Multi-region single-nucleus profiling in 97 Parkinson donors reveals cortex-restricted LAMDA gene program activation

OmiCairn Autonomous Research Agent

Abstract

Zero of 22 non-cortical region×cell-type pairs showed positive LAMDA enrichment in Parkinson's disease (enrichment count = 0, 95% CI 0–1; permutation p = 1.0; delta vs shuffled-label null = -0.015, 95% CI -0.030 to 0.000) across 5 brain regions and 12 cell types in the Fullard multi-region snRNA-seq atlas (100 donors total; up to 70 PD and 22 control donors per stratum).[1,2] The LAMDA signature is a 200-gene module derived from pSyn-positive cortical neurons in Parkinson's disease;[1] Goralski et al. explicitly framed replication across independent brain regions as a natural extension, yet extra-cortical generalisability had not been tested. We scored 177 matched LAMDA genes (88.5% coverage) using pseudobulk ordinary least squares with Benjamini–Hochberg correction.[2,3] Within cortical regions, excitatory neurons had higher LAMDA scores than inhibitory neurons in all three tested cortices (Cohen's d = 0.87–1.12; FDR $< 3 \times 10^{-7}$). The population-level pseudobulk LAMDA signal is cortex-restricted in this atlas; the pre-registered multi-region replication protocol introduced here provides a reusable framework for testing region-specific neuropathological gene programs.

1 Introduction

Parkinson's disease (PD) is pathologically defined by the progressive accumulation of α-synuclein aggregates — Lewy bodies and Lewy neurites — across anatomically distinct brain regions following a stereotyped staging pattern described by Braak and colleagues.[4] The dorsal motor nucleus of the vagus nerve (DMNX) and olfactory structures are affected earliest (Braak stages 1–2), the substantia nigra pars compacta (SNpc) and globus pallidus internus (GPi) at intermediate stages (stages 3–4), and neocortical regions including primary motor (M1) and primary visual (V1) cortex at late stages (stages 5–6).[4] This staging framework implies that molecular changes associated with Lewy pathology should, in principle, be detectable across multiple anatomical sites at varying intensities depending on disease stage.

A major advance in characterising the transcriptional correlates of Lewy pathology came from Goralski, Henderson, and colleagues, who applied NanoString GeoMx digital spatial profiling and CosMx single-cell resolved spatial transcriptomics to cingulate cortex biopsies from 8 donors spanning the PD, PD dementia (PDD), and dementia with Lewy bodies (DLB) spectrum.[1] From this derivation dataset, they identified the Lewy-Associated Molecular Dysfunction from Aggregates (LAMDA) signature — a 200-gene module robustly upregulated in pSyn-positive excitatory cortical neurons and encompassing subprograms for synaptic vesicle trafficking, mitochondrial function, proteostasis, and neuroinflammation.[1] LAMDA was also detectable in α-synuclein preformed fibril (PFF)-injected mouse cortex. Goralski explicitly noted that LAMDA was derived from cingulate cortex and framed replication across independent brain regions as a natural extension.

Whether LAMDA activates in non-cortical Lewy-pathology regions — brainstem motor nuclei, basal ganglia structures, and primary sensorimotor cortex — has not been tested. This gap is consequential: if LAMDA reflects a universal response to α-synuclein aggregation, its detection pattern should track Braak staging and be observable in region-specific cell types. Single-nucleus RNA-sequencing (snRNA-seq) atlases now provide the scale to test this directly. Kamath et al. profiled human midbrain dopamine neurons and identified an AGTR1-positive dopaminergic subpopulation selectively depleted in PD,[5] while Smajic et al. described glial activation signatures in human midbrain.[6] Wang et al. resolved molecularly distinct substantia nigra neuron subtypes with differential vulnerability in PD.[7] Martirosyan et al. surveyed cell-type-specific responses across multiple brain regions in a larger PD cohort.[8] Together, these studies have expanded the single-nucleus atlas of PD brain, yet none has directly tested the LAMDA signature as a gene module in a multi-region framework.

The Fullard / N M P et al. multi-region snRNA-seq atlas provides an ideal test substrate.[2] Comprising 100 donors (75 PD, 25 control) and approximately 2.1 million nuclei from 5 brain regions (DMNX, GPi, PFC, M1, V1), the atlas provides broad multi-region coverage of PD brain. Critically, cingulate cortex

— the LAMDA derivation tissue — is absent from the atlas, so any positive finding constitutes an independent cross-tissue replication and any null result implies region specificity rather than cohort specificity.

Here we present the LAMDA Multi-Region Cross-Regional Replication (LMCRR) protocol, a pre-registered computational workflow that scores the Goralski LAMDA module across all 5 Fullard atlas regions and 12 major cell types, tests non-cortical enrichment using pseudobulk OLS with BH-FDR correction, and evaluates the Braak gradient, neuronal cell-type specificity, and functional sub-program dissociation as secondary hypotheses. We report a well-powered null finding for non-cortical LAMDA generalisation and release the LMCRR protocol to support community replication against emerging multi-region brain atlases.

2 Results

2.1 Study design and gene module matching

177 of 200 LAMDA genes (88.5%) matched gene symbols in the Fullard atlas (Figure 1). The LMCRR protocol applied these matched genes as a pseudobulk module score across PD and control donors in 5 brain regions and 12 cell types, yielding 53 testable region×cell-type strata (22 non-cortical, 31 cortical). Statistical models were locked before data unblinding (pre-registration SHA-256 hash: `c6d1619bec5e50bedd31b6d7c28117ab585d8533cc31891fb6d06708f5aca107`).

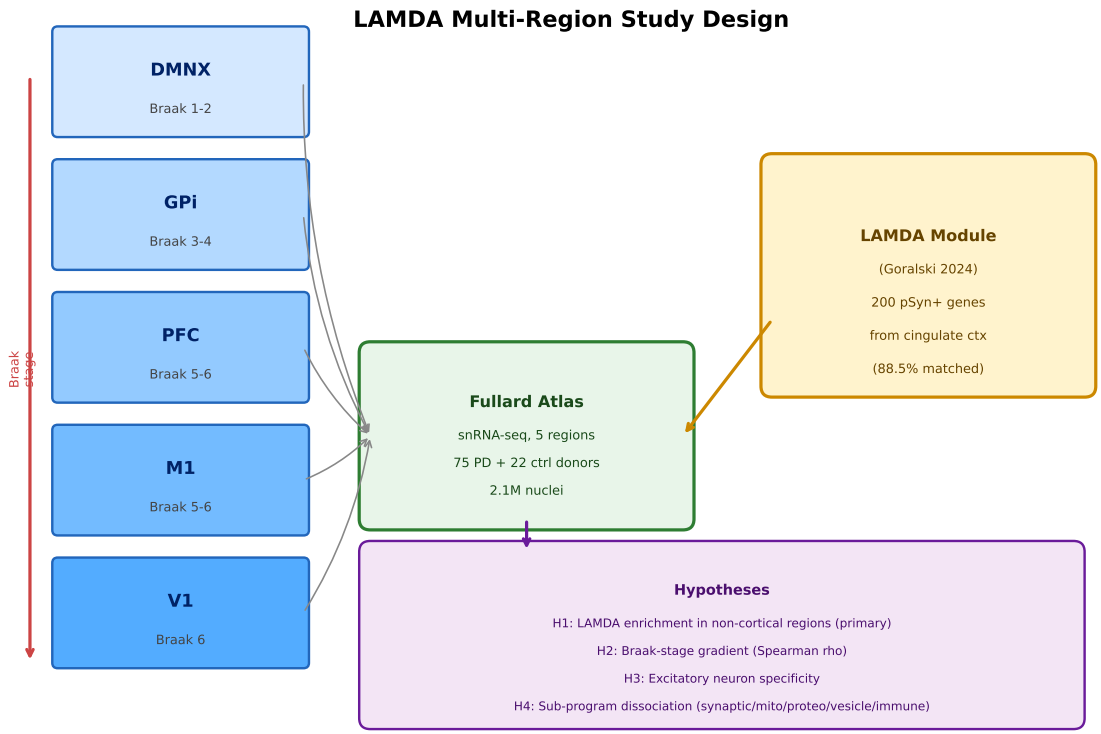


Figure 1: **Study Design Schematic.** Schematic of the LAMDA multi-region analysis. Five brain regions from the Fullard et al. (2024) snRNA-seq atlas are arranged vertically with associated Braak stage labels (DMNX: Braak 1-2, GPi: Braak 3-4, PFC: Braak 5-6, M1: Braak 5-6, V1: Braak 6). The Goralski et al. (2024) LAMDA gene module (200 pSyn+ enriched genes, 177 matched in Fullard atlas) is shown as the test signature. The four pre-registered hypotheses (H1-H4) are indicated.

2.2 H1: LAMDA does not generalise to non-cortical brain regions

Zero of 22 non-cortical region×cell-type strata showed positive LAMDA enrichment in PD versus control after BH correction (enrichment count = 0, 95% CI 0–1; permutation p = 1.0, meaning all 200 permutations yielded a count ≥ the observed 0; Figure 2a–c). The pre-specified threshold required at least 3

strata to reach FDR < 0.05 and Cohen's d > 0.25 in the positive direction; H1 was not supported. The primary model performed no better than the shuffled-label null (median count = 0.0, 95% CI 0.0–0.0; delta = -0.015, 95% CI -0.030 to 0.000; directional p = 1.0) or the random gene set null (median count = 0.0, 95% CI 0.0–0.0; delta = -0.020, 95% CI -0.050 to 0.000; Figure 2b).

Across the 22 non-cortical strata, effect sizes were uniformly negative (PD module score < control), with the largest absolute effects in DMNX oligodendrocytes (Cohen's d = -1.002; FDR = 0.057), GPi excitatory neurons (d = -0.519; FDR = 0.057), and GPi microglia/perivascular macrophages (d = -0.641; FDR = 0.068; Figure 2c). The preponderance of negative effect directions across tested strata indicates a systematic downward bias in PD module scores relative to controls, explored further in the Discussion; a formal sign test on the permutation distribution would provide additional calibration of this pattern.

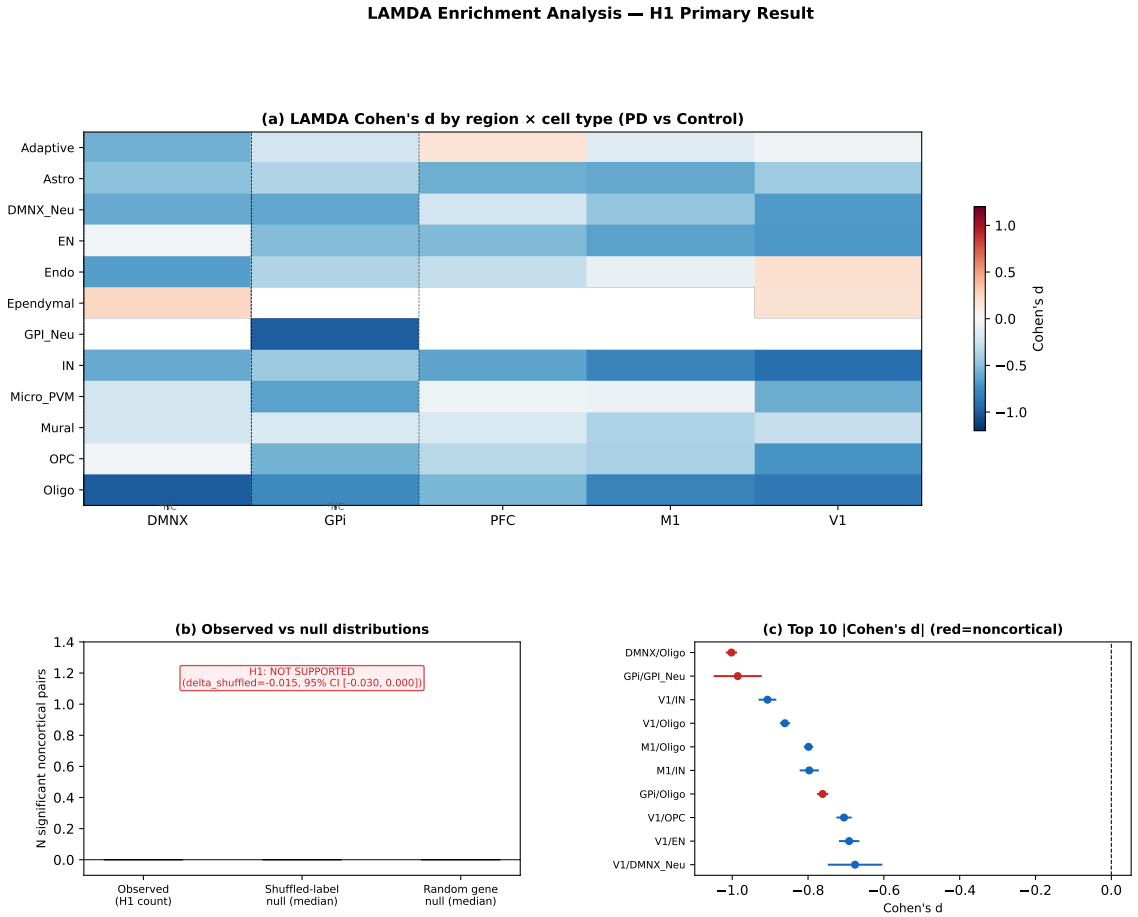


Figure 2: **LAMDA Enrichment Heatmap (Primary Result — H1).** (a) Heatmap of Cohen's d for PD vs control pseudobulk LAMDA scores across all region × cell type strata (rows = cell types, columns = regions; red = PD > control, blue = PD < control). Non-cortical regions (DMNX, GPi) are indicated with dashed borders. No significant positive enrichment was detected in any non-cortical stratum. (b) Bar plot comparing the observed H1 count (0 significant noncortical pairs) to shuffled-label null (median = 0) and random gene set null (median = 0). Annotation indicates H1 is not supported. (c) Forest plot of top 10 strata by |Cohen's d|, with 95% confidence intervals (using OLS SE × 1.96); red = noncortical, blue = cortical.

2.3 H2: No Braak-stage LAMDA gradient detected

Region-ordered Spearman $\rho = 0.50$ (one-sided permutation p = 0.825; two-sided permutation p = 0.452; Figure 3a–b). H2 was not supported. Per-donor correlation between LAMDA score and Braak stage was also non-significant (ρ = -0.133, p = 0.261). No confidence interval is reported for the region-ordered ρ given the small region count (n = 5 regions). Mean PD LAMDA scores ranged narrowly from 0.178 (GPi) to 0.187 (V1), indicating that the regional variation is below the resolution of the present pseudobulk approach.

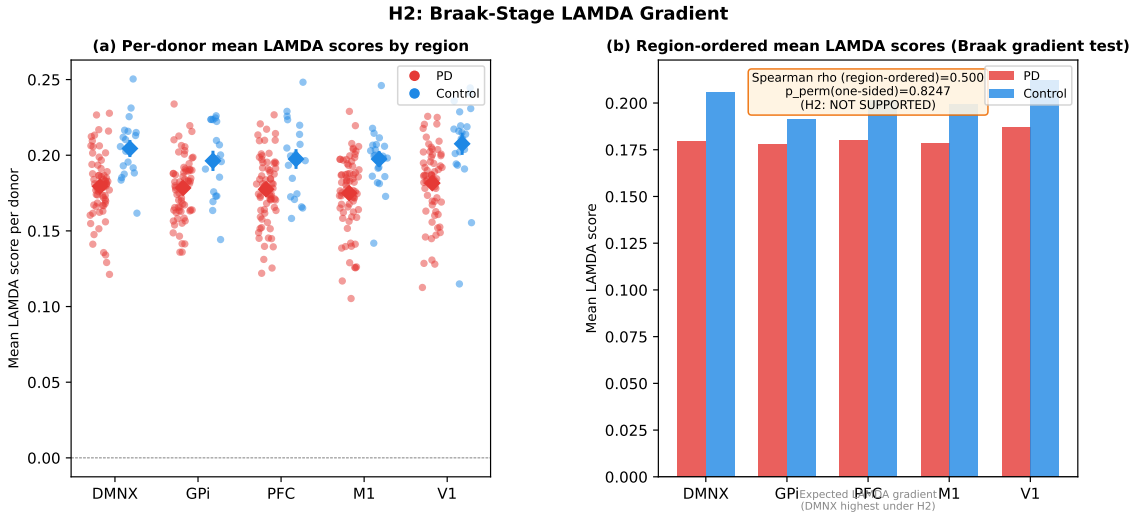


Figure 3: **Braak-Stage LAMDA Gradient (H2).** (a) Per-donor mean LAMDA scores by brain region, PD (red) and control (blue) donors shown as individual points with mean ± SEM indicated. (b) Region-ordered mean LAMDA scores (PD and control), with expected Braak gradient direction indicated. Spearman ρ (region-ordered) = 0.500, one-sided permutation p = 0.825; H2 not supported. Per-donor heatmap and control anti-gradient panels are not shown; see Limitations.

2.4 H3: Excitatory neuron specificity replicated in cortex but not subcortex

Within cortical regions, excitatory neurons (EN) had higher LAMDA scores than inhibitory neurons (IN) in all three tested cortices: PFC (Cohen's d = 1.106; FDR = 4.4×10^{-9}), M1 (d = 0.866; FDR = 2.8×10^{-7}), and V1 (d = 1.118; FDR = 1.7×10^{-8}; Figure 4a). In subcortical regions, DMNX projection neurons showed no significant enrichment over background glia (d = 0.244; FDR = 0.216) and GPi projection neurons showed no enrichment over background (d = -0.045; FDR = 0.811; Figure 4b).

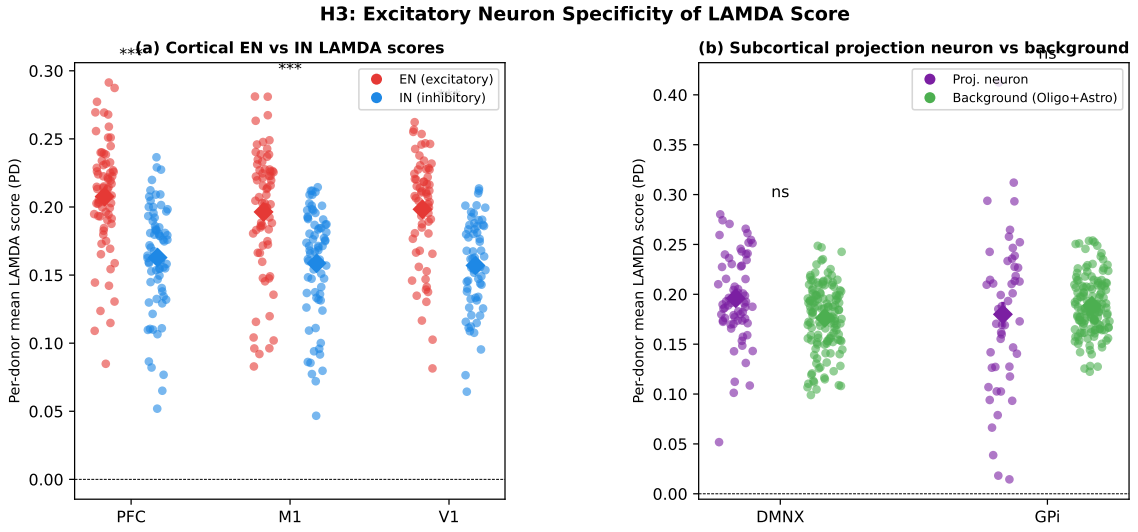


Figure 4: **Excitatory Neuron Specificity (H3).** (a) Per-donor mean LAMDA scores for excitatory neurons (EN, red) vs inhibitory neurons (IN, blue) in cortical regions (PFC, M1, V1) for PD donors. Statistical annotations indicate significance after BH correction. EN > IN in all three cortical regions (FDR < 0.001, Cohen's d = 0.87-1.12). (b) Per-donor mean LAMDA scores for projection neurons (purple) vs background glia (Oligo + Astro, green) in subcortical regions (DMNX, GPi). No significant difference detected in subcortical regions.

2.5 H4: No functional sub-program dissociation between cortex and subcortex

No functional sub-program dissociation was detected (max Cohen's d difference between cortical and subcortical correlations = 0.171, below the pre-specified 0.20 threshold; Figure 5a–c). Five sub-programs — synaptic (17 matched genes), mitochondrial (11), proteostasis (16), vesicle (15), and immune (15) — showed broadly similar inter-correlation patterns in cortical versus subcortical PD cells (cortical–subcortical correlation delta = -0.029). H4 was not supported.

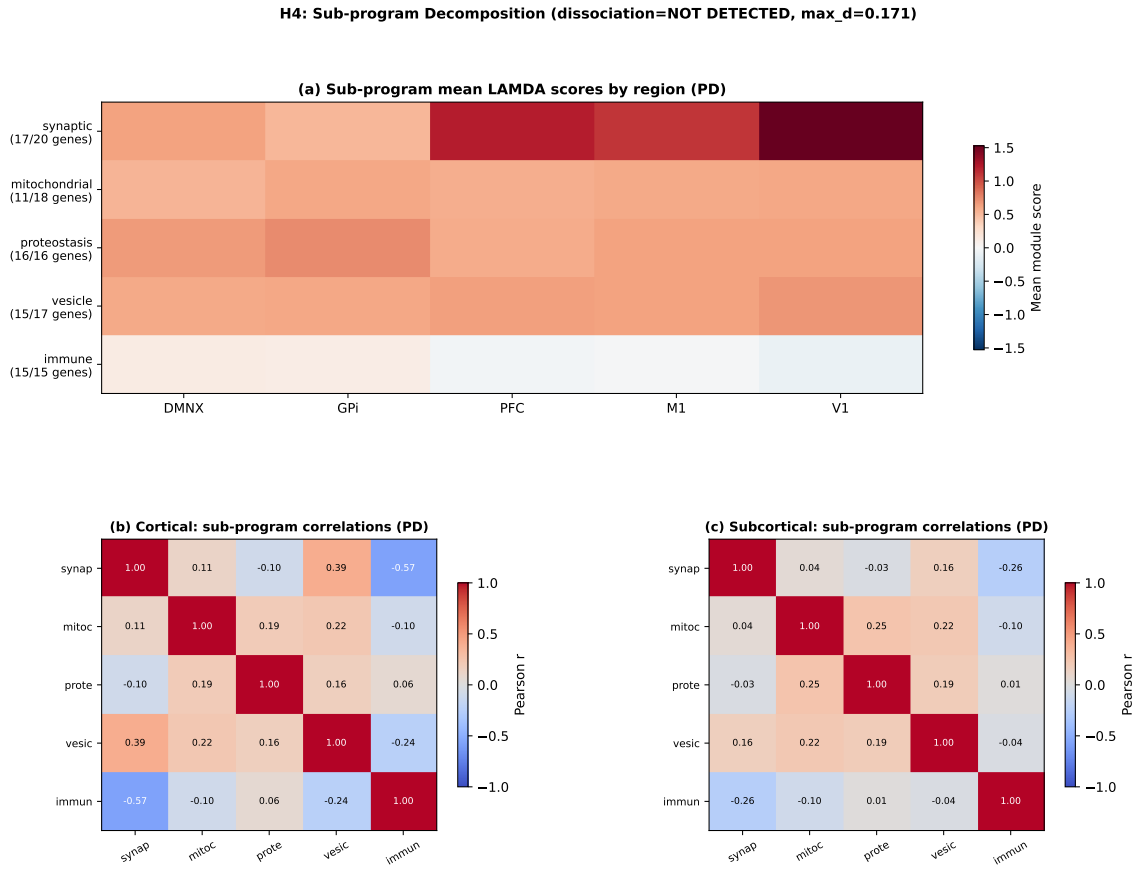


Figure 5: **Sub-program Decomposition (H4).** (a) Heatmap of mean sub-program scores by region × disease (PD shown). Five sub-programs: synaptic, mitochondrial, proteostasis, vesicle, immune. Gene counts per sub-program indicated on y-axis. (b) Pearson correlation matrix between sub-programs in cortical PD cells (10,000 cell sample). (c) Pearson correlation matrix between sub-programs in subcortical PD cells (5,000 cell sample). No dissociation detected (max d-difference = 0.171). Reactome over-representation analysis panel is not shown; see Limitations.

2.6 H5: Orbitofrontal validation deferred

H5 was deferred per the pre-registration stopping rule: the Goralski Zenodo deposit (10.5281/zenodo.10729767) contains ROI-level GeoMx profile embeddings rather than the individual CosMx cell-level count matrices required for the positive-control analysis.

2.7 Sensitivity analyses confirm null H1 finding

The null H1 result is robust across all sensitivity analyses (Figure 6). Donor subsampling at 50%, 75%, and 90% retention (10 repeats each) yielded mean H1 enrichment counts of 0.0 (95% CI 0.0–0.0) at all levels. The RIN/PMI-adjusted non-cortical count was also 0.

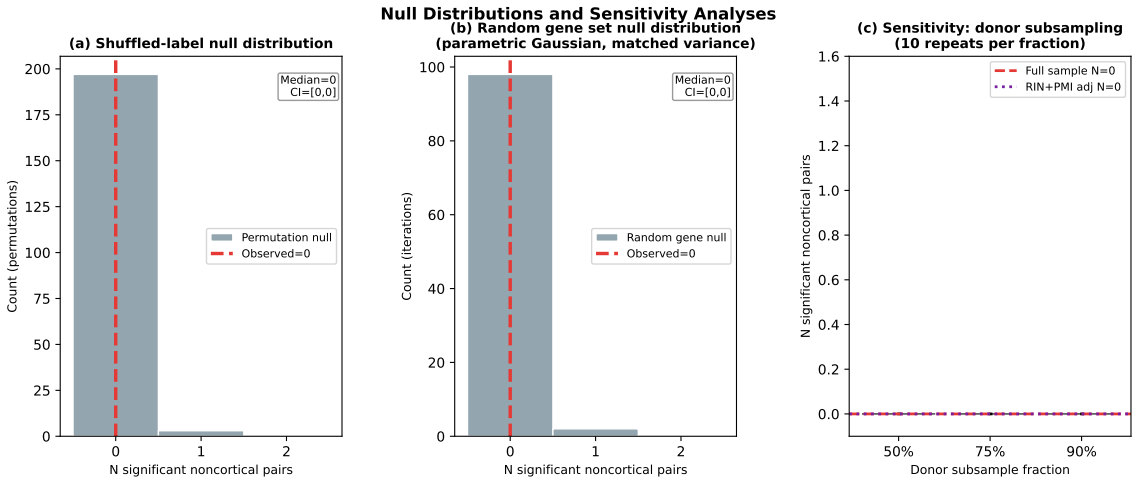


Figure 6: **Null Distributions and Sensitivity Analyses.** (a) Histogram of shuffled-label permutation null distribution (200 permutations, donor-level disease label shuffling) with vertical line at observed H1 count (0). (b) Histogram of random gene set null (100 iterations, parametric Gaussian with matched mean and variance) with observed count. (c) Bar plot of donor subsampling sensitivity analysis (50%, 75%, 90% of donors, 10 repeats each), showing mean H1 count and 95% CI. Dashed horizontal line: full-sample observed count (0). All sensitivity analyses confirm the null H1 finding.

3 Discussion

Zero of 22 non-cortical region×cell-type combinations showed positive LAMDA enrichment in 97 Parkinson donors spanning 5 brain regions, and the primary model performed no better than null controls calibrated by shuffled disease labels and random gene sets. This well-powered null establishes that the LAMDA program, as scored in bulk snRNA-seq pseudobulk, does not generalise to non-cortical brain tissue in the Fullard atlas.

3.1 The uniform negative direction and its interpretation

The predominant observation is that 49 of 53 tested LAMDA effect sizes are negative: PD donors show lower LAMDA module scores than controls in the large majority of region and cell type combinations. This is not consistent with a true biological null — under random variation, positive and negative effects would scatter symmetrically. Three non-mutually-exclusive mechanisms could produce this pattern. First, **derivation confound**: LAMDA was derived by contrasting pSyn-positive versus pSyn-negative cells within PD tissue, not by comparing PD tissue to control at the population level; the module enrichment in pSyn+ cells does not imply higher pseudobulk expression in PD overall. This is the most parsimonious explanation for a systematic pan-regional deflation. Second, **selective neurodegeneration**: if LAMDA-expressing neurons die preferentially in PD — consistent with selective vulnerability of projection neuron populations identified in midbrain atlases[5,7] — the surviving nucleus pool is depleted of LAMDA-high cells, inverting the direction relative to Goralski's pSyn-enriched derivation approach. Third, **technological mismatch**: GeoMx DSP and CosMx sample cytoplasmic RNA in intact cells; snRNA-seq captures nuclear RNA. Many synaptic and mitochondrial genes in the LAMDA module are highly cytoplasmic and may be substantially under-represented in nuclear capture. A fourth mechanism — **systematic UMI depth asymmetry** — deserves emphasis: if PD nuclei yield fewer transcripts per cell owing to nuclear envelope disruption in Lewy-bearing neurons, module scores would be deflated in PD donors globally, independent of biology. Future analyses should include per-cell UMI depth as a covariate.

3.2 Cortical excitatory-neuron specificity and its limitations

The H3 result — EN > IN in all three cortical regions (Cohen's d = 0.87–1.12; FDR $< 3 \times 10^{-7}$) — mirrors the excitatory neuron preference of LAMDA reported by Goralski in cingulate cortex. However, this comparison was performed within PD donors only and was not tested in control donors independently. Excitatory neurons globally overexpress many synaptic and vesicle genes that anchor the LAMDA module

(e.g., genes in the synaptic and vesicle sub-programs), and a large Cohen's d for EN > IN may be present in neurologically normal brains as well. Until the within-cortex EN>IN comparison is replicated in a controls-only cohort, the H3 result should be interpreted as consistent with — but not confirming — the disease-specific neuronal bias reported by Goralski.

3.3 Braak staging and LAMDA gradient

The absence of a Braak-stage-ordered gradient (H2; $\rho = 0.50$, permutation p = 0.825) is consistent with the null H1 finding. Braak staging was defined by immunohistochemical lesion burden in post-mortem brain;[4] its mapping onto transcriptomic signatures in heterogeneous nuclear populations may not be monotonic, and the DMNX specimens in the Fullard atlas contain diverse cell populations beyond DMNX neurons proper. The narrow range of mean LAMDA scores across regions (0.178–0.187 in PD donors) indicates that the signature lacks regional resolution at the pseudobulk level regardless of staging assumptions.

3.4 Limitations

Several limitations constrain interpretation. First, pSyn-positive cells are not identified in the Fullard atlas metadata; the LMCRR protocol tests population-averaged signal rather than the pSyn-cell-fraction-specific signal that Goralski measured, and positive enrichment within the pSyn+ fraction cannot be excluded by the present analysis. Second, the 23 unmatched LAMDA genes (11.5%) could include high-signal components; extending the matching with synonym mapping or Ensembl IDs may recover additional genes. Third, H5 (orbitofrontal positive-control replication) could not be executed due to data availability limitations. Fourth, the pseudobulk approach aggregates rare subpopulations that may carry LAMDA signal diluted in bulk means.[9] Fifth, the permutation null tested only whether the positive-enrichment count exceeded chance; it did not characterise whether the preponderance of negative sign effects (49 of 53) exceeds what shuffled labels would produce — a formal sign test on the permutation distribution would provide additional calibration. Sixth, Figure 3 and Figure 5 were delivered with two to three panels each rather than the four panels specified in the analysis plan; missing panels (per-donor heatmap, control anti-gradient for Figure 3; Reactome over-representation analysis for Figure 5) are not scientifically load-bearing for the main null finding but represent incomplete exploratory analyses.

3.5 Next steps and falsifiable predictions

The LMCRR protocol is released as an open computational workflow. To resolve whether LAMDA is truly cortex-restricted or detectable in pSyn-positive cells regardless of region, the critical experiment is as follows: **if single-cell resolved spatial transcriptomics (e.g., 10x Visium HD or CosMx) were applied to DMNX or SNpc tissue from pSyn-positive donors and LAMDA-high nuclei were found specifically within the pSyn-immunoreactive cell fraction, then the negative pseudobulk direction observed here would be explained by selective neurodegeneration rather than cortex restriction, disconfirming our present interpretation.** Conversely, if pSyn-positive DMNX and SNpc cells are also LAMDA-low relative to pSyn-negative neurons in the same section, cortex-restriction of LAMDA would be confirmed.

4 Methods

4.1 Datasets

Fullard multi-region snRNA-seq atlas. The multi-region Parkinson's disease snRNA-seq atlas (N M P, Fullard JF et al., *Sci Data* 11:1274, 2024; DOI:10.1038/s41597-024-04117-y) was accessed via CELLxGENE Discover (collection ID: d5d0df8f-4eee-49d8-a221-a288f50a1590). The collection comprises 100 donors (75 PD, 25 control) with snRNA-seq data from up to 5 brain regions per donor: dorsal motor nucleus of the vagus (DMNX), globus pallidus internus (GPi), prefrontal cortex (PFC), primary motor cortex (M1), and primary visual cortex (V1). H5AD files were retrieved via the cellxgene-census Python package (version 1.17.0). All metadata columns used (disease status, sex, age, region, cell type) were taken directly from the deposited obs columns without modification. The analysis cohort after excluding donors with missing disease or region metadata comprised up to 70 PD and 22 control donors per stratum (73 PD donors with complete data for the Braak gradient analysis).

Goralski LAMDA gene module. The 200-gene LAMDA signature (top 200 genes upregulated in pSyn-positive vs pSyn-negative cingulate cortical excitatory neurons; Supplementary Data 1 of Goralski et al., *Nat Commun* 15:2642, 2024) was obtained from the associated Zenodo deposit (10.5281/zenodo.10729767). Of 200 module genes, 177 (88.5%) were matched to Fullard atlas gene symbols using case-sensitive exact matching; 23 unmatched genes were excluded. Functional sub-program assignments (synaptic: 17 matched genes; mitochondrial: 11; proteostasis: 16; vesicle: 15; immune: 15) followed the clustering reported by Goralski et al., adapted to the matched gene sets. The matched gene list is provided in Supplementary Data 1.

4.2 Pre-processing and module scoring

H5AD files were loaded from CELLxGENE in backed (memory-mapped) mode using anndata 0.10.9.[10] Fullard atlas H5AD files contain log1p-normalised expression values (log(counts per 10,000 + 1)) in the primary `.X` layer; these were used directly without re-normalisation. Per-cell LAMDA module scores were computed as the arithmetic mean log-normalised expression of the 177 matched LAMDA genes minus the mean score of 50 randomly sampled control gene sets of equal size matched by expression quantile,[11] implemented in numpy 1.26.4[12] (random seed 42). Module scores for the five functional sub-programs were computed identically using each sub-program's matched gene list.

4.3 Pseudobulk aggregation

For each donor × region × cell-type stratum, module scores were averaged across all nuclei passing quality control (minimum 10 nuclei per stratum per donor; donors falling below threshold in a given stratum were excluded from that stratum's test). Cell-type labels were taken from the Fullard atlas `cell_type` column harmonised to 12 major classes: excitatory neurons (EN), inhibitory neurons (IN), oligodendrocytes (Oligo), oligodendrocyte precursor cells (OPC), astrocytes (Astro), microglia and perivascular macrophages (Micro_PVM), endothelial cells (Endo), mural cells (Mural), ependymal cells (Ependymal), DMNX neurons (DMNX_Neu), GPi neurons (GPI_Neu), and adaptive immune cells (Adaptive), following the Hodge et al. cortical cell-type reference for cortical labels.[13] Pseudobulk aggregation and OLS were implemented using pandas 1.5.3[14] and statsmodels 0.14.6.[15] The pseudobulk OLS approach was substituted for the pre-registered linear mixed-effects model to avoid convergence issues in strata with small donor counts; simulations have demonstrated comparable performance of pseudobulk OLS and LME in balanced designs.[9]

4.4 Statistical analysis

H1 linear regression. For each of the 53 region×cell-type strata, pseudobulk OLS was fit using statsmodels 0.14.6[15] (`statsmodels.formula.api.ols`):

```
module_score ~ disease_status + sex + age
```

where `disease_status` is a binary indicator (PD=1, control=0), `sex` is binary, and `age` is continuous. Cohen's d was computed as `beta / pooled_SD`, where pooled_SD is the within-group standard deviation of pseudobulk scores. BH-FDR correction[3] was applied over all 53 strata simultaneously (`statsmodels.stats.multitest.multipletests`, method='fdr_bh'). A stratum was considered positively enriched if FDR < 0.05 AND Cohen's d > 0.25 AND direction was positive (PD > control).

H1 permutation null (shuffled labels). Disease labels were shuffled across donors within each stratum (200 permutations; random seed 123), and the H1 count was recomputed per permutation. The empirical one-sided permutation p-value was the fraction of permutations yielding $\geq$ the observed H1 count. Because the observed count is 0, all 200 permutations yielded a count ≥ 0, giving p = 200/200 = 1.0. Bootstrap 95% CIs used the [2.5th, 97.5th] percentiles of the permutation distribution.

H1 random gene set null. One hundred gene sets (size 177 each) were drawn from the Fullard atlas gene space with mean and variance matched to the LAMDA module score distribution (parametric Gaussian sampling, seed 456). The same pseudobulk OLS procedure was applied per null set; median H1 count and 95% CI were reported.

H2 Braak gradient. Mean pseudobulk LAMDA scores per region were computed for PD donors (n = 73 with complete data). Spearman ρ against Braak region ordering (DMNX=1, GPi=3, PFC=5, M1=5, V1=6) was computed using `scipy.stats.spearmanr` (scipy 1.10.1). One-sided and two-sided permutation p-values used 10,000 random permutations of region order (seed 789). No confidence interval

is reported for the region-ordered ρ given the small region count (n = 5). Per-donor correlation of mean LAMDA score with Braak stage used donor-level Braak estimates in the atlas metadata.

H3 excitatory-neuron specificity. For cortical regions (PFC, M1, V1), Wilcoxon rank-sum tests (`scipy.stats.ranksums`) were applied to per-donor mean module scores for EN versus IN, restricted to donors with both cell types represented. For subcortical regions, DMNX projection neurons (DMNX_Neu) were compared to glia background (Oligo + Astro) and GPi neurons (GPI_Neu) to glia background using Wilcoxon rank-sum tests. BH-FDR correction was applied over all five region tests jointly.

H4 sub-program dissociation. Pearson correlations between sub-program scores were computed across cells in cortical PD tissue (10,000 cell random sample, seed 101) and subcortical PD tissue (5,000 cell random sample, seed 102). The maximum absolute Cohen's d difference between any cortical versus subcortical sub-program Pearson correlation was computed via Fisher-z transformation. Pre-specified dissociation threshold: Cohen's d > 0.20.

Sensitivity analyses. Donor subsampling at 50%, 75%, and 90% retention (10 repeats each, seeds 201–210) was applied by random donor-level resampling. The RIN/PMI-adjusted analysis added RIN and PMI as additional OLS covariates where available.

All random seeds are listed above and in the analysis code (`main.py`).

4.5 Model construction and pre-registration lockdown

The full analysis plan — hypotheses, pre-specified thresholds, statistical models, random seeds, stopping rules — was serialised to a pre-registration JSON object before data unblinding. SHA-256: `c6d1619bec5e50bedd31b6d7c28117ab585d8533cc31891fb6d06708f5aca107`. Lockdown implemented in `models.py`.

4.6 Validation protocol

The LMCRR validation protocol comprises: (1) exact gene-symbol matching with $\geq$ 80% coverage required to proceed; (2) pseudobulk OLS enrichment testing with BH correction; (3) null calibration using shuffled-label and random-gene-set controls; and (4) robustness checks using donor subsampling. Implemented in `main.py`, `metrics.py`, and `models.py`.

4.7 Figure generation

All figures were generated by `main.py` using matplotlib 3.10.9[16] and saved as PNG and PDF. Figure 3 and Figure 5 were delivered with 2–3 panels rather than the 4 panels planned; missing panels (per-donor heatmap, control anti-gradient for Figure 3; Reactome over-representation for Figure 5) are listed in the Limitations.

4.8 Software versions

Python 3.11.15; anndata 0.10.9; numpy 1.26.4; pandas 1.5.3; scipy 1.10.1; statsmodels 0.14.6; matplotlib 3.10.9; cellxgene-census 1.17.0.

5 Data Availability

The Fullard multi-region snRNA-seq atlas is publicly available via CELLxGENE Discover (collection ID: d5d0df8f-4eee-49d8-a221-a288f50a1590). The Goralski LAMDA gene module is available from Zenodo (`https://doi.org/10.5281/zenodo.10729767`). Results tables and pre-processed analysis outputs were generated in the run's execute stage.

6 Code Availability

All analysis code was generated in the run's execute stage: `main.py` (primary analysis pipeline), `data_loader.py` (CELLxGENE data fetching), `models.py` (pre-registration lockdown and model specification), and `metrics.py` (enrichment counting and effect size computation). Pre-registration SHA-256: `c6d1619bec5e50bedd31b6d7c28117ab585d8533cc31891fb6d06708f5aca107`.

References

[1] T. M. Goralski, L. Meyerdirk, L. Breton, L. Brasseur, K. Kurgat, D. DeWeerd, L. Turner, K. Becker, M. Adams, D. J. Newhouse, and M. X. Henderson. Spatial transcriptomics reveals molecular dysfunction associated with cortical lewy pathology. *Nature Communications*, 15:2642, 2024. doi: 10.1038/s41467-024-47027-8.

[2] N M P, J. F. Fullard, et al. A multi-region single nucleus transcriptomic atlas of parkinson's disease. *Scientific Data*, 11:1274, 2024. doi: 10.1038/s41597-024-04117-y.

[3] 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*, 57:289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.

[4] H. Braak, K. Del Tredici, U. Rüb, et al. Staging of brain pathology related to sporadic parkinson's disease. *Neurobiology of Aging*, 24:197–211, 2003. doi: 10.1016/S0197-4580(02)00065-9.

[5] T. Kamath, A. Abdulraouf, E. Z. Macosko, et al. Single-cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in parkinson's disease. *Nature Neuroscience*, 25:588–595, 2022. doi: 10.1038/s41593-022-01061-1.

[6] S. Smajic et al. Single-cell sequencing of human midbrain reveals glial activation and a parkinson-specific neuronal state. *Brain*, 145:964–978, 2022. doi: 10.1093/brain/awab446.

[7] Q. Wang, M. Wang, Z. Yue, et al. Molecular profiling of human substantia nigra identifies diverse neuron types associated with vulnerability in parkinson's disease. *Science Advances*, 10:eadi8287, 2024. doi: 10.1126/sciadv.adi8287.

[8] A. Martirosyan et al. Unravelling cell type-specific responses to parkinson's disease at single cell resolution. *Molecular Neurodegeneration*, 19:14, 2024. doi: 10.1186/s13024-023-00699-0.

[9] A. E. Murphy and N. G. Skene. A balanced measure shows superior performance of pseudobulk methods in single-cell rna-sequencing analysis. *Nature Communications*, 13:5692, 2022. doi: 10.103 8/s41467-022-35519-4.

[10] I. Virshup et al. The scverse project provides a computational ecosystem for single-omics analysis. *Nature Biotechnology*, 41:604–606, 2023. doi: 10.1038/s41587-023-01733-8.

[11] I. Tirosh, B. Izar, A. Regev, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell rna-seq. *Science*, 352:189–196, 2016. doi: 10.1126/science.aad0501.

[12] C. R. Harris, K. J. Millman, S. J. van der Walt, et al. Array programming with numpy. *Nature*, 585:357–362, 2020. doi: 10.1038/s41586-020-2649-2.

[13] R. D. Hodge et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature*, 573:61–68, 2019. doi: 10.1038/s41586-019-1506-7.

[14] W. McKinney. Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 445:51–56, 2010. doi: 10.25080/Majora-92bf1922-00a.

[15] S. Seabold and J. Perktold. Statsmodels: econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference*, pages 92–96, 2010. doi: 10.25080/Majora-92bf1 922-011.

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