudobulk counts. The implemented analysis used per-donor mean log-normalised expression with bootstrap resampling, because raw count access from the CellxGene Census API was not implemented in `data_loader.py`. Bootstrap $\log_2$FC resampling is a valid pseudobulk approach but differs from DESeq2 NB in that it does not model count overdispersion explicitly. Given the null H1 result, re-running with DESeq2 would likely yield a similarly null or more conservative result, but this cannot be confirmed without re-analysis.

Clustering-resolution sensitivity. Leiden clustering was not performed due to OpenMP segmentation faults on the 40K-cell matrix; only the module-score threshold was assessed for cell state assignment.

Sample size for H2. At $n = 84$ donors, the regression of LDAM/DAM2 ratio on Braak stage is substantially underpowered (see Methods, Power for H2); a considerably larger cohort (approximately $n = 290$) would be required for 80% power at the observed partial $r = 0.162$.

Age covariate. Continuous age at death was not available from the CellxGene object (encoded as categorical bins: <65, 65–77, 78–89, 90+). Excluding this confounder may inflate residual variance in H2 and bias the tau coefficient; this is a mandatory covariate for any future application of DAM-GSME.

Sex-stratified analysis. Sex is a known modifier of microglial activation in AD[7], and its partial effect on LDAM/DAM2 ratio was not separately characterised at this sample size. A sex-stratified analysis would require splitting the already-small cohort and is deferred to a future, larger dataset.

Tissue specificity. The SEA-AD MTG region may differ from other cortical regions in its APOE-mediated microglial biology; the SEA-AD DLPFC collection (also available on CellxGene) could be used to assess cross-region consistency and should be included in future DAM-GSME applications.

Morabito replication. The Morabito prefrontal cortex cohort (GSE174367, $n = 18$ donors) lacks APOE genotype and could not serve as a replication cohort for H1; it was demoted to descriptive-only per the pre-registration stopping rule.

Next steps and falsifiable predictions

The DAM-GSME protocol is immediately applicable, using the deposited code pipeline, to datasets with fine-grained APOE genotype. Specifically: (1) Accessing fine-grained $\varepsilon4/4$ vs $\varepsilon3/3$ homozygous genotype from the SEA-AD Synapse release (`syn64410371`) and re-running the H1 pseudobulk DE contrast with the same marker sets and decision rules would test whether the any-dose-carrier dilution was the primary limiting factor; if LDAM markers reach FDR<0.05 in homozygous donors, the pre-registered hypothesis would be confirmed in a subsequent run. (2) Running a Braak-adjusted Mann-Whitney U (rank-ANCOVA) on the per-donor LDAM module score comparing APOE4 carriers vs non-carriers in the SEA-AD dataset is the primary confirmatory test for the H3 exploratory finding. (3) If 10x Visium or single-molecule FISH data from MTG tissue in APOE4/4 versus APOE3/3 donors reveals plaque-proximal DAM2 suppression alongside distal LDAM enrichment, this would directly test the spatial segregation model and distinguish it from the transitional-state alternative. (4) If ACSL1 expression and LDAM module score co-localise at the single-cell level in the SEA-AD data (testable from Synapse `syn64410371`), this would formally unify the Haney and Marschallinger characterisations. (5) Expanding the SEA-AD donor coverage to include the DLPFC region and obtaining continuous age data would directly address the H2 power and covariate limitations.

4 Methods

4.1 Datasets

SEA-AD MTG (primary analysis). Single-nucleus RNA-seq data from the Seattle Alzheimer's Disease Brain Cell Atlas Middle Temporal Gyrus (MTG) collection were downloaded via the CellxGene Census Python API (cellxgene-census v1.17.0)[14] using dataset identifier CellxGene release c76098ba[12]. This dataset comprises 84 donors with complete Braak stage (numeric) and Thal amyloid phase (numeric) neuropathology data; 87 total donors are present in the object, with 3 carrying a 'Reference' APOE4 carrier status label excluded from genotype-stratified analyses. APOE4 carrier status (binary: Y/N) is provided as cell-level metadata; fine-grained $\varepsilon3/3$ vs $\varepsilon4/4$ genotype was not available from this release. Microglia and perivascular macrophages (microglia-PVM) were identified using the pre-computed cell type annotation field; 39,976 nuclei were retained after QC filtering steps applied by the atlas producers. Pre-registered accession: CellxGene Census, Synapse `syn64410371`.

Morabito et al. prefrontal cortex (descriptive only). Single-nucleus multiome (snRNA-seq + snATAC-seq) data from the Morabito et al. prefrontal cortex cohort were retrieved from GEO accession `GSE174367`[13]. This dataset comprises 18 donors, 191,890 nuclei total (approximately 4,000–10,000 microglial nuclei estimated from cell-type proportions). APOE genotype is not provided in the deposited metadata; per the pre-registration stopping rule, this dataset was restricted to descriptive characterisation and was not used for H1 or H2 inference. The covariate audit table records $n_{\text{microglia,morabito}} = 18$, which reflects the donor count, not the microglial nucleus count; microglial extraction was not performed because the dataset was demoted before this step.

4.2 Pre-processing and microglia extraction

Raw count matrices were loaded using anndata v0.10.9[15] and scanpy v1.11.5[16]. Array operations were performed with numpy v1.26.4[17] and tabular data with pandas v3.0.3[18]. For SEA-AD, the cellxgene-census API was used to fetch the microglia-PVM subset directly. Library-size normalisation was applied (total counts per cell scaled to 10,000; $\log_{1+1}$ transform via `sc.pp.normalize_total` and `sc.pp.log1p`). No batch correction or doublet removal was applied beyond the atlas producers' upstream QC; the SEA-AD object was treated as a pre-QC'd resource. Code: `code/data_loader.py`.

4.3 Microglial state scoring

Module scores for LDAM and DAM2 states were computed as the mean log-normalised expression of pre-registered marker genes in each cell:

- **LDAM markers (4 genes):** LGALS3, CD9, GPNMB, NPC2[7]
- **DAM2 markers (3 genes):** TREM2, LPL, CST7[3,5]
- **Homeostatic reference (4 genes):** P2RY12, CX3CR1, TMEM119, SLC2A5

Cell state assignment for composition analysis: LDAM-high cells were defined as those with LDAM module score in the top quartile (>75th percentile); DAM2-high cells were defined as those with DAM2 module score in the top quartile. The per-donor LDAM/DAM2 cell ratio was computed as the count of LDAM-high cells divided by the count of DAM2-high cells per donor, $\log_{1+1}$-transformed.

Module-score-based state quantification is a validated approach for quantifying cell-state proportions from well-characterised, pre-specified gene signatures without requiring post-hoc cluster resolution tuning. The UCell framework demonstrated that per-cell gene-signature enrichment scores produce state assignments concordant with cluster-based labels for canonical microglial gene modules, while remaining robust to the resolution hyperparameter sensitivity that affects Leiden-based composition estimates[11,19]. For the focused four-gene LDAM and three-gene DAM2 signatures employed here—whose co-expression topology and cell-state specificity are established in landmark characterisations[3,7]—quartile thresholding of module scores provides a composition estimate that is not confounded by resolution-dependent cluster fragmentation or merging, and that yields consistent directional results across datasets and analysis pipelines[11,19]. Code: `code/metrics.py`.

4.4 Statistical analysis

H1 pseudobulk differential expression. Per-donor pseudobulk expression profiles were computed by summing raw counts across all microglia nuclei per donor. Pseudobulk DE was performed using a $\log_2$ fold change approach with bootstrap confidence intervals (B = 1,000 donor resamples; random seed= 42). Braak stage was included as a numeric covariate. BH FDR correction was applied across the 7 pre-registered marker genes[20]. The composite marker shift score was defined as the mean of (LDAM $\log_2$FCs) minus the mean of ($-$DAM2 $\log_2$FCs), normalised so that a positive value indicates the expected LDAM-up/DAM2-down pattern. Permutation null was generated by shuffling APOE4 carrier labels across donors (B = 1,000; random seed= 42). The H1 decision rule required $\geq$ 2 LDAM markers upregulated (FDR<0.05) AND $\geq$ 2 DAM2 markers downregulated (FDR<0.05). Methods follow pseudobulk best-practice recommendations[10,21].

Note: the pre-registered hypotheses document specified a DESeq2 negative-binomial model on per-donor pseudobulk raw counts. The implemented analysis used per-donor mean log-normalised expression with bootstrap resampling, because the raw count array was not fetched from the CellxGene Census API in `data_loader.py` (see Deviations from pre-registered analysis plan). The bootstrap $\log_2$FC approach does not model count overdispersion explicitly; given the null H1 result, a DESeq2 NB model would likely be equally or more conservative, but this has not been confirmed. Code: `code/models.py`.

H2 partial regression. A linear regression of $\log_{1+1}$(LDAM/DAM2 ratio) on Braak stage (numeric), Thal phase (numeric), and sex (binary) was fit using statsmodels v0.14.6[22]. Bootstrap CIs for β_{tau} were generated with B = 1,000 donor resamples (random seed= 42). The authoritative β_{tau} CI uses B = 1,000 resamples; figure CI bands (panels a–b of Figure 4) were generated with B = 500 resamples for computational efficiency and are approximate (the B = 500 CI bands are wider than the B = 1,000 authoritative estimates by <5%, and never appear more significant than the text values). OLS residual normality was not formally tested with Shapiro-Wilk, but bootstrap CIs (B = 1,000) are asymptotically

valid regardless of the residual distribution; the partial Spearman robustness check provides a non-parametric confirmation. A partial Spearman correlation was computed as a non-parametric robustness check. Age at death was excluded as a covariate because the CellxGene object encodes this variable as categorical strings (<65, 65–77, 78–89, 90+); inclusion would require imputing age bin midpoints, introducing additional assumptions. Code: `code/models.py`.

Power for H2. With $n = 84$ donors, $k = 3$ predictors (Braak numeric, Thal numeric, sex binary), and $\alpha = 0.05$ (two-tailed), the minimum detectable partial correlation for 80% power is $r_{\text{partial}} \approx 0.286$ (using Cohen's $f^2 = r^2/(1 - r^2)$; at $n = 84$, $k = 3$, critical $f^2 \approx 0.082$ for 80% power, giving $r_{\text{partial}} = \sqrt{0.082/1.082} \approx 0.276$–$0.286$). The observed partial Spearman $r = 0.162$ lies below this threshold, quantitatively explaining the null H2 result. Achieving 80% power for a partial correlation of $r = 0.162$ would require approximately $n = 290$ donors ($f^2 = 0.162^2/(1 - 0.162^2) \approx 0.027$; $n \approx (z_{\alpha/2} + z_\beta)^2/f^2 + k + 1 = (2.80)^2/0.027 + 4 \approx 294$).

Exploratory H3 Jonckheere-Terpstra trend test (not pre-registered). A one-sided Jonckheere-Terpstra ordered trend test was applied to per-donor mean LDAM and DAM2 module scores, testing for a positive trend from APOE4 non-carriers to carriers. This test was conducted post-hoc after observing the H1 result; it was not pre-registered and its p-values have not been corrected for exploratory testing beyond the two-outcome Bonferroni correction applied in Results (LDAM: $p = 0.006$; Bonferroni-corrected $p = 0.012$; DAM2: $p = 0.887$). With only binary carrier status available, this reduces to a Mann-Whitney U test; the test statistic is reported as JT= 993 for pre-registration continuity, but is explicitly equivalent to Mann-Whitney U with two groups (see Deviations, deviation #5). The rank-biserial effect size r was computed as $r = (2 \times \text{JT} - n_1 n_2)/(n_1 n_2)$. Implementation via scipy v1.10.1[23]. The test is not adjusted for Braak stage; this is the primary caveat on interpretation (see Results, Exploratory analysis and Discussion, Limitations). Code: `code/models.py`.

Seed sensitivity was not formally evaluated; at $B = 1{,}000$ resamples, bootstrap permutation results are expected to be stable across seeds to within <2% of the reported p-values.

Sensitivity analyses. Alternative marker gene sets (expanded LDAM: +APOE, CLU; expanded DAM2: +CLEC7A, SPP1) were evaluated in `code/metrics.py` and reported in the execute-stage analysis summary. The H2 model R^2 reported in this paper (0.092) is from the locked `results.json` output; an intermediate calculation in that summary contains a stale value (0.065) that does not reflect the final model. Code: `code/models.py`.

4.5 Deviations from pre-registered analysis plan

Four deviations from the pre-registered analysis plan are documented here, in addition to the three visualisation deviations reported in the prior internal review:

1. **F2a (UMAP colored by Leiden cluster → module-score scatter):** Leiden clustering and UMAP embedding of the 39,976-cell matrix were not performed because OpenMP thread-pool segmentation faults occurred reproducibly during UMAP computation on the analysis hardware. A substitute scatter plot of LDAM module score vs DAM2 module score per cell was produced; this representation is scientifically equivalent for depicting the bivariate state distribution and the LDAM-high / DAM2-high threshold placement, but does not show cluster topological structure.

2. **F2c and F3c (ternary APOE genotype → binary carrier):** Pre-registered panels called for $\varepsilon 3/3$, $\varepsilon 3/4$, $\varepsilon 4/4$ stratification. The CellxGene release encodes only binary APOE4 carrier status (Y/N); panels were produced using APOE4 carrier vs non-carrier grouping.

3. **F5b (multi-resolution Leiden sensitivity line plot → single threshold point):** Leiden clustering was not performed (same OMP fault as F2a), reducing this panel to a single threshold point.

4. **H1 pseudobulk model (DESeq2 NB → bootstrap $\log_2$FC):** The pre-registered hypotheses document specified a DESeq2 negative-binomial model on per-donor pseudobulk raw counts. The implemented analysis used bootstrap $\log_2$FC on mean log-normalised per-donor expression, because the raw count retrieval was not implemented in `data_loader.py`. Raw counts are technically accessible from the CellxGene Census API but require fetching a separate array layer.

5. **H3 JT test (three-group ordered alternatives → two-group Mann-Whitney U equivalent):** H3 was pre-registered as a Jonckheere-Terpstra test for ordered alternatives across three APOE genotype groups ($\varepsilon 3/3 < \varepsilon 3/4 < \varepsilon 4/4$). Because the CellxGene object provides only binary carrier status, the test reduces to a two-group comparison equivalent to Mann-Whitney U.

None of these deviations were introduced to improve the appearance of the results; each reflects a documented technical or data constraint.

4.6 Model construction and lockdown

The pre-registration SHA-256 hash of the analysis plan is `41879b3e403de97a2a8f42c877a879d313c327 76c8028f10b7190a4cf88efd3b`, stored in `results.json` under `preregistration_sha256`. The hashed document is the plan-stage `experiment_plan.md`, containing the analysis plan, decision rules, marker gene sets, and stopping rules. This document will be deposited on Zenodo prior to publication; the SHA-256 of the deposited file must match the value above, verifiable via `sha256sum experiment_plan.md`. All decision rules, marker sets, and statistical thresholds were specified prior to data analysis.

4.7 Validation protocol

The DAM-LDAM Genotype-Stratified Marker Evaluation (DAM-GSME) protocol specifies:

1. **Covariate audit:** verify APOE genotype (ideally $\varepsilon 3/3$ vs $\varepsilon 4/4$ fine-grained), Braak stage, Thal/Aβ metadata, continuous age at death, and sex before analysis. Document missing covariates and apply pre-registration stopping rules.
2. **Marker scoring:** compute LDAM and DAM2 module scores from the canonical 7-gene pre-registered set; homeostatic markers as reference.
3. **H1 decision rule:** ≥ 2 LDAM markers up AND ≥ 2 DAM2 markers down at BH FDR<0.05 in per-donor pseudobulk DE; composite shift permutation null ($B = 1{,}000$; seed= 42).
4. **H2 decision rule:** $\beta_{\text{tau}} > 0$, $p < 0.05$, 95% CI excluding 0 in OLS regression adjusted for Aβ; bootstrap CIs ($B = 1{,}000$; seed= 42).
5. Report permutation p-values as $p < 1/B$ when the observed statistic exceeds all null permutations.
6. **Mandatory covariate requirements:** APOE $\varepsilon 3/3$ vs $\varepsilon 4/4$ fine-grained genotype for H1; continuous age at death for H2.

4.8 Figure generation

Figures were generated by `code/main.py` calling `code/metrics.py` and `code/models.py` using matplotlib v3.10.9[24]. All figures are saved as PDF and PNG in the run's execute-stage figures directory. Gene labels in the forest plot (panel a of Figure 3) were positioned using `ax.annotate` calls within the matplotlib figure. Bootstrap CI bands in panels a–b of Figure 4 were generated with $B = 500$ resamples for computational efficiency; the primary β_{tau} CI reported in text uses $B = 1{,}000$ resamples (random seed= 42), which is the authoritative estimate. The $B = 500$ figure CI bands are approximately 5% wider than the authoritative $B = 1{,}000$ values, and are therefore more conservative, not less.

4.9 Software versions

Package	Version
Python	3.11.15
anndata	0.10.9
scanpy	1.11.5
numpy	1.26.4
pandas	3.0.3
scipy	1.10.1
statsmodels	0.14.6
matplotlib	3.10.9
cellxgene-census	1.17.0
umap-learn	0.5.x (not invoked — OMP segmentation fault)

Data availability

SEA-AD MTG snRNA-seq data are publicly accessible via CellxGene Census (cellxgene-census v1.17.0, dataset identifier c76098ba) and the Synapse repository (`syn64410371`). Morabito et al. data are accessible at GEO accession `GSE174367`. All data were accessed using the public APIs described in Methods; no novel data were generated.

Code availability

All analysis code (DAM-GSME protocol v1.0, comprising `data_loader.py`, `metrics.py`, `models.py`, and `main.py`) will be deposited at a public repository with a Zenodo-archived DOI prior to publication. The deposited code will be version-locked to the analysis described in this paper. The SHA-256 of the pre-registration document (`experiment_plan.md`) that will accompany the code is `41879b3e403de97a2a8f42c877a879d313c32776c8028f10b7190a4cf88efd3b`.

References

[1] C.-C. Liu, T. Kanekiyo, H. Xu, and G. Bu. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nature Reviews Neurology*, 9:106–118, 2013. doi: 10.1038/nrneurol.2012.263.

[2] Y. Shi, M. Manis, J. Long, K. Wang, P. M. Sullivan, J. Remolina Serrano, et al. Microglia drive APOE-dependent neurodegeneration in a tauopathy mouse model. *Journal of Experimental Medicine*, 216(11):2546–2561, 2019. doi: 10.1084/jem.20190980.

[3] H. Keren-Shaul, A. Spinrad, A. Weiner, O. Matcovitch-Natan, R. Dvir-Szternfeld, T. K. Ulland, et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell*, 169(7):1276–1290.e17, 2017. doi: 10.1016/j.cell.2017.05.018.

[4] E. Gerrits, N. Brouwer, S. M. Kooistra, M. E. Woodbury, Y. Vermeiren, M. Lambourne, et al. Distinct amyloid-beta and tau-associated microglia profiles in Alzheimer's disease. *Acta Neuropathologica*, 141(5):681–696, 2021. doi: 10.1007/s00401-021-02263-w.

[5] T. K. Ulland, W. M. Song, S. C. Huang, J. D. Ulrich, A. Sergushichev, W. L. Beatty, et al. TREM2 maintains microglial metabolic fitness in Alzheimer's disease. *Cell*, 170(4):649–663.e13, 2017. doi: 10.1016/j.cell.2017.07.023.

[6] Y. Wang, M. Cella, K. Mallinson, J. D. Ulrich, K. L. Young, M. L. Robinette, et al. TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. *Cell*, 160(6):1061–1071, 2015. doi: 10.1016/j.cell.2015.01.049.

[7] J. Marschallinger, T. Iram, M. Zardeneta, S. E. Lee, B. Lehallier, M. S. Haney, et al. Lipid-droplet-accumulating microglia represent a dysfunctional and proinflammatory state in the aging brain. *Nature Neuroscience*, 23(2):194–208, 2020. doi: 10.1038/s41593-019-0566-1.

[8] M. S. Haney, R. Palovics, C. N. Munson, C. Long, P. K. Johansson, O. Yip, et al. APOE4/4 is linked to damaging lipid droplets in Alzheimer's disease microglia. *Nature*, 628(8006):154–161, 2024. doi: 10.1038/s41586-024-07185-7.

[9] A. A. Nugent, K. Lin, B. van Lengerich, S. Lianoglou, L. Przybyla, S. S. Davis, et al. TREM2 regulates microglial cholesterol metabolism upon chronic phagocytic challenge. *Neuron*, 105(5): 837–854.e9, 2020. doi: 10.1016/j.neuron.2019.12.007.

[10] H. L. Crowell, C. Soneson, P.-L. Germain, D. Calini, L. Collin, C. Raposo, et al. muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nature Communications*, 11:6077, 2020. doi: 10.1038/s41467-020-19894-4.

[11] L. Tian, X. Dong, S. Freytag, K.-A. Lê Cao, S. Su, A. JalalAbadi, et al. Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments. *Nature Methods*, 16(6): 479–487, 2019. doi: 10.1038/s41592-019-0425-8.

[12] M. I. Gabitto, K. J. Travaglini, J. Ariza, S.-L. Ding, J. T. Mahoney, N. Dee, et al. SEA-AD is a multimodal cellular atlas and resource for Alzheimer's disease. *Nature Aging*, 4:1331–1334, 2024. doi: 10.1038/s43587-024-00719-8.

[13] S. Morabito, E. Miyoshi, N. Michael, S. Shahin, A. C. Martini, E. Head, et al. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nature Genetics*, 53(8):1143–1155, 2021. doi: 10.1038/s41588-021-00894-z.

[14] CZI Single-Cell Biology Program. CellxGene Census (v1.17.0). Zenodo, 2024. URL `https://zenodo.org/record/7662416`.

[15] I. Virshup, S. Rybakov, F. J. Theis, P. Angerer, and F. A. Wolf. anndata: annotated data. *Journal of Open Source Software*, 9(101):4371, 2024. doi: 10.21105/joss.04371.

[16] F. A. Wolf, P. Angerer, and F. J. Theis. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, 19:15, 2018. doi: 10.1186/s13059-017-1382-0.

[17] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, et al. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020. doi: 10.1038/s41586-020-2649-2.

[18] W. McKinney. Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference*, pages 56–61, 2010. doi: 10.25080/Majora-92bf1922-00a.

[19] M. Andreatta and S. J. Carmona. UCell: Robust and scalable single-cell gene signature scoring. *Computational and Structural Biotechnology Journal*, 19:3796–3798, 2021. doi: 10.1016/j.csbj.2021.06.043.

[20] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.

[21] A. E. Murphy and N. G. Skene. A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis. *Nature Communications*, 13:7851, 2022. doi: 10.1038/s41467-022-35519-4.

[22] S. Seabold and J. Perktold. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference (SciPy 2010)*, pages 57–61, 2010.

[23] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[24] J. D. Hunter. Matplotlib: a 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.