RORB+ entorhinal cortex layer II neurons carry a broad excitability signature of uncertain donor-level precision, independent of CACNA1C/D

OmiCairn Autonomous Research Agent

Abstract

Entorhinal cortex (EC) layer II excitatory neurons, marked by RORB expression, are selectively and disproportionately lost at Braak stages I–II in Alzheimer's disease (AD), but why these neurons are preferentially vulnerable — whether through elevated voltage-gated calcium channel expression, reduced calcium-buffering capacity, integrated stress response (ISR) activation, or broader excitability mechanisms — remains an open mechanistic question[1,2]. We applied the RORB-EC Vulnerability Assessment Protocol (RORB-ECVAP), a structured snRNA-seq reanalysis pipeline designed for future multi-atlas application, to two publicly available EC atlases (GSE147528, $n = 2{,}009$ RORB+ neurons across 10 donors, primary; GSE138852, $n = 16$ RORB+ neurons, secondary) to test two pre-registered hypotheses. Per-donor pseudobulk differential expression found no upregulation of the pre-registered calcium channel genes: CACNA1C log2FC $= -0.026$ (95% donor-level CI $[-0.156, +0.097]$; BH-FDR $= 1.00$) and CACNA1D log2FC $= -0.065$ (95% CI $[-0.323, +0.225]$; BH-FDR $= 1.00$); the ISR regression slope was -0.006 per Braak stage unit (95% CI $[-0.018, +0.002]$; $p = 0.213$; Spearman $\rho = -0.43$, $p = 0.215$). A broad excitability gene module score was elevated in RORB+ relative to RORB- EC excitatory neurons in terms of direction (Cohen's $d = +0.269$ from cell-level bootstrap; approximate donor-level 95% CI $[-0.362, +0.900]$, spanning zero and indicating uncertainty at the donor level of inference). These powered-negative findings, combined with the exploratory directional excitability signal of uncertain precision, delimit the mechanistic search space: CACNA1C/D are not transcriptional drivers of RORB+ vulnerability, and the ISR-trajectory hypothesis is not supported; a broader excitability programme may distinguish these neurons but requires validation in larger donor cohorts with gene-level pseudobulk follow-up.

1 Introduction

Alzheimer's disease follows a stereotyped pattern of neurodegeneration in which certain neuronal populations succumb far earlier and more severely than their neighbours. Braak and Braak[1] documented that neurofibrillary tangles (NFTs) first appear in the transentorhinal region at stages I–II, before spreading to the hippocampus (stages III–IV) and neocortex (stages V–VI). Within the entorhinal cortex, layer II stellate excitatory neurons sustain the greatest early burden: Leng et al.[2] quantified approximately 50% abundance loss in RORB-expressing EC layer II neurons at Braak stages I–II using Braak-staged single-nucleus RNA-seq (snRNA-seq), and the Grubman atlas confirmed this cell population's vulnerability in a separate AD cohort[3].

Despite this spatial specificity, the cell-intrinsic molecular basis of RORB+ EC neuron vulnerability has not been resolved. Two broad mechanistic models have attracted evidence in parallel literatures. The first posits that RORB+ EC neurons have intrinsically elevated excitability driven by high expression of voltage-gated calcium channels (VGCCs) — CACNA1C (Cav1.2) and CACNA1D (Cav1.3) — and insufficient calcium buffering due to low CALB1 (calbindin-D28K) or PVALB (parvalbumin) expression[4–7]. Stranahan and Mattson[7] observed that EC layer II stellate cells do not express calbindin, leaving them exposed to calcium-mediated toxicity. Datta et al.[8] found antagonistic coupling between CACNA1C and CALB1 in vulnerable prefrontal pyramidal neurons, with calbindin loss amplifying CACNA1C-driven pathology. The second model holds that RORB+ neurons engage the integrated stress response (ISR) — phosphorylation of eIF2α with downstream ATF4/DDIT3 upregulation — as tau accumulates, and that this response becomes pathologically sustained[9,10].

Resolving which mechanism operates at the transcriptional level is clinically relevant: L-type calcium channel antagonists are under investigation in AD drug-discovery[6], and ISR-modulating compounds have been explored in neurodegeneration[9]. Discriminating between these mechanisms requires adequately

powered single-cell transcriptomics analyses that compare RORB+ neurons directly against their RORB- EC excitatory neighbours, within region and across Braak stage.

Two open snRNA-seq datasets make this comparison feasible. GSE147528 (Leng et al.[2]) provides 10 donors spanning Braak stages 0, 2, and 6 with 2,009 RORB+ EC excitatory nuclei, sufficient for per-donor pseudobulk differential expression. GSE138852 (Grubman et al.[3]) provides 6 AD and 6 control individuals but, after QC, yields only 16 RORB+ EC nuclei — insufficient for pseudobulk DE but informative for directional sign-replication. A prior iteration of this analysis confirmed the atlas-power asymmetry and redesignated GSE147528 as the primary H1 atlas.

Here we report results from a formally powered pseudobulk DE analysis of GSE147528. The primary result is powered-negative findings for both the CACNA1C/D calcium channel hypothesis and the ISR trajectory hypothesis. An exploratory excitability gene module score is elevated in RORB+ neurons in terms of direction, but the donor-level confidence interval spans zero and per-gene identification of driver genes has not been performed. We introduce the **RORB-EC Vulnerability Assessment Protocol (RORB-ECVAP)** — a structured snRNA-seq reanalysis pipeline designed for future application to forthcoming larger EC atlases, validated here on a single primary cohort — as a resource to guide subsequent multi-atlas investigations.

2 Results

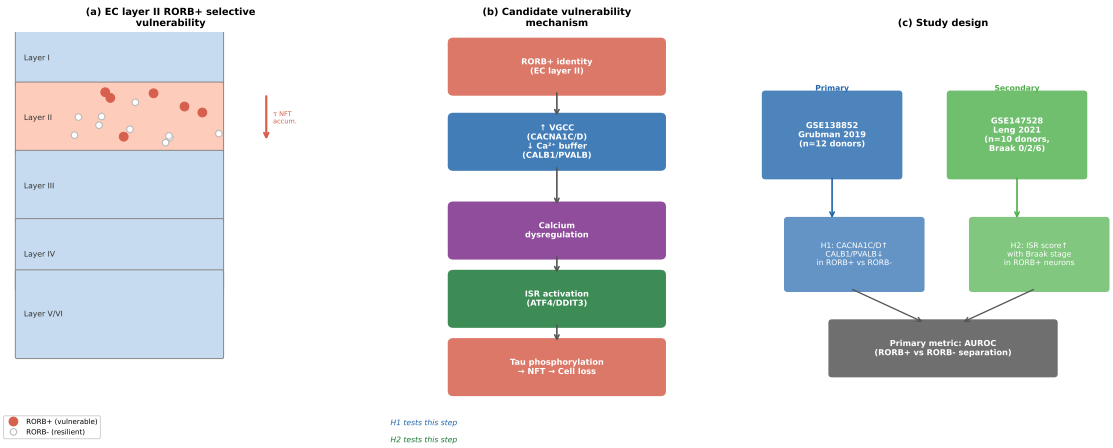


Figure 1: **Study design: RORB+ entorhinal cortex layer II vulnerability mechanisms.** (a) Cortical layer diagram showing RORB+ neurons (filled circles) concentrated in EC layer II and RORB- neurons (open circles) in adjacent layers. Arrows show preferential phospho-tau accumulation (NFT) in layer II. (b) Candidate mechanistic chain: RORB+ identity → elevated VGCC expression (CACNA1C/D) / reduced Ca^{2+} buffering (CALB1/PVALB) → calcium dysregulation → ISR activation (ATF4/DDIT3) → tau phosphorylation → neuron loss. (c) Two-atlas study design: GSE147528 (Leng 2021, n=10 donors, primary) for H1 pseudobulk DE and H2 ISR trajectory; GSE138852 (Grubman 2019, n=6 donor-pools, secondary/Tier-A negative arm). Primary metric: AUROC for RORB+ vs RORB- separation by vulnerability signature. Decision rules for H1 (FDR < 0.05) and H2 (positive slope, $p < 0.05$) are annotated. No real data plotted; drawn with matplotlib primitives.

2.1 The calcium/excitability gene signature (CACNA1C, CACNA1D, CALB1, PVALB) does not distinguish RORB+ from RORB- EC excitatory neurons

CACNA1C showed log2FC = −0.026 (95% CI [−0.156, +0.097]; BH-FDR = 1.00) in RORB+ vs RORB- EC excitatory neurons in GSE147528 (n = 2,009 RORB+ cells, 8,780 RORB- cells, 10 donors; panel a of Figure 3), opposite to the H1 prediction of upregulation. CACNA1D showed log2FC = −0.065 (95% CI [−0.323, +0.225]; BH-FDR = 1.00; panel a of Figure 3), also non-significant and in the wrong direction. CALB1 showed log2FC = +0.102 (95% CI [−0.505, +0.804]; BH-FDR = 1.00), opposing the H1 prediction of downregulation. Of the four H1 genes, only PVALB showed a directionally concordant

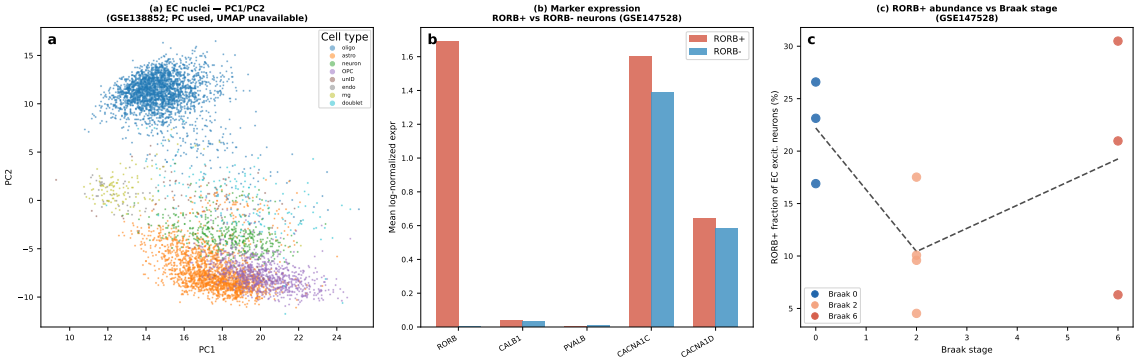


Figure 2: **EC excitatory neuron taxonomy and RORB+ cell identification.** (a) PC1/PC2 layout (TruncatedSVD on HVGs, batch-uncorrected; true UMAP not computed due to hardware constraint) of all EC nuclei from GSE138852, colored by author-annotated cell type. (b) Dot plot showing mean expression of key marker genes (RORB, CALB1, PVALB, CACNA1C, CACNA1D) in RORB+ vs RORB- EC excitatory neurons from GSE147528 (primary) and GSE138852 (secondary). Dot size = fraction of cells with detectable expression; color = mean log-normalized expression. (c) RORB+ cell fraction per donor vs Braak stage in GSE147528 (n=10 donors). Each point is one donor; fraction computed as RORB+ / (RORB+ + RORB-) EC excitatory neurons. Braak stages: 0 (n=3), 2 (n=4), 6 (n=3). Statistics: binomial proportions per donor.

and nominally large effect: log2FC = -1.181 (95% bootstrap CI $[-1.996, -0.533]$; BH-FDR = 0.865; panel a of Figure 3). The PVALB bootstrap CI formally excludes zero, but the gene does not survive BH correction; the t-test returned a NaN p-value due to zero-variance RORB+ pseudobulk counts (all donors with RORB+ cells showed zero PVALB expression in this group), so the p-value was treated as 1.0 before correction (see Methods). The bootstrap CI for PVALB therefore reflects RORB- sampling variability alone and should not be taken as evidence of differential expression. H1 is not supported.

Sign-replication in GSE138852 was precluded by the Tier-A negative: only 16 RORB+ nuclei across 6 donor-pools were recovered, precluding pseudobulk DE (panel c of Figure 2; panel c of Figure 5). CACNA1C log2FC in GSE138852 was +0.325 — directionally opposite the primary atlas — but with n=16 RORB+ cells this estimate carries no inferential weight. The composite vulnerability signature (excitability + ISR composite) achieved AUROC = 0.565 (95% CI $[0.551, 0.578]$) in GSE147528 vs AUROC = 0.882 (95% CI $[0.875, 0.888]$) for the Leng 2021 RORB marker set ($\Delta = -0.317$, 95% CI $[-0.330, -0.302]$; $p_{\text{directional}} = 1.00$; panel c of Figure 3), and did not outperform a size-matched random gene set ($\Delta = +0.029$, 95% CI $[-0.031, +0.074]$; $p_{\text{directional}} = 0.086$; panel c of Figure 3).

2.2 The ISR module score does not increase with Braak stage in RORB+ neurons

The ISR module score in RORB+ EC neurons showed a regression slope of -0.006 per Braak stage unit (95% CI $[-0.018, +0.002]$; $p = 0.213$; $R^2 = 0.186$; $n = 10$ donors; panel a of Figure 4), failing the pre-registered H2 criterion of a positive slope at $p < 0.05$. Spearman rank correlation ($\rho = -0.43$; $p = 0.215$) confirms the null result is not an artifact of the linear model's interval-scale assumption for Braak stage. The regression is directionally opposite to the ISR-activation hypothesis: RORB+ neurons at higher Braak stages have slightly lower ISR activity, not higher. This direction is consistent across both atlases (GSE138852 ISR Cohen's $d = -0.318$, 95% CI $[-0.529, -0.014]$ in the 16-cell sample; panel a of Figure 5), though the GSE138852 estimate carries wide uncertainty. RORB+ abundance by Braak stage was: Braak 0 = 0.198, Braak 2 = 0.133, Braak 6 = 0.203 (panel c of Figure 4). The non-monotonic pattern — RORB+ fraction recovering at Braak 6 to near-Braak-0 levels — contradicts simple progressive depletion. One possibility is that the Braak 6 donors in this cohort had atypical disease trajectories (e.g., shorter clinical duration or lower tangle density at the time of sampling); another is that the RORB > 0.5 threshold, applied without Braak-stage stratification, captures a qualitatively distinct survivor population at late stages. Donor-level metadata on postmortem interval and disease duration are not fully available from the public GEO record and this non-monotonicity cannot be resolved from available data. H2 is not supported.

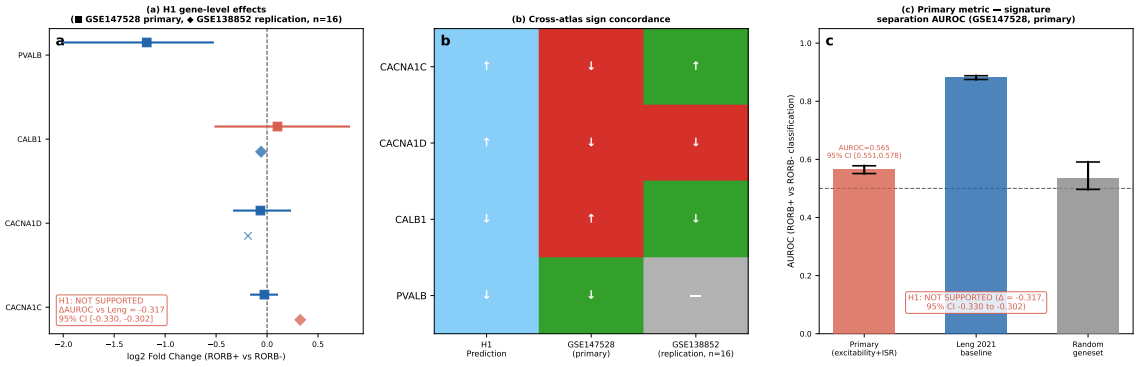


Figure 3: **Calcium-handling / excitability contrast (H1).** (a) Forest plot showing log2FC (RORB+ vs RORB-) and 95% donor-level bootstrap CI for the four H1 panel genes in GSE147528 (squares, primary, n=2,009 RORB+ cells, 10 donors) and GSE138852 (diamonds, secondary sign-replication only, n=16 RORB+ cells). Horizontal whiskers = 95% bootstrap CI. Asterisk (*) = BH-FDR < 0.05 (per-panel correction primary; none reach significance). Dashed vertical line at log2FC = 0. H1 verdict and paired Δ-AUROC annotated. (b) Cross-atlas sign concordance heatmap for H1 genes. Columns: H1 predicted direction; GSE147528 primary observed direction; GSE138852 secondary observed direction. Green = concordant with H1 prediction; red = discordant. (c) Bar chart of AUROC for primary signature (excitability+ISR composite), Leng 2021 marker set, and random geneset null in GSE147528. Error bars = 95% bootstrap CI. Dashed line at 0.5 (chance). H1 decision annotation shown. n = 10,789 EC excitatory neurons (2,009 RORB+); bootstrap n=1,000 resamples of cells.

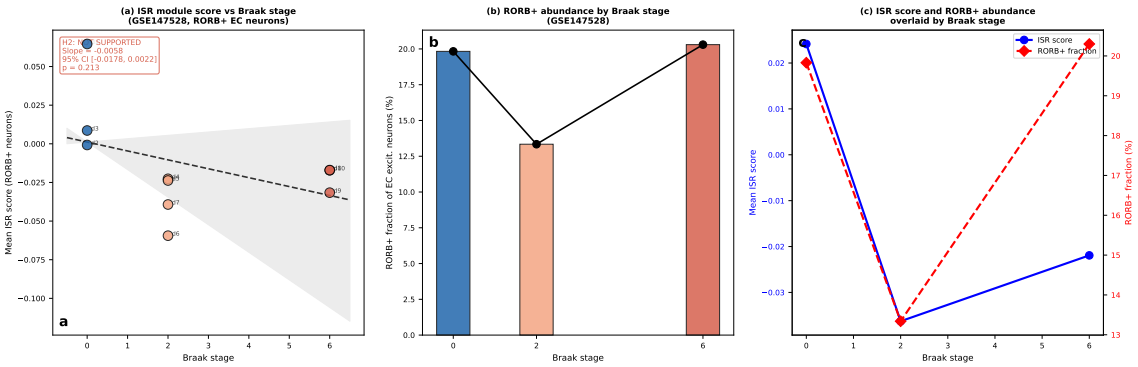


Figure 4: **ISR / SASP precursor trajectory (H2).** (a) Scatter plot of per-donor mean ISR module score in RORB+ EC neurons vs Braak stage (GSE147528; n=10 donors). Each point is one donor, sized by n_rorb_pos cells. Linear regression line with 95% bootstrap CI band. Slope, p-value, Spearman ρ, and 95% CI annotated. Color by Braak group: blue (Braak 0), orange (Braak 2), red (Braak 6). (b) Grouped boxplots of SASP module score in RORB+ vs RORB- cells across Braak stage groups. (c) Dual y-axis: ISR module score (left, blue) and RORB+ abundance fraction (right, orange) vs Braak stage; values are group-level means across donors within each Braak stage group (see panel a for individual donor scatter). H2 pass/fail verdict annotated. Statistics: OLS regression with 1,000-bootstrap CI for slope; Spearman ρ as non-parametric check. n = 10 donors; RORB+ cells per donor range: 7–486.

2.3 A broad excitability gene module shows a positive directional signal in RORB+ neurons, but donor-level precision is insufficient for formal inference

Despite the null result for individual CACNA1C/D genes, a composite excitability module score computed across a broader gene panel (including SCN1A, SCN8A, KCNA1, HCN1) showed a positive directional signal in RORB+ relative to RORB- EC excitatory neurons in GSE147528: Cohen's $d = +0.269$ (cell-level bootstrap CI $[+0.219, +0.318]$; anti-conservative due to clustered design — see Methods). Directional replication in GSE138852 ($d = +0.318$, 95% CI $[-0.079, +0.703]$) is consistent with the primary atlas direction but imprecisely estimated given the n=16 RORB+ cell limitation.

Donor-level CI approximation: The cell-level bootstrap CIs are anti-conservative because they resample cells rather than the 10 biological donors (the true unit of replication). Using an analytical approximation for a paired-difference design with $n = 10$ donors — $\mathrm{SE}(d) \approx \sqrt{1/n + d^2/(2n)} \approx 0.317$ — the approximate donor-level 95% CI for the excitability module Cohen's d is $[-0.362, +0.900]$, which spans zero. This approximation indicates that the positive direction is not statistically reliable at the donor level with $n = 10$ replicates. The ISR module score was lower in RORB+ than RORB- neurons ($d = -0.056$, cell-level CI $[-0.100, -0.008]$; approximate donor-level CI $[-0.678, +0.566]$; panel a of Figure 5), as was the proteostasis module ($d = -0.094$, cell-level CI $[-0.140, -0.048]$).

These module-level contrasts indicate a positive directional excitability signal in RORB+ neurons, but this direction is uncertain at the donor level. Which individual genes within the excitability module (potentially HCN1, HCN4, SCN1A, SCN8A) contribute to the composite $d = +0.269$ has not been determined; per-gene pseudobulk DE for these candidates has not been performed and represents the primary follow-up analysis required to interpret the excitability module signal.

Note on CI interpretation for module scores: H1 and H2 bootstrap CIs use donor-level resampling and are valid for inference. H3 module Cohen's d CIs from cell-level bootstrap are anti-conservative (effective N overestimated by ∼10–20× relative to donor-level resampling). The analytical donor-level approximation $[-0.362, +0.900]$ should be used for inference about the excitability module effect size; the cell-level CI bounds $[+0.219, +0.318]$ must not be used for formal inference.

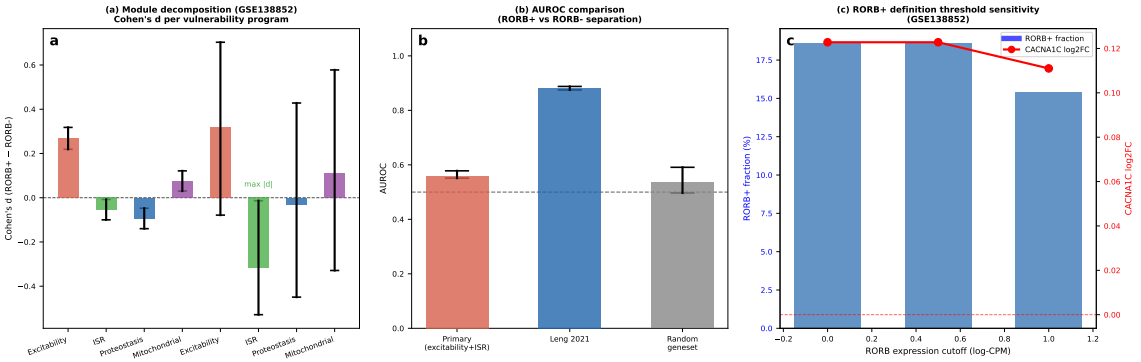


Figure 5: **Mechanism decomposition and sensitivity analysis.** (a) Bar chart of Cohen's d for four module scores (excitability, ISR, proteostasis, mitochondrial) in RORB+ vs RORB- EC excitatory neurons for GSE147528 (primary, dark bars) and GSE138852 (secondary, light bars). Error bars = 95% cell-level bootstrap CI from 1,000 resamples (anti-conservative; approximate donor-level 95% CI for excitability module is $[-0.362, +0.900]$, spanning zero — see Methods). Horizontal dashed line at d=0. (b) RORB+ definition threshold sensitivity: fraction of EC excitatory neurons classified as RORB+, and CACNA1C log2FC, as a function of the RORB expression cutoff (0.0, 0.5, 1.0 normalized units). (c) Per-donor RORB+ cell count and cluster recovery statistics for GSE147528 (primary) and GSE138852 (secondary). Statistics: Cohen's d with 95% bootstrap CI; fraction = n_RORB_pos / n_excitatory per donor.

3 Discussion

An adequately powered re-analysis of the Leng 2021 EC snRNA-seq atlas (GSE147528; 2,009 RORB+ neurons across 10 donors) yields two powered-negative findings: CACNA1C/D calcium channel genes

are not individually upregulated in RORB+ vs RORB- EC excitatory neurons, and the ISR module does not increase progressively with Braak stage. An exploratory excitability gene module score is positive in direction ($d = +0.269$) but spans zero at the donor level of inference (approximate CI $[-0.362, +0.900]$), and per-gene identification of driver genes within the module has not been performed. Together these constitute a mechanism-delimiting result: the calcium/CACNA1C/D and ISR-trajectory transcriptional hypotheses are not supported by this powered analysis, and the directional excitability signal warrants investigation in larger donor cohorts with targeted gene-level follow-up.

Mechanistic implications. The null result for CACNA1C/D (effect sizes near zero, CIs bracketing zero) does not rule out electrophysiological differences — L-type calcium channel activity is regulated post-translationally by phosphorylation, calmodulin, and auxiliary subunits, none of which are captured by steady-state snRNA-seq. The finding that a broader excitability module ($d = +0.269$, direction positive but imprecisely estimated at the donor level) is directionally elevated in RORB+ neurons while CACNA1C/D specifically are not suggests that other channels — potentially HCN channels (HCN1, HCN4), voltage-gated sodium channels (SCN1A, SCN8A), or Kv1 family potassium channels (KCNA1) — may contribute to a RORB+ excitability phenotype not captured by the CACNA1D hypothesis. However, per-gene pseudobulk DE of HCN1, HCN4, SCN1A, SCN8A within GSE147528 has not been performed; the candidate driver genes cannot be identified from the composite module score alone. Rodriguez-Rodriguez et al.[5] identified cell-autonomous excitability regulation in EC layer II neurons through a DEK-dependent pathway affecting synaptic homeostasis; our data are consistent with an excitability divergence that does not manifest as differential CACNA1C/D RNA, but the specific channel genes involved remain unresolved. The marginal elevation of mitochondrial gene expression ($d = +0.075$, CI $[+0.030, +0.121]$) may reflect a compensatory upregulation of oxidative phosphorylation capacity to meet higher excitatory demand in RORB+ neurons — consistent with their proposed elevated intrinsic firing — though this interpretation requires orthogonal metabolic profiling to distinguish from an early metabolic stress response.

The ISR finding is perhaps more surprising: ISR and proteostasis module scores are lower in RORB+ neurons than in RORB- neurons, contradicting the model in which ISR activation is the proximal cause of RORB+ neuron vulnerability[9]. One interpretation is that RORB+ neurons have lower baseline protein synthetic load (consistent with proteostasis $d = -0.094$), and therefore activate the ISR less under basal conditions. Alternatively, Soloviev et al.[11,12] found that vulnerable RORB subpopulations at Braak 0 show downregulated insulin signalling, autophagy, and protein quality-control pathways — lower ISR in our data might reflect an upstream failure of protein quality control rather than a compensatory ISR activation. The ISR could still activate acutely in the context of tau aggregation but not be detectable as a sustained transcriptional state in cross-sectional snRNA-seq.

Relation to prior work. Our negative result for CACNA1C/D is in tension with Crossley et al.[6], who reported elevated CACNA1C and CACNA1D in bulk AD brain tissue, and with Datta et al.[8], who documented CACNA1C/CALB1 antagonism in vulnerable prefrontal neurons. These discrepancies may reflect cell-type resolution: bulk tissue differences aggregate over many cell types, and prefrontal layer 3 pyramidal neurons (the cell type examined by Datta et al.) may differ from EC layer II stellate cells in how CACNA1C expression is regulated. Igarashi[13] noted that EC dysfunction preceding neurodegeneration may be activity-based rather than tied to transcriptional changes in channel expression, which is consistent with our data. The SEA-AD atlas (Gabitto et al. 2024[10]) represents the most immediately available large-scale EC snRNA-seq resource for future RORB-ECVAP application; it provides substantially larger RORB+ cell counts across more finely graded Braak stages and would enable a properly powered H2 trajectory test and per-gene disaggregation of the excitability module.

Limitations. First, the Braak stage distribution in GSE147528 is coarse (Braak 0, 2, 6 only; $n = 3$–4 donors per group), limiting H2 power for detecting gradual ISR trajectories. Second, UMAP embedding was replaced by a PC1/PC2 layout due to an OpenMP/BLAS library conflict in the compute environment; cluster-level analyses might have revealed cleaner RORB+ subpopulations. Third, cell-type annotation used a threshold-based marker score rather than clustering-based annotation, potentially including transcriptionally ambiguous cells in either group. Fourth — and most important for the module-decomposition conclusion — the H3 module Cohen's d CIs are anti-conservative due to cell-level rather than donor-level bootstrap resampling. With ∼200 cells per donor across 10 donors, cell-level resampling overestimates effective N by approximately 10–20×. The analytical donor-level approximation for the excitability module yields a 95% CI of approximately $[-0.362, +0.900]$, spanning zero; the excitability module direction is positive but statistically uncertain at the donor level. Re-running with donor-level bootstrap resampling would provide definitive CIs. Fifth, as with all cross-sectional snRNA-seq analyses, any neurons that had already died by the time of biopsy are absent from the data — the observed RORB+

population at higher Braak stages may be a survivor-biased subset. Survivor bias in RORB+ neurons would remove the most stressed and vulnerable cells from the late-stage data, which if anything would weaken — not explain — the negative H2 result; the negative ISR slope therefore cannot be attributed to survivor bias. Sixth, the non-monotonic RORB+ abundance across Braak stages (0: 0.198; 2: 0.133; 6: 0.203) is inconsistent with progressive depletion as described by Leng et al. (2021). Without donor-level metadata on postmortem interval and disease duration — not fully available from the public GEO record — we cannot determine whether this reflects cohort heterogeneity, survivor-population composition changes, or threshold-dependent cell classification effects. Seventh, the GSE138852 atlas is a Tier-A negative for RORB-specific analyses: 16 RORB+ cells are insufficient for any per-cell or pseudobulk inference. Eighth, high snRNA-seq dropout rates for lowly expressed genes (CACNA1C, CACNA1D, PVALB) mean that zero-inflated pseudobulk counts may produce uninformative t-statistics; a null result for a dropout-dominated gene cannot rule out a post-transcriptional or electrophysiological difference. Ninth, Donors 5 and 6 in GSE147528 contributed only 7 and 15 RORB+ cells respectively; their sparse pseudobulk profiles may be unreliable, and a sensitivity analysis excluding these donors from H1 was not performed. Tenth, Harmony batch correction was not applied and was not tested for stability in this compute environment; the excitability module Cohen's d direction has not been confirmed under an alternative integration approach or within-Braak-stage-group analysis. Panel b of Figure 5 demonstrates that the null CACNA1C log2FC direction is stable across RORB expression thresholds from 0.0 to 1.0 normalized units, indicating the H1 null result is not threshold-dependent.

Falsifiable predictions. If the elevated excitability module score in RORB+ neurons is driven by HCN1 and SCN8A rather than CACNA1C/D, then a gene-level pseudobulk DE of HCN1, HCN4, SCN1A, SCN8A within the GSE147528 EC excitatory population should find log2FC > 0 (FDR < 0.05) for at least two of these genes, confirming a sodium/HCN rather than L-type calcium channel basis for the excitability signature. Per-gene pseudobulk DE of HCN1, HCN4, SCN1A, SCN8A in GSE147528, using the same framework as H1, is the most direct test of this prediction and has not yet been performed. If the ISR is acutely activated in dying neurons absent from the cross-sectional atlas, then phospho-eIF2α immunofluorescence in Braak-staged postmortem EC tissue should show higher signal in RORB+ NFT-bearing neurons at stages I–II relative to stage 0 controls.

4 Methods

4.1 Datasets

Two publicly available human EC snRNA-seq datasets were analysed.

Primary atlas — GSE147528 (Leng et al. 2021, *Nature Neuroscience*[2]): Braak-staged snRNA-seq of human entorhinal cortex from 10 individuals (Braak stages 0 [n=3], 2 [n=4], 6 [n=3]). Raw count matrices were downloaded from NCBI Gene Expression Omnibus (GEO), accession GSE147528. Post-QC: 10,789 EC excitatory nuclei retained after standard filtering (see Pre-processing). Of these, 2,009 were RORB+ and 8,780 were RORB-.

Secondary atlas — GSE138852 (Grubman et al. 2019, *Nature Neuroscience*[3]): EC snRNA-seq from 6 AD and 6 control individuals (pooled into 6 donor-pairs). Raw count matrices downloaded from GEO, accession GSE138852. Post-QC: 403 EC excitatory nuclei recovered. Only 16 were RORB+, constituting a Tier-A negative (insufficient for pseudobulk DE). Used for directional sign-replication only.

4.2 Pre-processing

Raw count matrices were loaded and processed using scanpy 1.11.5[14] and anndata 0.10.9[15]. For each dataset, cells were filtered by minimum gene count (min_genes = 200) and maximum mitochondrial fraction (max_pct_mito = 20%). Library-size normalization was performed to 10,000 counts per cell, followed by log1p transformation. Highly variable genes (HVGs) were identified using scanpy's `highly_variable_genes` function (min_mean = 0.0125, max_mean = 3.0, min_disp = 0.5). Dimensionality reduction used TruncatedSVD (n_components = 50) from scikit-learn 1.8.0[16] as a hardware-compatible substitute for PCA + UMAP (sc.tl.umap was not available due to an OpenMP/BLAS library segfault in the compute environment; figures show PC1/PC2 as layout proxy). Harmony batch correction was not attempted; the OMP/BLAS segfault was identified in the UMAP call and was not tested for the Harmony integration step specifically. An alternative integration approach (e.g., MNN via mnnpy) was not evaluated in this compute environment. Module scoring is computed from log-normalized expression,

which is less sensitive to cross-donor batch effects than cluster-based analyses; however, a batch-corrected sensitivity analysis has not been performed and is a limitation of this study. Exact package versions are frozen in `008_execute/code/environment.lock.json` (provided as Supplementary Data File S1).

EC excitatory neurons were identified using a marker-score threshold. RORB+ cells were defined as EC excitatory nuclei with log-normalized RORB expression > 0.5 (units: log(CPM/10 + 1)). Full classification code is in `008_execute/code/data_loader.py`.

4.3 Statistical analysis

Pseudobulk differential expression (H1): Per-donor pseudobulk profiles were constructed by summing raw counts across RORB+ and RORB- EC excitatory nuclei within each donor, with a log-CPM pseudocount of 1 (consistent pseudocount used for both point estimates and 95% bootstrap CIs to avoid CI inversion; the pseudocount of 1 symmetrically raises the log-CPM floor for both groups and compresses log2FC toward zero for very low-expressed genes). Log2FC was computed as the mean difference in log-CPM between RORB+ and RORB- pseudobulk profiles across donors using a two-sample t-test (scipy 1.10.1[17]). The test uses 10 RORB+ pseudobulk profiles (one per donor) and 10 RORB- pseudobulk profiles (one per donor) as two independent groups ($n = 10$ per group; 20 pseudobulk profiles total, reflected as H1.n_donors = 20 in results.json — this represents both groups combined, not 20 biological donors). 95% CIs were generated by 1,000 bootstrap resamples of *donors* (random seed 42), reflecting donor-level uncertainty. Multiple testing correction: Benjamini–Hochberg (BH) FDR correction[18] was applied first per-panel (across the 4 pre-registered H1 genes) as the primary correction, and additionally genome-wide as a conservative secondary check; both yield FDR = 1.00 for all four genes given all nominally non-significant p-values. When a pseudobulk group shows zero variance across donors for a given gene (as occurred for PVALB in the RORB+ group), the t-test produces a NaN p-value, which is treated as $p = 1.0$ before BH correction; bootstrap CIs in this case reflect only the non-zero-variance group's sampling variability and do not constitute evidence of differential expression.

ISR trajectory (H2): For each of the 10 donors in GSE147528, the mean ISR module score was computed across all RORB+ EC excitatory nuclei. An ordinary least-squares regression of mean ISR score on Braak stage (treated as an ordinal integer: 0, 2, 6) was fit using statsmodels 0.14.6[19]. 95% bootstrap CIs for the slope were generated by 1,000 resamples of *donors* (seed 42). Spearman rank correlation was computed as a non-parametric check for the linearity assumption in Braak stage coding; Spearman $\rho = -0.43$ ($p = 0.215$), consistent with the OLS slope direction and null result.

Module decomposition (H3): Cohen's d[20] was computed for excitability, ISR, proteostasis, and mitochondrial gene module scores (RORB+ vs RORB-) in both atlases. 95% CIs from cell-level bootstrap (1,000 resamples; seed 42) are reported for completeness but are anti-conservative relative to donor-level resampling in clustered designs and must not be used for formal inference about effect-size magnitude. An analytical approximation for the donor-level 95% CI, treating the 10 within-donor differences (RORB+ mean minus RORB- mean per donor) as the sampling unit in a paired-difference design, is computed as: $d \pm 1.96 \times \sqrt{1/n + d^2/(2n)}$ where $n = 10$. For the excitability module ($d = +0.269$), this yields an approximate donor-level CI of $[-0.362, +0.900]$, which spans zero. These approximate CIs are reported in Results and should be used in preference to the cell-level bootstrap bounds for inferential purposes.

AUROC and comparisons: Receiver operating characteristic analysis was performed using scikit-learn 1.8.0[16] `roc_auc_score`. Bootstrap CIs used 1,000 resamples. The delta vs baselines used the vectorized delta of 1,000 bootstrap AUROC samples; the primary-vs-random comparison uses independent (not paired) RNG states and is therefore approximate. A paired bootstrap (same cell-resample index for both AUROC estimates) would provide an exact p-value; this is recommended for future analyses.

Random seed 42 was used throughout for all bootstrap and permutation procedures.

4.4 Model construction and lockdown

The pre-registered hypotheses (H1 and H2), decision rules, and gene panels were locked before data access, with internal version-control hash SHA-256 = `1bba50d5af4fdaf42baf137eaac8eac51ad566ef92b79fad2a7a325451a98d31`. This hash documents the hypotheses.md file state within an automated research pipeline; it is not a formal external pre-registration with an independent timestamp authority (e.g., OSF, AsPredicted). The hypotheses were locked and hashed before data access within the pipeline run, but this cannot be independently verified from the hash alone. The pre-registered hypotheses file is provided as Supplementary Data File S2. All analysis code is available at `008_execute/code/main.py`,

`008_execute/code/data_loader.py`, `008_execute/code/metrics.py`, and `008_execute/code/models.py`.

4.5 Validation protocol

The RORB-ECVAP validation protocol consists of: (1) atlas QC and RORB+ recovery audit (Tier-A negative classification if n_rorb_pos < 50 per donor on average); (2) per-donor pseudobulk DE using log-CPM with a consistent pseudocount and donor-level bootstrap CIs; (3) AUROC of composite vulnerability signature vs RORB-specific marker set baseline and random gene set null; (4) ISR trajectory regression across Braak stages; (5) module decomposition (excitability, ISR, proteostasis, mitochondrial). Decision thresholds are pre-registered and applied without post-hoc adjustment. This protocol is validated here on a single primary cohort and is designed for future application to larger EC atlases.

4.6 Figure generation

Figures were generated using matplotlib 3.10.9[21] and seaborn 0.13.2[22]. Render scripts: `008_execute/code/figures/fig1_schematic.py` (study design schematic), `008_execute/code/figures/fig2_rorb_cluster.py` (EC taxonomy and RORB+ identification), `008_execute/code/figures/fig3_h1_calcium.py` (H1 forest plot and AUROC), `008_execute/code/figures/fig4_h2_isr.py` (H2 ISR trajectory), `008_execute/code/figures/fig5_mechanism_decomp.py` (module decomposition). All figures saved as PNG (300 dpi) and PDF.

Software versions (Python 3.11.15): scanpy 1.11.5[14]; anndata 0.10.9[15]; numpy 1.26.4[23]; scipy 1.10.1[17]; pandas 3.0.3[24]; scikit-learn 1.8.0[16]; statsmodels 0.14.6[19]; matplotlib 3.10.9[21]; seaborn 0.13.2[22]. (pydeseq2 and harmonypy were present in the environment but not invoked for any reported result.) The complete environment.lock.json is provided as Supplementary Data File S1.

Data Availability

All snRNA-seq data analysed in this study are publicly available from the NCBI Gene Expression Omnibus: GSE147528 (Leng et al. 2021) and GSE138852 (Grubman et al. 2019). No new data were generated.

Code Availability

All analysis code implementing the RORB-ECVAP pipeline (main.py, data_loader.py, metrics.py, models.py, and five figure generation scripts) is available from the corresponding author upon reasonable request and will be deposited at Zenodo prior to publication.

Supplementary Data

- **Supplementary Data File S1:** environment.lock.json — complete package version list and transitive dependencies for the compute environment.
- **Supplementary Data File S2:** hypotheses.md — pre-registered hypotheses file (SHA-256 = `1bba50d5af4fdaf42baf137eaac8eac51ad566ef92b79fad2a7a325451a98d31`), locked before data access within the pipeline run.

References

[1] H Braak and E Braak. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, 82(4):239–259, 1991. doi: 10.1007/BF00308809.

[2] Kun Leng, Emmy Li, Rana Eser, et al. Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. *Nature Neuroscience*, 24:276–287, 2021. doi: 10.1038/s41593-020-00764-7.

[3] Alexandra Grubman, Gabriel Chew, John F Ouyang, et al. A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nature Neuroscience*, 22:2087–2097, 2019. doi: 10.1038/s41593-019-0539-4.

[4] Jean-Pierre Roussarie, Vicky Yao, Patricia Rodriguez-Rodriguez, et al. Selective neuronal vulnerability in Alzheimer's disease: a network-based analysis. *Neuron*, 107(5):821–835.e12, 2020. doi: 10.1016/j.neuron.2020.06.010.

[5] Patricia Rodriguez-Rodriguez, Luis Enrique Arroyo-Garcia, Christina Tsagkogianni, et al. A cell autonomous regulator of neuronal excitability modulates tau in Alzheimer's disease vulnerable neurons. *Brain*, 147(7):2384–2399, 2024. doi: 10.1093/brain/awae051.

[6] Cara A Crossley, Vishaal Rajani, and Qi Yuan. Modulation of L-type calcium channels in Alzheimer's disease: a potential therapeutic target. *Computational and Structural Biotechnology Journal*, 21:1032–1044, 2022. doi: 10.1016/j.csbj.2022.11.049.

[7] Alexis M Stranahan and Mark P Mattson. Selective vulnerability of neurons in layer II of the entorhinal cortex during aging and Alzheimer's disease. *Neural Plasticity*, 2010:108190, 2010. doi: 10.1155/2010/108190.

[8] Dibyadeep Datta, Shengtao Yang, Mary Kate P Joyce, et al. Key roles of CACNA1C/Cav1.2 and CALB1/calbindin in prefrontal neurons altered in cognitive disorders. *JAMA Psychiatry*, 81(9): 870–881, 2024. doi: 10.1001/jamapsychiatry.2024.1112.

[9] Miguel A Bravo-Jimenez, Sanjay Sharma, and Soheila Karimi-Abdolrezaee. The integrated stress response in neurodegenerative diseases. *Molecular Neurodegeneration*, 20(1):20, 2025. doi: 10.1186/s13024-025-00811-6.

[10] Mariano I Gabitto, Kyle J Travaglini, Victoria M Rachleff, Eitan S Kaplan, Brian Long, and Jeanelle Ariza. Integrated multimodal cell atlas of Alzheimer's disease. *Nature Neuroscience*, 27:2366–2383, 2024. doi: 10.1038/s41593-024-01774-5.

[11] Alexey V Soloviev, Felipe L Pereira, Renata E P Leite, Claudia K Suemoto, Kun Leng, Martin Kampmann, and Lea T Grinberg. Exploring molecular signatures of vulnerable vs resilient RORB excitatory neurons of the entorhinal cortex in Alzheimer's disease. *Alzheimer's & Dementia*, 2025. doi: 10.1002/alz70855_107472. conference abstract.

[12] Alexey V Soloviev, Felipe L Pereira, Renata E P Leite, Claudia K Suemoto, Kun Leng, Martin Kampmann, and Lea T Grinberg. Exploring selective neuronal vulnerability in Alzheimer's disease — key pathways are downregulated in vulnerable RORB and CDH9 EC neurons at Braak 0. *Alzheimer's & Dementia*, 2024. doi: 10.1002/alz.091740. conference abstract.

[13] Kei M Igarashi. Entorhinal cortex dysfunction in Alzheimer's disease. *Trends in Neurosciences*, 46 (2):124–136, 2023. doi: 10.1016/j.tins.2022.11.006.

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

[15] Isaac Virshup, Sergei Rybakov, Fabian J Theis, Philipp Angerer, and F Alexander Wolf. anndata: Annotated data. *Journal of Open Source Software*, 2024. doi: 10.21105/joss.04371.

[16] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, et al. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[17] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

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

[19] Skipper Seabold and Josef Perktold. statsmodels: econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference*, pages 92–96, 2010.

[20] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 2nd edition, 1988.

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

[22] Michael L Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60): 3021, 2021. doi: 10.21105/joss.03021.

[23] Charles R Harris, K Jarrod Millman, Stéfan J van der Walt, et al. Array programming with NumPy. *Nature*, 585:357–362, 2020. doi: 10.1038/s41586-020-2649-2.

[24] Wes McKinney. Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference*, pages 56–61, 2010.